Preface

From what I know, there are two "correct" ways to teach linear algebra: *Linear Algebra Done Right* and *Linear Algebra Done Wrong*. These are named after the books written about them by Sheldon Axler and Sergei Treil respectively. The difference between the two books is the use of matrices. In *Linear Algebra Done Right*, a focus on linear operators/transformations is presented to justify all the elements of linear algebra one may be familiar with, with little attention to matrices. *Linear Algebra Done Wrong* wants a more practical understanding of linear algebra and introduces matrices early in order to justify many other elements of linear algebra. Both are completely valid ways of teaching linear algebra and they simply serve different audiences. These notes will largely be based on *Linear Algebra Done Wrong* with a few elements of *Linear Algebra Done Right* here and there where intuition for a topic is deemed necessary. So we will only discuss finite dimensional vector spaces (mostly $\mathbb{R}^N$ and $\mathbb{C}^N$) and will use matrices to justify many of the ideas such as determinants and eigenvalues.

Contents

Chapter 1

Basics

1.1 Fields

Definition 1.1.1. A field $\mathbb{F}$ is a set endowed with operations addition and multiplication satisfying:

1. Associativity: $\forall a, b, c \in \mathbb{F},\ a + (b + c) = (a + b) + c,\ (ab)c = a(bc)$
2. Commutativity: $\forall a, b \in \mathbb{F},\ a + b = b + a,\ ab = ba$
3. Additive Identity: $\forall a \in \mathbb{F}, a + 0 = a$
4. Multiplicative Identity: $\forall a \in \mathbb{F},\ a \times 1 = a$
5. Additive Inverse: $\forall a, \exists b, a + b = 0$
6. Multiplicative Inverse: $\forall a \neq 0, \exists b, a \times b = 1$
7. Distributive: $a(b + c) = ab + ac$

You should notice that there isn't really much that's special about fields. It is essentially just a fancy name for the number systems we are used to and the rules we've been using since elementary school.

Example 1.1.2. The following are examples of fields:

1. Real Numbers
2. Complex Numbers
3. $\mathbb{F}_2$ (this is the set of all integers modulo 2)

Importantly, the naturals and integers are not fields. The naturals lack an additive inverse and both lack a multiplicative inverse. It should also be noted that $\mathbb{F}_p$ for some prime p is a vector field since you are guaranteed to have a multiplicative inverse when you are mod a prime.

1.2 Vector Spaces

Definition 1.2.1. A vector space V over some field $\mathbb{F}$ is a set V endowed with the operations of vector addition and scalar multiplication and satisfies the following axioms:

1. Associativity: $(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$
2. Commutativity: $\vec{u}+\vec{v}=\vec{v}+\vec{u}$
3. Additive Identity: $\exists\vec{0}\in V, \vec{0}+\vec{u}=\vec{u}$
4. Additive Inverse: $\forall\vec{u}, \exists\vec{v}, \vec{u}+\vec{v}=0$, we denote $\vec{v}$ as $-\vec{u}$
5. Multiplicative Associativity: $\alpha,\beta\in\mathbb{F}, \alpha(\beta\vec{v})=(\alpha\beta)\vec{v}$
6. Identity Multiplication: $1\times\vec{v}=\vec{v}$
7. Vector Over Scalar Distributivity: $\alpha,\beta\in\mathbb{F}, (\alpha+\beta)\vec{v}=\alpha\vec{v}+\beta\vec{v}$
8. Scalar Over Vector Distributivity: $\alpha\in\mathbb{F}, \alpha(\vec{u}+\vec{v})=\alpha\vec{u}+\alpha\vec{v}$
9. Closure Under Scalar Multiplication: $\forall\alpha\in\mathbb{F}, \vec{v}\in V, \alpha\vec{v}\in V$
10. Closure Under Vector Addition: $\forall\vec{v},\vec{u}\in V, \vec{u}+\vec{v}\in V$

Remark 1.2.2. What it means for a vector space to be over some field $\mathbb{F}$ means that the scalars are elements from $\mathbb{F}$

Example 1.2.3. These are examples of vector spaces:

1. $\mathbb{R}^2=\left\{\begin{pmatrix}x\\y\end{pmatrix}: x,y\in\mathbb{F}\right\}$
2. The set of all continuous functions $f:[0,1]\to\mathbb{R}$ over $\mathbb{R}$
3. The set of all polynomials of degree 5 over $\mathbb{R}$

This example should make it clear that not all vector spaces contain what we normally think of as vectors. The set of all continuous functions is very strange to think of as a vector space since you can't really represent a function in our typical $\begin{pmatrix}\cdot\\\cdot\end{pmatrix}$ format. There are also many other abstract vector spaces so when you are given an arbitrary vector space V over an arbitrary field $\mathbb{F}$ you can't make any assumptions outside of the axioms. Now sometimes you can still represent abstract vector spaces in our standard vector format as the below remark shows.

Remark 1.2.4. Take the polynomial $4x^5+x^4-x^3+2x^2+7x+15$. We can encode this as a vector via $(4,1,-1,2,7,15)$ where each component represents the coefficient of the respective degrees in the polynomial.

Vector space axioms do not explicitly contain everything about vectors. However, they are written in such a way that any standard rule of vectors can be derived from them. Below is an example:

Theorem 1.2.5. Let V be a vector space. For all $\vec{v} \in V$, we have that $0 \cdot \vec{v} = \vec{0}$

Proof.

$$\begin{aligned}
0 \cdot \vec{v} &= 0 \cdot \vec{v} \\
0 \cdot \vec{v} &= (0+0) \cdot \vec{v} && \text{(from Additive Identity)} \\
0 \cdot \vec{v} &= 0 \cdot \vec{v} + 0 \cdot \vec{v} && \text{(from Vector Over Scalar Distributivity)} \\
0 \cdot \vec{v} - 0 \cdot \vec{v} &= 0 \cdot \vec{v} \\
\vec{0} &= 0 \cdot \vec{v} && \text{(from Additive Inverse)}
\end{aligned}$$

□

Remark 1.2.6. If a problem says a vector space V over a field $\mathbb{F}$, you can make no assumptions about the vector space and field outside of the axioms. However, in general, it is safe to assume that if a vector space and field are not specifically listed, you are in $\mathbb{R}^N$ over $\mathbb{R}$

Vector spaces are often contain other vector spaces. For example, if we consider the $\mathbb{R}^2$ over $\mathbb{R}$, it contains the smaller vector space of all vectors which have the same first and second coordinate. One can check that the vector space axioms hold for this. This leads us to the definition of a subspace.

Definition 1.2.7. Let V be a vector space. U is a subspace of V if and only if $U \subseteq V$ and U satisfies all the vector space axioms.

Remark 1.2.8. When proving something is a subspace, you don't actually need to go through and check all the axioms again. We already know that axioms like commutativity and associativity hold because U is a subset of V. Thus, you don't need to check all of the axioms when proving that U is a subspace.

In general, to prove that $U \subseteq V$ is a subspace, you need to check that $\vec{0} \in U$, U contains inverses, and U is closed under operations. You should convince yourself that, as mentioned in the previous remark, the other axioms follow from the fact that $U \subseteq V$.

Remark 1.2.9. Every vector space V has two "canonical" subspaces: $\{\vec{0}_V\}$ and V itself.

Example 1.2.10. Let $V = \mathbb{R}$. Is $U = \mathbb{N}$ a subspace of V? No, since $1 \in U$ does not have an additive inverse.

1.3 Linear Combinations and Span

It's important for vector spaces to have some notion of relationships between vectors. Generally, this is captured via the idea of linear combinations.

Definition 1.3.1. Given a list of vectors $L = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ for some vector space V over some field $\mathbb{F}$. A linear combination of L is defined as $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_n\vec{v}_n$ where $\alpha_i \in \mathbb{F}$

Example 1.3.2. Consider $L = \left(\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right)$ over the field $\mathbb{R}$. Can $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ be written as a linear combination of L?

Proof. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$. Consider $\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. We can see that this equals $\begin{pmatrix} 2\alpha_1 + \alpha_2 \\ \alpha_1 + \alpha_2 \\ 0 \end{pmatrix}$. Now you set up a system of equations with $\begin{pmatrix} 2\alpha_1 + \alpha_2 \\ \alpha_1 + \alpha_2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ component wise. This gives the following equations:

$$\begin{aligned} 2\alpha_1 + \alpha_2 &= 0 \\ \alpha_1 + \alpha_2 &= 0 \end{aligned}$$

Solving this system yields that $0 = 1$ which is impossible so we know that $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ cannot be a linear combination of L. $\square$

We also want to consider the idea of all possible linear combinations since this tells us, in some sense, all vectors related to the vectors in our list.

Definition 1.3.3. Let L be a list of vectors. We define the span of L to be the set of all vectors that can be represented as a linear combination of L. We denote this as *span*(L). We define the span of the empty list to be $\{\vec{0}\}$.

Theorem 1.3.4. Let $L = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$. Then we have that *span*$(L) = \mathbb{R}^2$ over $\mathbb{R}$

Proof. Note that we are trying to prove equality between two sets namely *span*(L) and $\mathbb{R}^2$. Thus, we will use double containment. First, we will show that $\mathbb{R}^2 \subseteq$ *span*(L). Given an arbitrary vector $(x, y) \in \mathbb{R}^2$. Let $a_1, a_2 \in \mathbb{R}$. Consider $a_1(1, 0) + a_2(0, 1)$ which equals (a_1, a_2). This gives $(a_1, a_2) = (x, y)$. Thus, taking $a_1 = x$ and $a_2 = y$ will show that $(x, y) \in$ *span*(L). Next, we want to show $span(L) \subseteq \mathbb{R}^2$. We know this is true by Closure Under Scalar Multiplication and Vector Addition. $\square$

1.4 Exercises 1

Exercise 1.4.1. Let $\vec{v} \in V$ for some vector space V for a field $\mathbb{F}$. Prove that $-1 \cdot \vec{v} = -\vec{v}$

Proof. We will begin using the fact that $0 \cdot \vec{v} = \vec{0}$

$$\begin{aligned} 0 \cdot \vec{v} &= \vec{0} \\ (-1+1) \cdot \vec{v} &= \vec{0} \\ -1 \cdot \vec{v} + 1 \cdot \vec{v} &= \vec{0} && \text{(by Vector over Scalar Distributivity)} \\ -1 \cdot \vec{v} + \vec{v} &= \vec{0} && \text{(by Identity Multiplication)} \\ -1 \cdot \vec{v} &= -\vec{v} && \text{(by Additive Inverse)} \end{aligned}$$

□

Exercise 1.4.2. Let V be a vector space over a field $\mathbb{F}$. Prove that for any $\vec{v} \in V$ and $c \in \mathbb{F}$, if $c\vec{v} = \vec{0}$ then either $c = 0$ or $\vec{v} = \vec{0}$.

Proof. We will assume that $c\vec{v} = \vec{0}$ and prove that if $c \neq 0$ then we must have $\vec{v} = \vec{0}$. Otherwise, we have that $c = 0$ which satisfies our claim. Since $c \in \mathbb{F}$ and $c \neq 0$, we know that c must have a multiplicative inverse. Denote c's multiplicative inverse as c^{-1}.

$$\begin{aligned} c\vec{v} &= \vec{0} \\ c^{-1}c\vec{v} &= c^{-1}\vec{0} \\ \vec{v} &= \vec{0} && (1) \end{aligned}$$

We note that there is no vector space axiom telling us that a scalar times the zero vector is the zero vector so we need a proof to justify (1).

$$\begin{aligned} c\vec{0} &= c\vec{0} \\ c(\vec{0}+\vec{0}) &= c\vec{0} && \text{(Additive Identity)} \\ c\vec{0} + c\vec{0} &= c\vec{0} && \text{(Scalar Over Vector Distributivity)} \\ c\vec{0} &= \vec{0} && \text{(Additive Inverse)} \end{aligned}$$

□

Exercise 1.4.3. Let $L = ((1,-1),(-\sqrt{2},\sqrt{2}))$ from $\mathbb{R}^2$ over $\mathbb{R}$. Find, with proof, $span(L)$

Proof. First, one may note that both vectors are in the form $(a,-a)$. This should hopefully give you an inkling as to what the span may be. We start by letting $a_1, a_2 \in \mathbb{R}$. Consider $a_1(1,-1) + a_2(-\sqrt{2},\sqrt{2})$. Computing the sum of these two vectors we get $(a_1 - a_2\sqrt{2}, -a_1 + a_2\sqrt{2})$. By letting

$b = a_1 - a_2\sqrt{2}$, we can see that all vectors in the span of L have the form $(b, -b)$. We now make the claim that $span(L) = S = \{(x, -x) \in \mathbb{R}^2\}$. We recall that since we are showing set equality we must use double containment and the above shows that $span(L) \subseteq S$. We now want to show that $S \subseteq span(L)$. Consider some arbitrary vector $(x, -x) \in S$. We want to show that this is a linear combination of L. We note that $(x, -x) = x(1, -1)$ thus we have what we want. So by double containment, $span(L) = S$ □

Exercise 1.4.4. Let $L = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ be a list of vectors from a vector space V over a field $\mathbb{F}$. Show that $span(L)$ is a vector space.

Proof. We first observe that since all the vectors are already in V which is a vector space, they must follow all the axioms like commutativity, associativity, and distributivity. It remains to check axioms like additive identity, additive inverse, and closure. For the additive inverse, for any vector in $span(L)$, we simply consider the negative of all the coefficients in the linear combination to get the additive inverse. For additive identity, we observe that $\vec{0} \in span(L)$ for any list L. This is because we can consider the linear combination where all coefficients are zero.

It remains to check closure. Clearly, closure under scalar multiplication holds since all it does it scale the coefficients in the linear combination which is still a linear combination. Closure over vector addition also holds since we are simply adding the coefficients of the linear combination of the vectors we are adding. □

1.5 Linear Independence

We want an idea that tells us in a sense how redundant our list. This is describing a list which contains multiple vectors that contribute the same vectors to the span.

Definition 1.5.1. A list L of vectors is linearly independent if for each $\vec{w} \in span(L)$, $\vec{w}$ is a unique linear combination of L. Lists that do not have this property are called linearly dependent.

Example 1.5.2. Let $L = ((1,0),(2,0))$. We observe that $(2,0) = 2*(1,0) = 1*(2,0)$ which are two different linear combination of L. Thus, L cannot be linearly independent.

The idea in this example is that $(1,0)$ and $(2,0)$ contribute the same vectors to the span. Consider $\vec{v} = a_1(1,0) + a_2(2,0)$. We can simply rewrite this as $\vec{v} = a_1(1,0) + a_2(2,0) = a_1(1,0) + 2a_2(1,0) = (a_1 + 2a_2)(1,0)$. This shows us that including the vector $(2,0)$ in the list does not change the span at all.

Example 1.5.3. Let $L = ((1,0,0),(0,1,0),(1,1,0))$. Consider the vector $(2,2,0)$. We observe that $(2,2,0) = 2*(1,0,0) + 2*(0,1,0)$ and that $(2,2,0) = 2*(1,1,0)$. Since we have two different linear combination that produce $(2,2,0) \in span(L)$, this list is not linearly independent.

Similarly, we have here that $(1,1,0) = (1,0,0) + (0,1,0)$ so having $(1,1,0)$ in the list does not affect the span at all. It should be noted though that having the list $((1,0,0),(0,1,0))$ and the list $((1,1,0))$ would not have the same span despite the fact that combining them produces a linearly dependent list.

We will now consider some definitions that are equivalent to our definition of linearly independent but often much more useful for proving linear independence/dependence.

Theorem 1.5.4. Let $L = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ be a list of vectors. The following are equivalent:

1. L is linearly independent
2. $\vec{0}$ has only the trivial representation
3. No $\vec{v}_j \in L$ can be written as a linear combination of $L \setminus \{\vec{v}_j\}$

Remark 1.5.5. The following are equivalent means that if you have any one of the statements then you immediately have all of the statements.

Remark 1.5.6. The trivial representation means that all the coefficients in the linear combination are zero.

Proof. We will first prove that (1) and (2) imply each other.

We want to show (1) implies (2). We assume that L is linearly independent. Note that $\vec{0} \in span(L)$ and that setting all coefficients to zero gives a linear combination of L which equals $\vec{0}$.

Since the linear combinations of L are unique and we have found one linear combination of L that equals $\vec{0}$, it must be the only linear combination that equals $\vec{0}$.

We now want to show (2) implies (1). We assume that $\vec{0}$ only has the trivial representation. Let $\vec{w} \in span(L)$. Suppose that we have $\vec{w} = a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_n\vec{v}_n$ and $\vec{w} = b_1\vec{v}_1 + b_2\vec{v}_2 + \ldots + b_n\vec{v}_n$.

$$\begin{aligned} \vec{w} - \vec{w} &= \vec{0} \\ a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_n\vec{v}_n - (b_1\vec{v}_1 + b_2\vec{v}_2 + \ldots + b_n\vec{v}_n) &= \vec{0} \\ (a_1 - b_1)\vec{v}_1 + (a_2 - b_2)\vec{v}_2 + \ldots + (a_n - b_n)\vec{v}_n &= \vec{0} \end{aligned}$$

Now we use our assumption which tells that $\vec{0}$ only has the trivial representation. This means that $a_i - b_i = 0$ for all $i \in \{1, \ldots, n\}$. Thus, $a_i = b_i$ for all $i \in \{1, \ldots, n\}$ and so the linear combinations must be unique.

We will now show that (2) and (3) imply each other.

First, we will show that (2) implies (3) via the contrapositive. So we will assume that (3) does not hold. This means that there exists a $\vec{v}_j \in L$ that can be written as a linear combination of $L \setminus \{\vec{v}_j\}$. We will say that $\vec{v}_j = \sum_{i=1, i\neq j}^{n} a_i\vec{v}_i$.

$$\begin{aligned} \vec{v}_j &= \sum_{i=1, i\neq j}^{n} a_i\vec{v}_i \\ \vec{0} &= \sum_{i=1, i\neq j}^{n} a_i\vec{v}_i - \vec{v}_j \end{aligned}$$

We observe that the above equation represents $\vec{0}$ as a linear combination of L. Importantly, the coefficient of $\vec{v}_j = -1 \neq 0$ thus $\vec{0}$ can be written as not the trivial representation.

Next, we will show that (3) implies (2) again via the contrapositive. Assume that $\vec{0}$ has a nontrivial representation. This means that we have $\vec{0} = \sum_{i=1}^{n} a_i\vec{v}_i$ such that there exists an $a_j \neq 0$.

$$\begin{aligned} \vec{0} &= a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_n\vec{v}_n \\ -a_j\vec{v}_j &= a_1\vec{v}_1 + \ldots + a_{j-1}\vec{v}_{j-1} + a_{j+1}\vec{v}_{j+1} + \ldots + a_n\vec{v}_n \\ \vec{v}_j &= \frac{a_1}{a_j}\vec{v}_1 + \ldots + \frac{a_{j-1}}{a_j}\vec{v}_{j-1} + \frac{a_{j+1}}{a_j}\vec{v}_{j+1} + \ldots + \frac{a_n}{a_j}\vec{v}_n \qquad \text{(we can divide since } a_j \neq 0) \end{aligned}$$

The above equation shows that $\vec{v}_j$ can be written as a linear combination of $L \setminus \{\vec{v}_j\}$. Thus, we have shown that (3) does not hold.

The proof that (1) and (3) imply each other is left as an exercise. □

Remark 1.5.7. To prove linear independence, it suffices to show any of the three conditions. Often the easiest to prove is (2). Since you consider the equation $\vec{0} = \sum_{i=1}^{n} a_i\vec{v}_i$ and solve the system given by each coordinate to see if there are nontrivial solutions. (3) is often most useful for proving linear dependence.

Corollary 1.5.8. A result of (2) is this slightly stronger statement about linear dependence. If a list L is linearly dependent, then there exists $\vec{v}_j \in L$ such that $\vec{v}_j$ can be written as a linear combination of $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{j-1})$. In other words, there exists a vector $\vec{v}_j \in L$ which can be written as a linear combination of the vectors that come before it in the list.

Proof. We know that there exists a nontrivial linear combination that equals $\vec{0}$. Let this be $\sum_{i=1}^{n} a_i\vec{v}_i$. Since it is nontrivial, let a_j be the last nonzero coefficient.

$$\begin{aligned} \vec{0} &= \sum_{i=1}^{n} a_i\vec{v}_i \\ \vec{0} &= \sum_{i=1}^{j} a_i\vec{v}_i && \text{(since } a_j \text{ is the last nonzero coefficient)} \\ -a_j\vec{v}_j &= \sum_{i=1}^{j-1} a_i\vec{v}_i \\ \vec{v}_j &= \sum_{i=1}^{j-1} -\frac{a_i}{a_j}\vec{v}_j \end{aligned}$$

□

Theorem 1.5.9. For any list of vectors L, there exists a sublist L' such that L' is linearly independent and $span(L) = span(L')$.

Proof. Let n be the size of the list L. We will prove this statement via induction on n.

Base Case: When $n = 0$, our list is empty so our statement is vacuously true.

Induction Step: Let $n \in \mathbb{N}$. Assume that the statement holds for lists of size n. We want to show that the statement holds for all lists of size $n+1$. Let $L = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n+1})$. If L is linearly independent, we're done. Now assume that L linearly dependent. We will use (3) to get that $\vec{v}_j = \sum_{i=1, i\neq j}^{n+1} a_i\vec{v}_i$. Consider the list $L' = L \setminus \{\vec{v}_j\}$. Note that $|L'| = n$ so we can apply the Induction Hypothesis. This gives a list L'' which is linearly independent and $span(L') = span(L'')$.

We want to show that the $span(L) = span(L') = span(L'')$. Let $\vec{w} \in span(L)$. By definition, we know that $\vec{w} = \sum_{i=1}^{n+1} b_i\vec{v}_i$. We use the fact that $\vec{v}_j = \sum_{i=1, i\neq j}^{n+1} a_i\vec{v}_i$ to get that:

$$\vec{w} = \sum_{i=1}^{n+1} b_i\vec{v}_i = \sum_{i=1, i\neq j}^{n+1} b_i\vec{v}_i + b_j\vec{v}_j = \sum_{i=1, i\neq j}^{n+1} b_i\vec{v}_i + b_j \sum_{i=1, i\neq j}^{n+1} a_i\vec{v}_i$$

We have shown that $\vec{w}$ can be written as a linear combination of $L \setminus \{\vec{v}_j\}$ so $\vec{w} \in span(L')$ and $span(L) \subseteq span(L')$. We know that $span(L') \subseteq span(L)$ because L' is a sublist of L. So by double containment $span(L) = span(L')$ and thus $span(L) = span(L'')$ so we have found a sublist L'' which satisfies the desired conditions. □

1.6 Basis and Dimension

Since there are so many possible vector spaces, we want to introduce a concept to ground the idea of a vector space.

Definition 1.6.1. Let V be a vector space. A basis of V is a list L which satisfies the following two properties:

1. L is linearly independent
2. $span(L) = V$

Example 1.6.2. Let's find a basis for $\mathbb{R}^3$ over $\mathbb{R}$. Consider $L = ((1,0,0),(0,1,0),(0,0,1))$. You may notice that you can generalize this type of basis over $\mathbb{R}^N$

Remark 1.6.3. In general, $\mathbb{R}^N$ has a basis called the elementary basis. This is the basis where each vector has a 1 at exactly one coordinate and 0 everywhere else.

Definition 1.6.4. Let V be a vector space. The dimension of V, denoted $dim(V)$, is the number of vectors in its basis.

There are infinite dimensional vector space, but those are much more complicated than finite-dimensional ones so we will only focus on finite dimensional vector spaces. We now want to prove a very important theorem, or rather lemma, known as the Exchange Lemma. It gives us a relationship between linearly independent lists and spanning lists which is immensely useful for finding properties of bases.

Theorem 1.6.5. Let V be a finite dimensional vector space. All linearly independent sets must have size less than or equal to all spanning sets.

Proof. Let $(\vec{v}_1, \ldots, \vec{v}_n)$ be a linearly independent set of vectors and let $(\vec{w}_1, \ldots, \vec{w}_m)$ be a spanning set of vectors. We want to show that $n \le m$.

Step 1: Consider the list $(\vec{v}_1, \vec{w}_1, \ldots, \vec{w}_m)$. We know that this list is linearly dependent because $(\vec{w}_1, \ldots, \vec{w}_m)$ is a spanning list so $\vec{v}_1$ can be written as a linear combination of the other vectors. We will use the fact that now there must exist a vector in the list that can be written as a linear combination of the vectors which come before it in the list. We know that this vector cannot be $\vec{v}_1$ since $span(\emptyset) = \vec{0}$ and $\vec{v}_1 \neq \vec{0}$ since $(\vec{v}_1, \ldots, \vec{v}_n)$ is linearly independent. This means that there exists some $\vec{w}_j$ that can be written as a linear combination of $(\vec{v}_1, \vec{w}_1, \ldots, \vec{w}_{j-1})$. Remove $\vec{w}_j$ from the list so we are now left with $(\vec{v}_1, \vec{w}_1, \ldots, \vec{w}_{j-1}, \vec{w}_{j+1}, \ldots, \vec{w}_m)$. We note that this new list still spans V since any linear combination of $(\vec{w}_1, \ldots, \vec{w}_m)$ we can simply replace the $\vec{w}_j$ with the linear combination of $(\vec{v}_1, \vec{w}_1, \ldots, \vec{w}_{j-1})$ which makes $\vec{w}_j$.

Step k: Consider the list $(\vec{v}_1, \ldots, \vec{v}_k, \ldots)$ which is the first k $\vec{v}$'s followed by the remaining $\vec{w}$'s. We know that this list is linearly dependent by the same logic as Step 1. We also know that none of the $\vec{v}$'s can be written as a linear combination of what comes before them since $(\vec{v}_1, \ldots, \vec{v}_n)$ is linearly independent. Thus, there exists some $\vec{w}_j$ that can be written as a linear combination of what comes before it in the list. Remove this $\vec{w}_j$ from the list. The span of the list is still V by, again, the same logic as Step 1.

We've shown that we can slowly replace the $\vec{w}$'s with $\vec{v}$'s and the list must always be spanning. Now, we skip forward to the final step where we add $\vec{v}_n$. We claim that there still remains at least one $\vec{w}$ in the list. We know that this list must be linearly dependent since the list $(\vec{v}_1, \ldots, \vec{v}_{n-1}, \ldots)$ without $\vec{v}_n$ is spanning. If there was not a $\vec{w}$ in the list, we would have that $\vec{v}_n$ can be written as a linear combination of $(\vec{v}_1, \ldots, \vec{v}_{n-1})$ which contradicts the linear independence of $(\vec{v}_1, \ldots, \vec{v}_n)$. Thus, we have shown that when we add the final $\vec{v}_n$ there is at least one $\vec{w}$ still in the list, so $n \leq m$. □

We will now explore how one can apply the Exchange Lemma to get some standard facts about bases and dimension. It should be noted that there is another way to reach these conclusions via matrices and the "Fundamental Theorems of Linear Algebra." However, it seems like there is a lack of intuition without knowing the Exchange Lemma for why these facts are true.

Corollary 1.6.6. Let V be a vector space with $dim(V) = n$. All sets in V with size greater than n must be linearly dependent.

Proof. From Theorem 8.5, we showed that linearly independent sets are at most the size of any spanning set. Since a basis is a spanning set and it has size n, all linearly independent lists must be at most size n. So any list larger than size n must be linearly dependent. □

There exists a similar statement to Corollary 8.6 regarding span which is left as an exercise. You should realize that these statements are immensely powerful. You need not know a single thing about any of the vectors than the list other than there are "too many" or "too few" vectors in a list and you immediately know facts about span and linear independence.

Theorem 1.6.7. Let V be a vector space such that $dim(V) = n$. Every basis of V has the same size.

Proof. Let B_1, B_2 be bases of V. We know that B_1 is linearly independent and B_2 is spanning. This means that $|B_1| \leq |B_2|$. We also know that B_2 is linearly independent and B_1 is spanning. This means that $|B_2| \leq |B_1|$. Thus, we must have that $|B_1| = |B_2|$. □

We have had all this discussion about bases but have yet to discuss how to actually find them. We will now show that you can always find a basis starting from any arbitrary list. Note that this is only theory and in most practical applications the elementary basis will suffice.

Theorem 1.6.8. Let V be a vector space with $dim(V) = n$. Let L be a linearly independent list of vectors in V. You can always extend L into a basis of V.

Proof. Let $L = (\vec{v}_1, \ldots, \vec{v}_k)$ be a linearly independent list. If it already spans V, then we have a basis are we are done. Assume $span(L) \neq V$. Since $span(L) \neq V$, there exists some $\vec{v}_{k+1} \in V$ and $\vec{v}_{k+1} \notin span(L)$. Consider the list $(\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1})$. We want to show that this list is linearly independent. Consider a linear combination of the above list which equals $\vec{0}$.

$$a_1\vec{v}_1 + \ldots + a_k\vec{v}_k + a_{k+1}\vec{v}_{k+1} = \vec{0}$$

We split into two cases. The first case is when $a_{k+1} = 0$. This would reduce our equation to $a_1\vec{v}_1 + \ldots + a_k\vec{v}_k = \vec{0}$ which by linear independence of L has only the trivial combination. So if $a_{k+1} = 0$ then all the coefficients must be 0.

The second case is when $a_{k+1} \neq 0$. We want to show this case is impossible.

$$\begin{aligned} a_1\vec{v}_1 + \ldots + a_k\vec{v}_k + a_{k+1}\vec{v}_{k+1} = \vec{0} \\ a_1\vec{v}_1 + \ldots + a_k\vec{v}_k = -a_{k+1}\vec{v}_{k+1} \\ \frac{a_1}{-a_{k+1}}\vec{v}_1 + \ldots + \frac{a_k}{-a_{k+1}}\vec{v}_k = \vec{v}_{k+1} \end{aligned}$$

This is a contradiction since we have shown that $\vec{v}_{k+1} \in span(L)$. Thus, the second case is not possible and we must have $a_{k+1} = 0$ which we showed implies all the coefficients are zero. This proves linear independence.

You can repeat this process until the $span(L) = V$. We've shown that at every step the new list is still linearly independent. Now we must show that it stops when $k = n$ and that the $span(L) = V$ at that point. If the process did not stop at $k = n$, then we would add another vector and the list would still be linearly dependent. However, this is impossible since we would then have a list of $n + 1$ linearly independent vectors. The only way for the process to stop is if there are no more vectors in V that are not in $span(L)$. Thus, the process stops when $k = n$ and $span(L) = V$. □

This theorem is also immensely powerful. It tells you that any linearly independent list is a component of a basis. In fact, it is a component of infinitely many bases and you can build a basis by slowly adding vectors that are not in the span of your list so far. We leave the theorem regarding shrinking spanning lists as an exercise.

Next, we will explore what this notion of dimension tells us outside of the size of a basis. This can make checking if a list is spanning/linearly independent easier as if its size is too small/too big you won't have to do any work.

Corollary 1.6.9. Let V be a vector space with $dim(V) = n$. Any linearly independent list of size n must be a basis.

Proof. If the list is not a basis, we know that we can extend it to a basis. However, if we add any vectors to the list, it becomes linearly dependent since its size is greater than the dimension. Thus, the list must already be a basis. □

Corollary 1.6.10. Let V be a vector space with $dim(V) = n$. Any spanning list of size n must be a basis.

Proof. If the list is not a basis, we know that we can reduce it to a basis. However, if we remove any vectors from the list, it is no longer spanning since its size is less than the dimension. Thus, the list must already be a basis. $\square$

We will now go over a notation that will become much more useful when we discuss linear transformations and matrices.

Definition 1.6.11. Let V be a finite dimensional vector space and $B = (\vec{v}_1, \ldots, \vec{v}_n)$ be a basis of V. The coordinate vector of a vector $\vec{v} \in V$ is $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_B$ where $\vec{v} = \sum_{i=1}^{n} a_i \vec{v}_i$. We denote the coordinate vector of $\vec{v}$ with respect to a basis B as $[\vec{v}]_B$.

Remark 1.6.12. The coordinate vector essentially describes the vector $\vec{v}$ as a vector of its coefficients with respect to a basis B.

Remark 1.6.13. Consider $(1, 2, 3) \in \mathbb{R}^3$. This is actually also a coordinate vector with respect to the elementary basis. The way we write vectors that we are used to is as the coordinate vectors of the elementary basis. So if no subscript basis is written, you are safe to assume that it is using the elementary basis.

Coordinate vectors possess certain properties that you will recognize when we discuss linear transformations in the next section. This makes them very useful representations when you're moving between vector spaces and different bases.

Theorem 1.6.14. Let V be an $\mathbb{F}$-vector space and B be a basis of V. The following two properties hold:

1. $\forall \vec{v}, \vec{w} \in V$, $[\vec{v}]_B + [\vec{w}]_B = [\vec{v} + \vec{w}]_B$
2. $\forall \vec{v} \in V, \forall \alpha \in \mathbb{F}, \alpha[\vec{v}]_B = [\alpha\vec{v}]_B$

Proof. First, we will prove (1). Let $B = (\vec{u}_1, \ldots, \vec{u}_n)$ and let $[\vec{v}]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_B$ and $[\vec{w}]_B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_B$.

Clearly, $[\vec{v}]_B + [\vec{w}]_B = \begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ \vdots \\ a_n+b_n \end{bmatrix}_B$. We know $\vec{v} = \sum_{i=1}^{n} a_i \vec{u}_i$ and $\vec{w} = \sum_{i=1}^{n} b_i \vec{u}_i$. So $\vec{v} + \vec{w} = \sum_{i=1}^{n} a_i \vec{u}_i + \sum_{i=1}^{n} b_i \vec{u}_i = \sum_{i=1}^{n} (a_i + b_i)\vec{u}_i$. Thus, $[\vec{v} + \vec{w}]_B = \begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ \vdots \\ a_n+b_n \end{bmatrix}_B$.

Next, we show (2). Let $B = (\vec{u}_1, \ldots, \vec{u}_n)$ and $[\vec{v}]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_B$. We can see that $\alpha[\vec{v}]_B = \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_n \end{bmatrix}_B$. We know that $\alpha[\vec{v}]_B = \sum_{i=1}^{n} \alpha a_i \vec{u}_i = \sum_{i=1}^{n} (\alpha a_i)\vec{u}_i = [\alpha \vec{v}]_B$. $\square$

1.7 Exercises 2

Exercise 1.7.1. Prove that (1) and (3) imply each other in Theorem 5.4.

Proof. For (1) implies (3), we observe that $\vec{v}_j \in L$ means that $\vec{v}_j \in span(L)$. This means that it has a unique linear combination and since $1\vec{v}_j$ is a linear combination is must be the only one. Thus, $L \setminus \{\vec{v}_j\}$ cannot form $\vec{v}_j$ as a linear combination.

We will prove that (3) implies (1) via the contrapositive. We assume that L is not linearly independent. This means that there exists a vector $\vec{w} \in span(L)$ such that $\vec{w}$ can be represented as at least two different linear combinations. Let these two linear combinations be $\vec{w} = \sum_{i=1}^{n} a_i\vec{v}_i$ and $\vec{w} = \sum_{i=1}^{n} b_i\vec{v}_i$.

$$\begin{aligned} \vec{w} - \vec{w} &= \vec{0} \\ \sum_{i=1}^{n} a_i\vec{v}_i - \sum_{i=1}^{n} b_i\vec{v}_i &= \vec{0} \\ \sum_{i=1}^{n}(a_i - b_i)\vec{v}_i &= \vec{0} \end{aligned}$$

Technically it suffices to stop here. Since we have shown that not (1) implies not (2) which we have already proved to imply not (3), but we will complete the direct proof anyway. Since the two linear combinations are different, we know that there exists some j such that $a_j \neq b_j$.

$$\begin{aligned} \sum_{i=1}^{n}(a_i - b_i)\vec{v}_i &= \vec{0} \\ \sum_{i=1, i\neq j}^{n}(a_i - b_i)\vec{v}_i &= -(a_j - b_j)\vec{v}_j \\ \sum_{i=1, i\neq j}^{n} \frac{a_i - b_i}{b_j - a_j}\vec{v}_i &= \vec{v}_j \end{aligned}$$

Thus, we have shown that $\vec{v}_j$ can be written as a linear combination of the other vectors. □

Exercise 1.7.2. Prove that in the vector space $\mathbb{R}$ over $\mathbb{R}$, any list of two vectors is linearly dependent.

Proof. Let $\vec{v}, \vec{u} \in \mathbb{R}$. We note that if either of the vectors are $\vec{0}$ then we immediately have that they are linearly dependent since you can multiply the other vector by zero to get $\vec{0}$ which violates (3). Now we assume that neither of the vectors are $\vec{0}$. Let $|\vec{v}|$ denote the real number contained in $\vec{v}$. We note that $\vec{u} = \frac{|\vec{u}|}{|\vec{v}|}\vec{v}$. Thus, we have shown that $\vec{u}$ can be written as a linear combination of $\vec{v}$ which violates (3). Thus, the list is linearly dependent. □

Exercise 1.7.3. In the vector space $\mathbb{R}$ over $\mathbb{Q}$, find a list of two vectors that are linearly independent.

Proof. Since we are over $\mathbb{Q}$, we will want to consider an irrational number in our list. Consider the list $L = (\vec{1}, \vec{\sqrt{2}})$. We will show that (3) holds. Clearly, we cannot write $\vec{1}$ as a linear combination of $\vec{\sqrt{2}}$ since $1 = \frac{1}{\sqrt{2}}\sqrt{2}$ and $\frac{1}{\sqrt{2}}$ is irrational so it is not in our field. Similarly, we cannot write $\vec{\sqrt{2}}$ as a linear combination of $\vec{1}$ since $\sqrt{2} = \sqrt{2} \cdot 1$ but $\sqrt{2}$ is also irrational. Thus, since (3) holds this list must be linearly independent. $\square$

Exercise 1.7.4. Is the list $L = \left(\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}7\\8\\9\end{pmatrix}\right)$ linearly independent or dependent?

Proof. We note that $2 * \begin{pmatrix}4\\5\\6\end{pmatrix} - \begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}7\\8\\9\end{pmatrix}$. This violates (3). Thus, this list is linearly dependent.

$\square$

Exercise 1.7.5. Let V be a finite dimensional vector space and L be a list that spans V. Prove that you can always reduce L into a basis of V.

Proof. If L is linearly independent, by definition you already have a basis. If L is linearly dependent, there exists a vector in the list that can be written as a linear combination of the other vectors in the list. Remove this vector. Notice that removing this vector does not change the span of the list since you can simply replace the vector with its linear combination. Continue this process until the list is linearly independent. The span will remain the same at each step and now your list is linearly independent thus you have a basis. $\square$

Exercise 1.7.6. Let V be a vector space with $dim(V) = n$. All sets that span V must have size greater than or equal to n.

Proof. Assume, for sake of contradiction, you have a spanning set of size strictly less than n. Since the $dim(V) = n$, consider any basis B which by definition is linearly independent and has size n. This means we have found a linearly independent list that has size greater than a spanning set which is a contradiction. $\square$

Chapter 2

Linear Transformations and Matrices

2.1 Linear Transformations

Definition 2.1.1. A linear transformation T from a $\mathbb{F}$-vector space V to a vector space W is a function with domain V and codomain W that satisfies the following two properties:

1. $\forall \vec{v}, \vec{u} \in V$, $T(\vec{v}) + T(\vec{u}) = T(\vec{v} + \vec{u})$
2. $\forall \vec{v} \in V, \forall \alpha \in \mathbb{F}, \alpha T(\vec{v}) = T(\alpha \vec{v})$

It suffices to prove that $T(\sum a_i \vec{v}_i) = \sum a_i T(\vec{v}_i)$ for all choices $a_i \in \mathbb{F}$ and $\vec{v}_i \in V$. This is simply a combination of properties (1) and (2) plus a little induction.

Remark 2.1.2. There are a lot of different words often used interchangeably with "transformation". You may hear terms such as "map", "operator", or simply "function". No one will misunderstand you if you use any one of these even though they technically have nuanced meanings. For example, "operator" is used to refer to linear transformations which have the same domain and codomain.

These two properties of linear transformations make them much easier to study than arbitrary transformations, but they are still just as useful. You've probably actually seen and used a lot of linear transformations just haven't realized that they were linear transformations.

Example 2.1.3. Let V be an $\mathbb{F}$-vector space. Define the transformation $T : V \to V$ via $T(\vec{v}) = \vec{v}$. This is known as the identity transformation.

Example 2.1.4. Let V be an $\mathbb{F}$-vector space. Define the transformation $T : V \to V$ via $T(\vec{v}) = \vec{0}$. This is known as the zero transformation.

The word "linear" in linear transformation can be a bit misleading. We are used to thinking of "linear" as meaning a straight line in the xy-plane. However, there are many of these straight lines which are not linear transformations.

Example 2.1.5. Consider the vector space $\mathbb{R}$ over $\mathbb{R}$. Define the transformation $T : \mathbb{R} \to \mathbb{R}$ via $T(x) = x+1$. This is not a linear transformation. Consider $T(x)+T(y) = x+1+y+1 = x+y+2 \neq x+y+1 = T(x+y)$.

There are also many linear transformations which are not obviously linear but you are likely very familiar with.

Example 2.1.6. Consider the vector space $\mathbb{P}_n$ over $\mathbb{R}$. This is the set of all polynomials with real coefficients that have degree less than or equal to n. Define the transformation $T : \mathbb{P}_n \to \mathbb{P}_{n-1}$ via $T(p) = p'$ where p' is the derivative of p. We know that for $p, q \in \mathbb{P}_n$ we have that $p' + q' = (p+q)'$ and $(\alpha p)' = \alpha p'$. These are simply facts from Calculus. Thus, this is a linear transformation.

Example 2.1.7. Consider the vector space $\mathbb{R}^2$ over $\mathbb{R}$. Define the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ via $T(x,y) = (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta))$. This is a counterclockwise rotation about the origin by an angle θ. This is also a linear transformation.

One of the conditions for a defined function is that is must be defined for the whole domain. This is still true of a linear transformation, however, linear transformations have a special property. Instead of needing to know what the function does to every single value, for example $f(x,y) = (x^2+y, x-y), \forall x, y \in \mathbb{R}$, we only need to know what the linear transformation does to a basis of the vector space. The following theorem shows why.

Theorem 2.1.8. Let V and W be $\mathbb{F}$-vector spaces. Let $(\vec{v}_1, \ldots, \vec{v}_n)$ be a basis for V and let $(\vec{w}_1, \ldots, \vec{w}_n)$ be an arbitrary list of vectors in W. There exists a unique linear transformation $T : V \to W$ such that $T(\vec{v}_i) = \vec{w}_i$ for each $i \in \{1, \ldots, n\}$.

Proof. We are going to prove this by constructing a linear transformation which satisfies the properties and then prove its unique.

First, we will prove the existence of a such a transformation and prove its linear. Since $(\vec{v}_1, \ldots, \vec{v}_n)$ is a basis, for all $\vec{v} \in V$, $\vec{v}$ can be written uniquely as a linear combination of $(\vec{v}_1, \ldots, \vec{v}_n)$. Define the transformation $T : V \to W$ via $T(\vec{v}) = \sum_{i=1}^n a_i\vec{w}_i$ where $\vec{v} = \sum_{i=1}^n a_i\vec{v}_i$. This is a well-defined function because of the fact that $(\vec{v}_1, \ldots, \vec{v}_n)$ is a basis. We want to show that $T(\vec{v}_i) = \vec{w}_i$ for each $i \in \{1, \ldots, n\}$ and that T is linear. Consider $T(\vec{v}_1)$. Since $\vec{v}_1 = 1 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2 + \ldots + 0 \cdot \vec{v}_n$, we know that $a_1 = 1$ and $a_j = 0$ for all $j \neq 1$. Thus, we have that $T(\vec{v}_1) = \sum_{i=1}^n a_i\vec{w}_i = a_1\vec{w}_1 = \vec{w}_1$. You can repeat this for all $\vec{v}_i$, so T satisfies $T(\vec{v}_i) = \vec{w}_i$. Now we want to show that T is linear. Let $\vec{v}, \vec{u} \in V$ such that $\vec{v} = \sum_{i=1}^n a_i\vec{v}_i$ and $\vec{u} = \sum_{i=1}^n b_i\vec{v}_i$.

$$T(\vec{v}) + T(\vec{u}) = \sum_{i=1}^n a_i\vec{w}_i + \sum_{i=1}^n b_i\vec{w}_i = \sum_{i=1}^n (a_i + b_i)\vec{w}_i = T(\vec{v} + \vec{u})$$
$$\alpha T(\vec{v}) = \alpha \sum_{i=1}^n a_i\vec{w}_i = \sum_{i=1}^n \alpha a_i\vec{w}_i = T(\alpha\vec{v})$$

Now we want to show uniqueness of T. Let $S : V \to W$ be a linear transformation that satisfies $S(\vec{v}_i) = \vec{w}_i$. We will show that $S = T$. Let $\vec{v} \in V$ such that $\vec{v} = \sum_{i=1}^{n} a_i \vec{v}_i$.

$$S(\vec{v}) = S(\sum_{i=1}^{n} a_i \vec{v}_i) = \sum_{i=1}^{n} a_i S(\vec{v}_i) = \sum_{i=1}^{n} a_i \vec{w}_i = \sum_{i=1}^{n} a_i T(\vec{v}_i) = T(\sum_{i=1}^{n} a_i \vec{v}_i) = T(\vec{v})$$

□

Remark 2.1.9. Let $\vec{v} \in V$ for some $\mathbb{F}$-vector space V and let $(\vec{v}_1, \dots, \vec{v}_n)$ be a basis of V. We know that $\vec{v}$ can be written as a linear combination of the basis. Let this be $\vec{v} = \sum_{i=1}^{n} a_i \vec{v}_i$. Consider a linear transformation $T : V \to W$ for some $\mathbb{F}$-vector space W. We want to show that it suffices to know how a linear transformation acts on the basis vectors to know how it acts on any vector.

$$\begin{aligned}
\vec{v} &= \sum_{i=1}^{n} a_i \vec{v}_i \\
T(\vec{v}) &= T(\sum_{i=1}^{n} a_i \vec{v}_i) \\
T(\vec{v}) &= \sum_{i=1}^{n} T(a_i \vec{v}_i) && \text{(use property (1))} \\
T(\vec{v}) &= \sum_{i=1}^{n} a_i T(\vec{v}_i) && \text{(use property (2))}
\end{aligned}$$

We use the properties of a linear transformation to show that $T(\vec{v})$ is actually a linear combination of the T applied to each of the basis vectors. Additionally, it has the exact same coefficients as $\vec{v}$ had in terms of the original basis. This shows that T applied to any vector in V is a linear combination of T applied to the basis vectors. This implies that it suffices to know only what T does to a set of basis vectors.

Theorem 2.1.10. Let V and W be $\mathbb{F}$-vector spaces. The set of linear transformations from $V \to W$ is a vector space over $\mathbb{F}$.

Proof. We need to define a notion of vector addition and scalar multiplication in our vector space of linear transformations. Let E be our vector space of linear transformations. Let $a \in \mathbb{F}$ and $T, S \in E$. Define vector addition via $(T + S)(\vec{x}) = T(\vec{x}) + S(\vec{x})$ for all $\vec{x} \in V$. Define scalar multiplication via $(aT)(\vec{x}) = aT(\vec{x})$ for all $\vec{x} \in V$. Checking the axioms is left as an exercise. □

Since we now understand how linear transformations work, we want to introduce a notion of "undoing" a linear transformation. This is not always possible but it can be useful to know if it is.

Definition 2.1.11. Let V be a $\mathbb{F}$-vector space. We define the identity linear transformation on V as $id : V \to V$ via $id(\vec{x}) = \vec{x}$ for all $\vec{x} \in V$.

Definition 2.1.12. Let V and W be $\mathbb{F}$-vector spaces. Let $T : V \to W$ and $S : W \to V$ be linear transformations.

1. S is a left inverse of T if $S \circ T = id_V$
2. S is a right inverse of T if $T \circ S = id_W$
3. S is a two sided inverse of T if $T \circ S = id_W$ and $S \circ T = id_V$

Theorem 2.1.13. Let $T : V \to W$, $R : W \to V$, $L : W \to V$ be linear transformations such that R is a right inverse of T and L is a left inverse of T. Prove that $R = L$.

Theorem 2.1.14. Let U, V, W be $\mathbb{F}$-vector spaces and $S : U \to V$ and $T : V \to W$ be invertible linear transformations. Then $T \circ S$ is invertible and its inverse is $S^{-1} \circ T^{-1}$.

Proof. We want to show that $S^{-1} \circ T^{-1}$ is a left inverse of $T \circ S$ and a right inverse of $T \circ S$.

$$\begin{aligned}(T \circ S) \circ (S^{-1} \circ T^{-1}) &= T \circ S \circ S^{-1} \circ T^{-1} \\ &= T \circ id_V \circ T^{-1} \\ &= T \circ T^{-1} \\ &= id_W\end{aligned}$$

You can do the same to show that it is a right inverse of $T \circ S$. $\square$

2.2 Exercises 3

Exercise 2.2.1. Let V and W be vector spaces. Prove that for any linear transformation $T : V \to W$ we have that $T(\vec{0}_V) = \vec{0}_W$.

Proof. We use additivity of linear transformations to get that:

$$\begin{aligned} T(\vec{0}_V) + T(\vec{0}_V) &= T(\vec{0}_V + \vec{0}_V) \\ T(\vec{0}_V) + T(\vec{0}_V) &= T(\vec{0}_V) \\ T(\vec{0}_V) &= \vec{0}_W \end{aligned}$$

We subtract $T(\vec{0}_V)$ from both sides to get $\vec{0}_W$ since T has codomain W so its outputs are vectors in W. $\square$

Exercise 2.2.2. Consider the vector space $\mathbb{R}$ over $\mathbb{R}$. Prove that $T : \mathbb{R} \to \mathbb{R}$ via $T(x) = x^2$ is not a linear transformation.

Proof. We will prove that $\alpha T(x) \neq T(\alpha x)$. Clearly, $\alpha T(x) = \alpha x^2$. However, $T(\alpha x) = \alpha^2 x^2$. Since these two are not always equal, this is not a linear transformation. $\square$

Exercise 2.2.3. Let $T : V \to W$, $R : W \to V$, $L : W \to V$ be linear transformations such that R is a right inverse of T and L is a left inverse of T. Prove that $R = L$.

Proof. By definition of left inverse, we know that $L \circ T = id_V$.

$$\begin{aligned} L \circ T &= id_V \\ L \circ T \circ R &= id_V \circ R \\ L \circ id_V &= R \\ L &= R \end{aligned}$$

$\square$

2.3 Linear Transformations as Matrices

This section is going to be different from the other ones. Instead of definitions, theorems, and corollarys, we are going to cover how to construct matrices and then prove that matrices are indeed the construction we want them to be.

First, we should start with some motivation about why its important to have another representation of linear transformations. Let's say I asked you to define the linear transformation rotate counterclockwise by 90 degrees about the origin in the vector space $\mathbb{R}^2$. I think you'd have a pretty hard time trying to define some sort of $T(\vec{x})$ equals something for all of $\mathbb{R}^2$. However, this has a very simple representation when using matrices. It is simply $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. This is just a simple example and motivation. Matrices also have properties which make understanding what abstract linear transformations do way easier such as determinants and eigenvalues. Overall, matrices help us understand, represent, and compute linear transformations much easier.

We've covered how linear transformations can be fully defined by a basis. Moreover, they are fully defined by a basis in the domain and an arbitrary same sized list in the codomain of the linear transformation. Let V and W be $\mathbb{F}$-vector spaces, $\vec{v} \in V$, and $B = (\vec{v}_1, \ldots, \vec{v}_n)$ be a basis of V. Let $T : V \to W$ be a linear transformation. We showed that the following is true:

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

This actually shows that the image of T, the set of all vectors in the codomain that T maps to, is spanned by the linear transformation applied to the vectors of B.

Now, if you recall, we discussed coordinate vectors some time ago and mentioned that it would be important for matrices and linear transformations. We noted that the standard way we write vectors is as the coordinate vector of the elementary basis. This is also true when we write $T(\vec{v}) = \sum_{i=1}^{n} a_i T(\vec{v}_i)$. $T(\vec{v})$ is written with respect to the elementary basis since the $T(\vec{v}_i)$'s are also written with respect to the elementary basis.

What if we wanted to consider the transformed vectors with respect to a different basis? The elementary basis is the most common basis to work in, but there are certain scenarios where having a different basis can actually make representing vectors much easier. One of these that we will cover later is Diagonalization which makes computing large powers of matrices substantially easier. Let C be a basis of W. From the properties of coordinate vectors, we can observe that:

$$[T(\vec{v})]_C = \sum_{i=1}^{n} a_i [T(\vec{v}_i)]_C$$

We still have this same property that the coordinate vector representation for each basis vector of the domain still spans the image. Let's try putting these together. What this means is consider lining up the coordinate vector representation for the basis B in a line.

$$\begin{bmatrix} | & | & \dots & | \\ [T(\vec{v}_1)]_C & [T(\vec{v}_2)]_C & \dots & [T(\vec{v}_n)]_C \\ | & | & \dots & | \end{bmatrix}_{CB}$$

Remark 2.3.1. You may notice that we are talking about a linear transformation from a basis B to C, but the subscript is CB instead of BC. This may seem unintuitive but we will soon see why the notation is like this in the context of composing linear transformations.

You might notice that this is actually a matrix! First, let's talk about the size of this matrix. It will have n columns where n is the size of B since there is one column per vector in the basis B. This is also just the dimension of the domain. It will have m rows where m is the size of the basis C. This is also just the dimension of the codomain. The number of rows is the size of C since the coordinate vector with respect to C has one entry per vector in C. This means that the linear transformation T from a n-dimensional vector space to an m-dimensional vector space can be contained in a $m \times n$ matrix.

We now want to show that this matrix does actually fully capture our linear transformation. That is, we want to show that in the same sense that we can apply a T to a vector in our domain, we can "apply" the matrix to a vector in our domain to get the same output. This means that we have to define a notion of "applying" a matrix so that it gives the same output as applying a linear transformation. This brings us back to the coordinate vectors we were working with earlier. Specifically, the equation:

$$[T(\vec{v})]_C = \sum_{i=1}^{n} a_i [T(\vec{v}_i)]_C$$

We can see that $[T(\vec{v})]_C$ can be expressed as a linear combination of the $[T(\vec{v}_i)]_C$'s. So all we have to do is figure out what the coefficients are. Now, we want to consider the vector of coefficients. This is simply $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_B$. Now it is important to discuss size again. We previously noted that our matrix is size $m \times n$. This vector is $n \times 1$. So what should we expect our output to be? If we want our matrix to be the same as our linear transformation, our output vector should be a coordinate vector in the vector space W. Since W has dimension m, our output should be $m \times 1$.

If you are already familiar with matrices and matrix multiplication then you already know that you can only multiply two matrices if the number of columns of the first matrix is equal to the number of rows of the second matrix. We can see here that our matrix is $m \times n$ and our vector is $n \times 1$ so it upholds this rule. Some intuition for why this rule makes sense is that the number of columns of the first matrix corresponds to the size of the domain vector space. Since we want to input a vector from the domain, if it has a different size then it isn't actually from our domain.

Let's consider "applying" our matrix to the vector we defined above. Quickly, we should discuss what is conceptually going on. The CB subscript that you may have noticed tells us that we are transforming from one basis to another. This is why we need to input a coordinate vector with respect to B since the way our matrix is set up is the transformation of the basis B written with respect to C. This yields:

$$\begin{bmatrix} | & | & \dots & | \\ [T(\vec{v}_1)]_C & [T(\vec{v}_2)]_C & \dots & [T(\vec{v}_n)]_C \\ | & | & \dots & | \end{bmatrix}_{CB} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_B$$

We know that $[T(\vec{v})]_C = \sum_{i=1}^{n} a_i [T(\vec{v}_i)]_C$ so one way we can define matrix multiplication is by taking the ith column of the matrix and scaling it by the ith row of the vector and then taking the sum across all the columns. This is commonly known as the column by coordinate method.

If you've learned how matrix multiplication works before, this is likely not the way you learned. You likely learned the row by column method. First, we will introduce some notation. Let our matrix be called A and our vector be called $\vec{x}$. We let A_{ij} denote the entry in the ith row and jth column of A and x_i denote the ith entry of $\vec{x}$. Row by column matrix multiplication is defined as follows:

$$(A\vec{x})_i = \sum_{j=1}^{n} A_{ij} x_j$$

In English, this tells us that the ith entry of the product is equal to the sum of the product of corresponding entries in the ith row of A.

Now let's think about why this is correct. We consider the following:

$$\begin{aligned} (A\vec{x})_1 &= A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n \\ (A\vec{x})_2 &= A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n \\ &\vdots \\ (A\vec{x})_m &= A_{m1}x_1 + A_{m2}x_2 + \ldots + A_{mn}x_n \end{aligned}$$

This part is especially important so really take some time to make sure you understand what is being said here. A_{ij} represents the jth coordinate of the vector $[T(\vec{v}_i)]_C$. This is the coefficient of the jth vector in C for the linear combination of $T(\vec{v}_i)$. x_j is the jth coordinate of the linear combination of $\vec{x}$ with respect to the basis B. You may notice that we are actually doing the exact same thing as the column by coordinate method. Notice that for each term in the equations above, if we simply consider the whole column of terms, it is literally just one of the $[T(\vec{v}_i)]_C$'s being scaled which is the exact same as the column by coordinate method.

Another way to think about this is to go back to the equation $[T(\vec{x})]_C = \sum_{i=1}^{n} a_i [T(\vec{v}_i]_C$. In this case, our a_i's have been denoted differently as x_i's. Take some time to convince yourself of this. The rest of the equation is the same. The row by column method simply deals with each entry of the output separately while the column by coordinate method deals with the output as a whole.

Example 2.3.2. Let's now work through a huge example with the derivative operator. Let $T : \mathbb{P}^2 \to \mathbb{P}^1$ be the derivative operator where $\mathbb{P}^n$ is the vector space of real coefficients polynomials of at most degree n. We consider the basis $B = (1, x, x^2)$ of $\mathbb{P}^2$ and the basis $C = (1, x)$ of $\mathbb{P}^1$. We want to find the matrix corresponding to T with respect to these two bases.

Let's start by determining the size of our matrix. We can see that we want a 2×3 matrix since $dim(\mathbb{P}^2) = 3$ and $dim(\mathbb{P}^1) = 2$. We now want to fill in our matrix with the coordinate vectors with respect to C of our basis B.

$$[T(1)]_C = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$[T(x)]_C = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$[T(x^2)]_C = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

So now let's combine these vectors to form our matrix A.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}_{CB}$$

We want to show that this is indeed the same as T. Let's consider $7x^2 + 5x + 3$. This can be represented as $\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}_B$. We can clearly see that the derivative $7x^2 + 5x + 3$ is $14x + 5$ which has coordinate vector $\begin{bmatrix} 5 \\ 14 \end{bmatrix}_C$. Let's see if our matrix gives us the same thing.

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}_{CB} \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}_B = \begin{bmatrix} 0 * 3 + 1 * 5 + 0 * 7 \\ 0 * 3 + 0 * 5 + 2 * 7 \end{bmatrix}_C = \begin{bmatrix} 5 \\ 14 \end{bmatrix}_C$$

Remark 2.3.3. Here we can already see a glimpse of why the subscript is written "backwards". We start from a vector written in a basis B and then apply the matrix. The matrix goes from basis B to basis C so if we simply follow the letters from right to left, we can determine what basis we end up in.

Example 2.3.4. We now want to start with a matrix and try to determine the linear transformation it comes from.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}_{CB}$$

Let $B = (\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix})$ and $C = (\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}7\\8\\9\end{bmatrix})$. From A, we can determine what happens to each of the basis vectors of B.

$$[T(\begin{bmatrix}1\\0\\0\\0\end{bmatrix})]_C = \begin{bmatrix}1\\5\\9\end{bmatrix}_C$$
$$[T(\begin{bmatrix}0\\1\\0\\0\end{bmatrix})]_C = \begin{bmatrix}2\\6\\10\end{bmatrix}_C$$
$$[T(\begin{bmatrix}0\\0\\1\\0\end{bmatrix})]_C = \begin{bmatrix}3\\7\\11\end{bmatrix}_C$$
$$[T(\begin{bmatrix}0\\0\\0\\1\end{bmatrix})]_C = \begin{bmatrix}4\\8\\12\end{bmatrix}_C$$

These are simply the columns of the matrix A. Let's say I wanted to use the elementary basis instead of C. How would I use these coordinate vectors in terms of C to get coordinate vectors for the elementary basis? We recall that the definition of a coordinate vector is that its coordinates and the coefficients in the linear combination representing that vector with respect to a basis. This means that $\begin{bmatrix}x\\y\\z\end{bmatrix}_C$ represents x times the first basis vector of C plus y times the second basis vector of C plus z times the third basis vector of C. Using this fact, we can determine what vectors each of the columns represent with respect to the elementary basis.

$$\begin{bmatrix}1\\5\\9\end{bmatrix}_C = 1 * \begin{bmatrix}1\\2\\3\end{bmatrix} + 5 * \begin{bmatrix}4\\5\\6\end{bmatrix} + 9 * \begin{bmatrix}7\\8\\9\end{bmatrix} = \begin{bmatrix}84\\99\\114\end{bmatrix}$$
$$\begin{bmatrix}2\\6\\10\end{bmatrix}_C = 2 * \begin{bmatrix}1\\2\\3\end{bmatrix} + 6 * \begin{bmatrix}4\\5\\6\end{bmatrix} + 10 * \begin{bmatrix}7\\8\\9\end{bmatrix} = \begin{bmatrix}96\\114\\132\end{bmatrix}$$
$$\begin{bmatrix}3\\7\\11\end{bmatrix}_C = 3 * \begin{bmatrix}1\\2\\3\end{bmatrix} + 7 * \begin{bmatrix}4\\5\\6\end{bmatrix} + 11 * \begin{bmatrix}7\\8\\9\end{bmatrix} = \begin{bmatrix}108\\129\\150\end{bmatrix}$$
$$\begin{bmatrix}4\\8\\12\end{bmatrix}_C = 4 * \begin{bmatrix}1\\2\\3\end{bmatrix} + 8 * \begin{bmatrix}4\\5\\6\end{bmatrix} + 12 * \begin{bmatrix}7\\8\\9\end{bmatrix} = \begin{bmatrix}120\\144\\168\end{bmatrix}$$

Now, I want to consider the linear transformation represented by A but instead of from $B \to C$, I want $B \to D$ where D is the elementary basis. Let this new matrix be A'. All I have to do now is swap the columns of A with the new coordinate vectors I computed in the previous step.

$$A' = \begin{bmatrix}84 & 96 & 108 & 120\\99 & 114 & 129 & 144\\114 & 132 & 150 & 168\end{bmatrix}_{DB}$$

Notice that A and A' capture the exact same linear transformation just between different bases. In general, when I want to switch between bases, I need to replace each of the columns with the coordinate vector for that column vector with respect to the new basis.

Now we have discussed how linear transformations can be represented as matrices. This means that every matrix is a linear transformation and vice versa. This means that we expect matrices to be able to capture all aspects of linear transformations. So far we have shown that $T(\vec{v}) = A\vec{v}$

where A is the matrix representing T. From this definition, we can derive how we should define scalar multiplication and vector addition for matrices. The proof is not shown here, but you should think about why these definitions make sense. When we scale a linear transformation/matrix, we simply scale each of the entries of the matrix. When we add linear transformations/matrices, we add the entries pointwise. An example is shown below:

$$c\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{bmatrix}$$

Remark 2.3.5. The set of all $m \times n$ matrices is actually a vector space. Typically denoted $\mathbb{M}_{m\times n}$. We define scalar multiplication and vector addition as above. You can check that it satisfies the axioms. This should make sense since we have already seen that the set of all linear transformations between two vector spaces is a vector space.

Now there remains one more aspect of linear transformations that we need to show that our matrices capture: composition. We need to somehow show that we can represent $T_A \circ T_B$ as some operation of the matrices A and B representing their respective transformations. Here, we will define matrix to matrix multiplication in almost the exact same manner as we did matrix to vector multiplication. This means that we will show $(T_A \circ T_B)(\vec{v}) = AB\vec{v}$.

Let A be an $m \times n$ and B be a $n \times r$ matrix. Note that we still require the number of columns in the first matrix to be equal to the number of columns in the second matrix (why?). We define the matrix product AB to be the $m \times r$ matrix satisfying:

$$AB_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \tag{2.1}$$

Reminder that A_{ik} denotes the entry in the ith row and kth column of A. Let's take a second to understand the above formula. As we increment k, we notice that A_{ik} is moving along the ith row. If k increases by one, it moves right one column. Similarly, B_{kj} is moving along the jth column. If k increases by one, it moves down one row.

Hopefully, this looks at least a little familiar. As mentioned previously, this is almost the exact same formula as the row by column matrix-vector multiplication we described previously. That formula is reproduced down below for an $m \times n$ matrix A and $n \times 1$ vector $\vec{x}$:

$$(A\vec{x})_i = \sum_{j=1}^{n} A_{ij}x_j$$

Let's think about why this might be the case. Remember that our goal is to show that multiplying two matrices is the same as composing their corresponding linear transformations. The way that you should think about this new formula is that we are performing matrix-vector multiplication on each of the column vectors. Take some time to convince yourself that this is indeed what (1) is doing.

Remark 2.3.6. Some intuition behind performing matrix-vector multiplication on each of the column vectors is that we are finding out where the original basis vectors end up. Recall that the columns tell us where each of the respective basis vectors end up. If we are composing linear transformations, we first see what the first linear transformation does to the basis vectors and then see what the second linear transformation does to that result.

Let A be a $m \times n$ matrix and B be a $n \times r$ matrix. We will say that A maps from a basis $\mathcal{B}$ to a basis $\mathcal{C}$ and that B maps from $\mathcal{A}$ to a basis $\mathcal{B}$. When we consider $AB\vec{x}$, we will consider the transformations one by one. First, $B\vec{x}$ will output a coordinate vector with respect to $\mathcal{B}$. Next, we do $A(B\vec{x})$ and since $B\vec{x}$ is a coordinate vector with respect to $\mathcal{B}$, we can simply apply A to it. Notice that this is actually equivalent to $(T_A \circ T_B)(\vec{x}) = T_A(T_B(\vec{x}))$.

When composing linear transformations, we have to be very careful that the codomain of the first transformation aligns with the domain of the second linear transformation. If we had previously had that A maps from a basis $\mathcal{D}$ to a basis $\mathcal{C}$, composing A and B would make no sense. $B\vec{x}$ outputs a coordinate vector with respect to $\mathcal{B}$. Trying to apply a linear transformation which only understands $\mathcal{D}$ coordinate vectors to a $\mathcal{B}$ coordinate vector wouldn't make any sense.

Remark 2.3.7. When we say "wouldn't make any sense", it doesn't mean that the math will magically break down and stop working. It actually means that you are composing a different linear transformation than the one you intended. If we did end up applying the $\mathcal{D}$ input A to $B\vec{x}$, all of the formulas would still work as long as the dimensions lined up. It's just that A would interpret the coordinates of $B\vec{x}$ with respect to $\mathcal{D}$. This is an issue because this would almost surely represent a different vector than when interpreted with $\mathcal{B}$.

Example 2.3.8. All that Remark 11.4 is saying is that if we had $\mathcal{D} = (\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix})$ and $\mathcal{B} = (\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix})$, they would interpret the coordinate vector $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$ differently. We would get that $\begin{bmatrix} 5 \\ 6 \end{bmatrix}_{\mathcal{D}} = 5 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 6 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 23 \\ 34 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 6 \end{bmatrix}_{\mathcal{B}} = 5 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$

Now let's consider an example of composing linear transformations. This example is a followup to Example 11.2. We're going to talk about how we can do that exact same process but now in the context of composing linear transformations.

Example 2.3.9. We will now find a way to generalize switching a vector between bases. Let V be a $\mathbb{F}$-vector space and $\vec{v} \in V$ with B, C, and D as bases. We want to find a matrix that when applied to $[\vec{v}]_B$ gives us $[\vec{v}]_C$. Since we are looking for a matrix, we know that this must be a linear transformation. One important thing to notice is that $[\vec{v}]_B$ and $[\vec{v}]_C$ actually encode the same vector for all $\vec{v} \in V$. This means that our linear transformation must be the identity transformation.

This gives us the linear transformation $Id : V \to V$ via $Id(\vec{v}) = \vec{v}$ for all $\vec{v} \in V$. We let A_{CB} be the corresponding matrix for this linear transformation. Like before, let's say we wanted A_{DB} when we already know what A_{CB} is.

We are actually going to find Id_{DC} and compose it with A_{CB} to get $Id_{DC} \circ A_{CB}$. There are a few things to notice about this composition. Firstly, if we just follow the letters from right to left we get that $B \to C \to D$ so at the very least it seems like we are getting the bases correctly. Secondly, since Id_{DC} is the identity transformation, this composition is still the same linear transformation as A.

Id_{DC} is known as the change of basis matrix from C to D. You would compute it in the same way that you would the matrix for any linear transformation.

Remark 2.3.10. If we write a matrix without a subscript say BA representing two bases, we automatically assume that we are mapping from the elementary basis of the domain to the elementary basis of the codomain.

There are many types of matrices and we define them below. The names are pretty accurate descriptions of what the matrices look like so they shouldn't be hard to remember.

Definition 2.3.11. We say that a matrix A is square if it is $n \times n$.

Definition 2.3.12. Let A be a $m \times n$ matrix. We define the transpose of A, denoted A^T to be the $n \times m$ matrix satisfying $A_{ij} = A^T_{ji}$.

Remark 2.3.13. The transpose simply takes the rows of A and makes them the columns of A^T.

Definition 2.3.14. Let A be a square matrix. The main diagonal of A are the entries a_{ij} such that $i = j$.

Definition 2.3.15. A diagonal matrix is a matrix where the only nonzero entries are on the main diagonal.

Definition 2.3.16. An upper triangular matrix is zero below the main diagonal. A lower triangular matrix is zero above the main diagonal.

Example 2.3.17.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1 & 3 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ 6 & 8 & 3 \end{bmatrix} \tag{3}$$

(1) is diagonal, lower triangular, and upper triangular. (2) is upper triangular. (3) is lower triangular.

2.4 Exercises 4

Exercise 2.4.1. Let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation of counterclockwise rotation about the origin by θ radians. Let $\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\vec{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

(a) Let $T_\theta(\vec{v}_1) = \vec{w}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$. Determine the numerical length of $\vec{w}_1$.

(b) Compute $\cos(\theta)$ and $\sin(\theta)$.

(c) Using part (b), determine $\vec{w}_1$.

(d) Let $T_\theta(\vec{v}_2) = \vec{w}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Determine the numerical length of $\vec{w}_2$.

(e) In trigonometry, we compute our sin and cos values with respect to angles starting from the positive x-axis which is not what we currently have as θ. What is the angle formed by $\vec{w}_2$ and the positive x-axis?

(f) Using part (e), find the cos and sin values of the angle formed by $\vec{w}_2$ and the x-axis.

(g) Use the following two identities to simplify your expression: $\cos(\frac{\pi}{2}+\theta) = -\sin(\theta)$ and $\sin(\frac{\pi}{2}+\theta) = \cos(\theta)$

(h) Determine $\vec{w}_2$.

(i) What is the matrix corresponding to T_θ from the elementary basis to the elementary basis?

Exercise 2.4.2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that represents reflection across the line $y = x$.

(a) Compute the matrix representing T from the standard basis to the standard basis.

(b) Prove $B = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$ is a basis of $\mathbb{R}^2$

(c) Compute the matrix representing T from B to B

Exercise 2.4.3. Let $B = \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$ and $C = \left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right)$ be bases. Find the change of basis matrix from B to C and use it to compute $[\vec{v}]_C$ where $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}_B$.

Exercise 2.4.4. Let A be an $m \times n$ matrix. We define the column space of A to be the span of the columns of A. Prove that the column space of A is the same as the image of T_A.

Exercise 2.4.5. Let A be a $m \times n$ matrix and B be a $n \times r$ matrix.

(a) Prove that matrix multiplication is not commutative. This means that, in general, $AB \neq BA$

(b) Give an example of an A and B which commute.

Exercise 2.4.6. Let $A = \left(\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 5 \\ 0 \end{pmatrix} \right)$ and $B = \left(\begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ -1 \end{pmatrix} \right)$. Note that these are both bases of $\mathbb{R}^3$. Find the change of basis matrix from $B \to A$.

Proof. $\begin{bmatrix} 31 & -15 & -49 \\ 17 & -7 & -25 \\ -6 & 3 & 10 \end{bmatrix}$ □

Exercise 2.4.7. Let A be an $m \times n$ matrix and B be an $n \times r$ matrix. Prove that $(AB)^T = B^T A^T$.

2.5 Fundamental Theorem of Linear Algebra

We will now state the Fundamental Theorem of Linear Algebra (technically three theorems but they're all really similar). One thing to keep in mind before we state the theorem is that you will not understand parts of the theorem. By the end of Chapter 15, you should understand and know how to prove the Fundamental Theorem of Linear Algebra.

Theorem 2.5.1. Given an $m \times n$ matrix A and the corresponding linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\vec{x}) = A\vec{x}$, the following are equivalent:

1. T_A is injective
2. $A\vec{x} = \vec{b}$ has at most one solution for all $\vec{b} \in \mathbb{F}^m$
3. $ker(T_A) = \{\vec{0}_{\mathbb{F}^n}\}$
4. The columns of A are linearly independent in $\mathbb{F}^n$
5. A has a left inverse
6. $rank(A) = n$

Moreover, these imply that $m \geq n$.

Theorem 2.5.2. Given an $m \times n$ matrix A and the corresponding linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\vec{x}) = A\vec{x}$, the following are equivalent:

1. T_A is surjective
2. $A\vec{x} = \vec{b}$ has at least one solution for all $\vec{b} \in \mathbb{F}^m$
3. $Im(T_A) = \mathbb{F}^m$
4. The columns of A span $\mathbb{F}^m$
5. A has a right inverse
6. $rank(A) = m$

Moreover, these imply that $n \geq m$.

Theorem 2.5.3. Given an $m \times n$ matrix A and the corresponding linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\vec{x}) = A\vec{x}$, the following are equivalent:

1. T_A is bijective
2. $A\vec{x} = \vec{b}$ has exactly one solution for all $\vec{b} \in \mathbb{F}^m$
3. $ker(T_A) = \{\vec{0}_{\mathbb{F}^n}\}$ and $Im(T_A) = \mathbb{F}^m$

4. The columns of A form a basis for $\mathbb{F}^m$
5. A is invertible
6. $rank(A) = m = n$

Moreover, these imply that $m = n$.

Hopefully, some of these are believable. Specifically, (1), (2), and (4) for each of the three theorems. Additionally, you should believe that Theorem 13.1 and Theorem 13.2 imply Theorem 13.3. Take the time to convince yourself that you, at least to some degree, believe the last two sentences. We will spend the rest of this section proving that (1), (2), and (4) imply each other.

We will only prove Theorem 13.1. Theorem 13.2 will be left as an exercise to test your understanding of the proofs used in Theorem 13.1.

Lemma 2.5.4. T_A is injective if and only if $A\vec{x} = \vec{b}$ has at most one solution for all $\vec{b} \in \mathbb{F}^m$.

Proof. We will prove the forward direction via the contrapositive. Assume that $A\vec{x} = \vec{b}$ has more than one solution for a $\vec{b} \in Im(T_A)$. Note that it must be in the image of T_A otherwise $A\vec{x} = \vec{b}$ simply has no solution. There exists $\vec{x}_1, \vec{x}_2 \in \mathbb{F}^m$ where $\vec{x}_1 \neq \vec{x}_2$ and $A\vec{x}_1 = A\vec{x}_2 = \vec{b}$. This shows that T_A is not injective.

We will also prove the reverse direction via the contrapositive. Assume that T_A is not injective and fix a $\vec{b} \in \mathbb{F}^m$. This means that there exists some $\vec{x}_1, \vec{x}_2 \in \mathbb{F}^n$ with $\vec{x}_1 \neq \vec{x}_2$ such that $T_A(\vec{x}_1) = T_A(\vec{x}_2) = \vec{b}$. By definition, $T_A(\vec{x}) = A\vec{x}$ so we have that $T_A(\vec{x}_1) = A\vec{x}_1 = \vec{b} = A\vec{x}_2 = T_A(\vec{x}_2)$. This shows that $A\vec{x} = \vec{b}$ there exists a $\vec{b} \in \mathbb{F}^m$ where $A\vec{x} = \vec{b}$ has more than one solution. □

Lemma 2.5.5. T_A is injective if and only if the columns of A are linearly independent.

Proof. We will prove the forward direction via the contrapositive. We assume that the columns of A are not linearly independent. Recall the column by coordinate definition of matrix-vector multiplication. We notice that if we have $A\vec{x} = \vec{b}$, $\vec{b}$ is a linear combination of the columns of A. Thus, if the columns of A are linearly dependent, there are multiple linear combinations which make $\vec{b}$. This also means that there are multiple $\vec{x}$'s which cause $A\vec{x} = \vec{b}$. Since, $T_A = A\vec{x}$, we have that T_A is not injective.

The reverse direction follows the exact same logic as the forward direction. We proceed by contrapositive. Assume that T_A is not injective and fix $\vec{b} \in \mathbb{F}^m$. This means that there exists $\vec{x}_1, \vec{x}_2 \in \mathbb{F}^n$ with $\vec{x}_1 = \vec{x}_2$ such that $T_A(\vec{x}_1) = T_A(\vec{x}_2) = \vec{b}$. Since $T_A(\vec{x}) = A\vec{x}$, we have that $A\vec{x}_1 = A\vec{x}_2 = \vec{b}$ which shows that there exists a $\vec{b} \in \mathbb{F}^m$ such that $A\vec{x} = \vec{b}$ has more than one solution. □

2.6 Systems of Linear Equations

We're going to take a step back from all the linear transformation stuff and transition into talking about systems of equations. This is a very very practical application of matrices as they make solving systems of linear equations highly algorithmic. Row reduction is an algorithm you can very easily define and execute. If you have taken a linear algebra course before, this is likely something you are familiar with.

Let's consider the following system of equations:

$$\begin{aligned} 5x - 2y + 9z &= -7 \\ -2x + y - 4z &= 5 \\ 3x - 10y - 8z &= 0 \end{aligned}$$

We can actually model this system using a 3×4 matrix shown below:

$$\left[\begin{array}{ccc|c} 5 & -2 & 9 & -7 \\ -2 & 1 & -4 & 5 \\ 3 & -10 & -8 & 0 \end{array}\right]$$

Notice that the first column corresponds to the coefficients of x, the second column for y, and the third for z. The last column holds the solutions. Moreover, the first row represents the first equation and so on.

When working with systems of linear equations in matrices, we have three elementary operations we can perform: scaling a row by a nonzero constant, swapping two rows, and adding one row to another row. Notice that these actually correspond with the operations we typically do when solving systems of equations. We can scale equations and add equations. Swapping rows/equations isn't actually necessary when we solve systems of equations but we'll soon see that it makes understanding what solving a system really means much easier.

Definition 2.6.1. We define a elementary matrix to be a matrix that is exactly one elementary operation away from the identity matrix.

Example 2.6.2.

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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{3}$$

(1) represents adding the third row to the first row. Notice that it simply adds the third row of the identity to the first row of the identity. (2) is the elementary matrix that represents scaling the second row by -5. Again, notice that it simply scales the second row of the identity by -5. (3) is the elementary matrix representing swapping the first and third row. Once again, it simply swaps the first and third rows of the identity matrix. Hopefully, you can see that to find the elementary matrix representing an operation, you simply perform that operation on the identity.

2.7 Matrix Inverses

One way to think about Gaussian/Gauss Jordan Elimination is that we are simply chaining together elementary matrices. In other words, we are finding $E_1, E_2, \ldots, E_k$ such that $E_k E_{k-1} E_{k-2} \ldots E_2 E_1 A = RREF(A)$. It might help to think about this as each E_i is a step that we are using to unwind the system of equations. This type of thinking will help when we discuss finding inverses.

Definition 2.7.1. Let A be a $n \times n$ matrix. The inverse of A is defined to be the matrix A^{-1} such that $A^{-1}A = AA^{-1} = I_n$

Let's say we had a $n \times n$ matrix A such that $RREF(A) = I_n$. When determining what $RREF(A)$ is using Gaussian/Gauss Jordan Elimination, we found a sequence of elementary matrices that when applied to A, leads to I_n. This means that we have some elementary matrices such that $E_k E_{k-1} E_{k-2} \ldots E_2 E_1 A = RREF(A) = I_n$. We define $A^{-1} = E_k E_{k-1} E_{k-2} \ldots E_2 E_1$. This definition should be believable since clearly $A^{-1}A = I_n$. Furthermore, this reveals something else about A, but first we should discuss the inverse of elementary matrices.

Elementary matrix inverses are actually really easy to find. If I wanted to find the matrix inverse of the elementary matrix corresponding to swapping row one and row five, what would I do? Well, I would simply swap them back. Thus, the inverse of swapping row one and five is to swap row one and five again. This same logic can be applied to the other elementary matrices. This actually also tells us that all elementary matrices are invertible and that their inverses are also elementary matrices.

Now, let's uncover what new information we have gained about A. Briefly, when we have an equation say $7x = 5$, we can "move" the 7 to the other side by multiplying both sides by $\frac{1}{7}$ which is the multiplicative inverse of 7. We can do the same with matrices.

Theorem 2.7.2. Let A be a $n \times n$ matrix. A is invertible if and only if A is a product of elementary matrices.

Proof. For the forward direction, we assume that A is invertible and use the $A^{-1} = E_k E_{k-1} \ldots E_2 E_1$ definition where E_i is the ith step of Gaussian Elimination.

$$\begin{aligned}
E_k E_{k-1} E_{k-2} \ldots E_2 E_1 A &= I_n \\
E_k^{-1} E_k E_{k-1} E_{k-2} \ldots E_2 E_1 A &= E_k^{-1} I_n \\
E_{k-1} E_{k-2} \ldots E_2 E_1 A &= E_k^{-1} \\
E_{k-1}^{-1} E_{k-1} E_{k-2} \ldots E_2 E_1 A &= E_{k-1}^{-1} E_k^{-1} \\
E_{k-2} \ldots E_2 E_1 A &= E_{k-1}^{-1} E_k^{-1} \\
&\vdots \\
A &= E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1}
\end{aligned}$$

To do the reverse direction, we actually just follow the above equations but backwards. $\square$

The way that we actually get all these E_i's is by performing Gaussian Elimination. So far, we have only examined what happens if $RREF(A) = I_n$, but what if it doesn't?

Theorem 2.7.3. Let A be a $n \times n$ matrix. A is invertible if and only if $RREF(A) = I_n$.

Proof. This follows from the previous theorem. $\square$

Based on this theorem, we can deduce that one way to check if A is invertible, is to perform Gaussian Elimination and see if you get $RREF(A) = I_n$. While doing Gaussian Elimination, each step you do is also one of the E_i's.

After all this talk about products of elementary matrices, you may have thought about the fact that this definition is quite impractical. We can get the E_i's simply from performing Gaussian Elimination, but to get a singular matrix A^{-1}, I would have to actually perform all the matrix multiplications. Even if these are elementary matrices, it can still be a pain since as matrices get larger you will have many more elementary matrices from your Gaussian Elimination.

To get around this, we notice that, no matter what, we are performing Gaussian Elimination. We need to perform Gaussian Elimination to check if A is invertible and, if it is, along the way we have also found all of the E_i's. One way to make this more efficient is to simply also multiply the E_i's as you are performing Gaussian Elimination.

Example 2.7.4. Let's prove that $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$ is invertible and find its inverse. We will do this by considering the augmented matrix $[A|I_3]$

$$\begin{aligned}
[A|I_3] &= \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] && \text{(swap rows 1 and 2)} \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array}\right] && \text{(add -4 times row 1 to row 3)} \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] && \text{(add 3 times row 2 to row 3)} \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] && \text{(add -1 times row 3 to row 2)} \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right] && \text{(scale row 3 by } \tfrac{1}{2}\text{)} \\
&= \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right] && \text{(add -3 times row 3 to row 1)}
\end{aligned}$$

Clearly, $RREF(A) = I_n$ so A is invertible. The right half of this augmented matrix is now actually A^{-1}. Let's discuss why.

We start with $[A|I_3]$. Since A is invertible, we can rewrite this as $[E_kE_{k-1}\dots E_1I_3|I_3]$. Every step we did in Gaussian Elimination, undoes one of these E_i's until we eventually got to E_1. Undoing one of the E_i's is equivalent to multiplying by E_i^{-1}. So, after the first step we get

$$[E_k^{-1}E_kE_{k-1}\dots E_1I_3|E_k^{-1}I_3] = [E_{k-1}\dots E_1I_3|E_k^{-1}I_3]$$

Then the second step gives

$$[E_{k-1}^{-1}E_{k-1}\dots E_1I_3|E_{k-1}^{-1}E_k^{-1}I_3] = [E_{k-2}\dots E_1I_3|E_{k-1}^{-1}E_k^{-1}I_3]$$

This should look familiar. It's the exact same process as we did in proving Theorem 2.7.2.

Remark 2.7.5. Let A be an invertible $n \times n$ matrix. If I want to solve $A\vec{x} = \vec{b}$, I can simply do

$A^{-1}A\vec{x} = A^{-1}\vec{b}$ to get $\vec{x} = A^{-1}\vec{b}$.

To recap, now you know how to check invertibility of square matrices and how to find their inverses. We use Gaussian Elimination to both check invertibility and compute the inverse. This is actually a really nice and simple way of understanding and computing inverses. Sadly, it is actually much more involved to find and prove left/right inverses for non-square matrices. Now let's find out how.

Let A be a $m \times n$ matrix and $T_A : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation via $T_A(\vec{x}) = A\vec{x}$. We will assume that T_A is surjective and thus that $n \geq m$. We will first attempt to construct a right inverse in an example and then use some algebra to show that this is indeed the right inverse.

Example 2.7.6. Assume A is 3×4. Let $A_r = RREF(A)$. We don't actually care what the values in A are, but assume that A_r is shown below. Moreover, $A_r = GA$ for some 3×3 matrix G. Note that G represents the sequence of row operations we performed in Gaussian Elimination to get $RREF(A)$.

$$A_r = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We are going to first find a right inverse for A_r and then show that from this right inverse, we can get the right inverse of A. Some motivation for why we are doing this is that when we found the inverse for A, we analyzed how $RREF(A)$ which was the identity matrix. We are hoping that we can do something kind of similar here. Additionally, finding an inverse for something in reduced row-echelon form is substantially easier than any arbitrary matrix since most of the entries are zero.

By definition of right inverse, we know that we want $A_r A_r^{-1} = I_3$. If A_r is 3×4 and I_3 is 3×3, what are the dimensions of A_r^{-1}? A_r^{-1} must be 4×3. This is because the dimension of the codomain of A_r^{-1} must match the dimension of the domain of A_r and the dimension of the domain of A_r^{-1} must match the dimensions of the domain of I_3.

Now let's consider what we have so far.

$$A_r A_r^{-1} = I_3$$

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Our goal will be to "copy" the leading 1's from A_r into the 1's on the main diagonal of I_3. The way we will do this is by considering something similar to the transpose of A_r. Consider the 4×3 matrix given by $B_{ji} = 1$ if A_{ij} is a leading one/pivot and 0 otherwise.

This gives us the following matrix:

$$A_r A_r^{-1} = I_3$$

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that we do not copy over the 3 because it is not a leading one. If you perform the matrix multiplication, you can see that this is clearly a right inverse for A_r. We will algebraically prove that this method of constructing the right inverse for A_r will always work.

Now that we have a right inverse of A_r, let's see if we can construct a right inverse of A. We want to exploit the fact that we know $A_r = GA$. Additionally, we actually know that G is invertible because it is a product of elementary matrices.

$$\begin{aligned} A_r &= GA \\ G^{-1}A_r &= A \\ G^{-1}A_r A_r^{-1} &= AA_r^{-1} \\ G^{-1} &= AA_r^{-1} \\ G^{-1}G &= AA_r^{-1}G \\ I_3 &= AA_r^{-1}G \end{aligned}$$

Thus, we have that $A_r^{-1}G$ is a right inverse of A.

We have now conjectured that a right inverse of A_r is this transpose like construct we defined earlier. Let's prove it algebraically. Let the "transpose like construct" be denoted B_r. Let A_r be a $m \times n$ matrix. Let's consider what $A_r B_r$ is.

$$(A_r B_r)_{ij} = \sum_{k=1}^{n} A_{r_{(ik)}} B_{r_{(kj)}}$$

We know that the desired product, I_m, has 1's along the main diagonal and zeros everywhere else. Thus, we know that $I_{m_{ij}} = (A_r B_r)_{ij}$. Moreover, if $i = j$ (on the main diagonal), then $I_{m_{ij}} = (A_r B_r)_{ij} = 1$. Otherwise, $I_{m_{ij}} = (A_r B_r)_{ij} = 0$. Let's consider the $i = j$ case first.

$$\begin{aligned} (A_r B_r)_{ij} &= \sum_{k=1}^{n} A_{r_{(ik)}} B_{r_{(kj)}} \\ (A_r B_r)_{ii} &= \sum_{k=1}^{n} A_{r_{(ik)}} B_{r_{(ki)}} \end{aligned}$$

We know that $B_{r_{ki}} = 1$ if and only if $A_{r_{ik}}$ is a leading one and $B_{r_{ki}} = 0$ otherwise. As we increment k, we are moving across the ith row of A_r. Since A_r is in reduced row echelon form and A_r is surjective, it has exactly one leading one per row. If we didn't have the fact that A is surjective, then we don't know that A_r has exactly one leading one per row. It could have a row of zeros.

From the definition of B_r, we know that B_r is zero in the ith column except for the corresponding leading one in row i of A_r. This means that all terms in the above summations are 0 except for the $A_{r_{(ik)}} B_{r_{(ki)}}$ where $A_{r_{(ik)}}$ is a leading one. Thus:

$$(A_r B_r)_{ii} = \sum_{k=1}^{n} A_{r_{(ik)}} B_{r_{(ki)}} = 1$$

Now let's consider what happens if we are off the main diagonal of I_m so $i \neq j$.

2.8 Exercises 5

Exercise 2.8.1. Consider an $m \times n$ matrix A and the corresponding linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\vec{x}) = A\vec{x}$. Prove the following:

(a) T_A is surjective if and only if $A\vec{x} = \vec{b}$ has at least one solution for all $\vec{b} \in \mathbb{F}^m$

(b) T_A is surjective if and only if the columns of A span $\mathbb{F}^m$

(c) Why would it follow that $n \geq m$? Moreover, why is this statement made separate from the "The following are equivalent" statements (why would it be false to place this as one of the "The following are equivalent" statements)?

Exercise 2.8.2. Let E_1 be the elementary matrix representing a row swap, E_2 be the elementary matrix representing scaling a row, and E_3 be the elementary matrix representing adding two rows. We now know what E_1A, E_2A, and E_3A do, but what do AE_1, AE_2, and AE_3 do?

Exercise 2.8.3. Find the inverse of $\begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$.

Exercise 2.8.4. Prove that the definition of A^{-1} as $E_kE_{k-1}\ldots E_2E_1$ is actually a two sided inverse. Note that we have already shown that it is a left inverse so you only need to show that it is a right inverse.

Exercise 2.8.5. Let A be a $n \times n$ matrix. Prove that A^{-1} is unique.

Exercise 2.8.6. Let A be a $m \times n$ matrix and $T_A : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation via $T_A(\vec{x}) = A\vec{x}$. Prove that if T_A is injective then A has a left inverse.

2.9 Image and Kernel

Definition 2.9.1. Let $T : V \to W$ be a linear transformation. The image of T, denoted $Im(T)$, is the set of all vectors $\vec{w} \in W$ such that there exists a $\vec{v} \in V$ where $T(\vec{v}) = \vec{w}$.

Remark 2.9.2. Although linear transformations and matrices are equivalent, different terminology is typically used for the image of a matrix. When talking about the image of a matrix, we call it the column space.

Definition 2.9.3. Let $T : V \to W$ be a linear transformation. The kernel of T, denoted $ker(T)$, is the set of all vectors $\vec{v} \in V$ such that $T(\vec{v}) = \vec{0}_W$.

Remark 2.9.4. Similarly, when discussing the kernel of a matrix, we instead call it the nullspace.

Theorem 2.9.5. Let $T : V \to W$ be a linear transformation. Then $\vec{0}_V \in ker(T)$.

Proof. We previously proved that for any vector spaces V, W and linear transformation $T : V \to W$, $T(\vec{0}_V) = \vec{0}_W$ □

Consider $A\vec{x} = \vec{b}$. The image of T_A is the set of all $\vec{b}$'s such that $A\vec{x} = \vec{b}$ has a solution. Consider $A\vec{x} = \vec{0}$. Notice that the solution space, the set of all $\vec{x}$'s which satisfy $A\vec{x} = \vec{0}$, is actually just the kernel of T_A.

Definition 2.9.6. Let A be a matrix. We define the rank of A, denoted $rank(A)$, to be the dimension of the image of T_A.

Definition 2.9.7. Let A be a matrix. We define the nullity of A, denoted $null(A)$, to be the dimension of the kernel of T_A. If $ker(T_A) = \{\vec{0}\}$, we say that $null(A) = 0$.

Example 2.9.8. Let $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. We will consider the equation $A\vec{x} = \vec{b}$. We are looking for all the $\vec{b}$'s such that the above equation has a solution since that it is the image.

$$A\vec{x} = \vec{b}$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

This tells us $Im(T_A) = \{\begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} : a \in \mathbb{F}\}$. Additionally, $rank(A) = 1$ since $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ forms a basis for $Im(T_A)$. We can now do something similar to find the kernel and nullity by looking at solutions to $A\vec{x} = \vec{0}$.

$$A\vec{x} = \vec{b}$$
$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} x_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We notice that we must have $x_3 = 0$, but x_1 and x_2 can be any value. Thus, the $ker(T_A) = \{\vec{x} : \vec{x} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}, a, b \in \mathbb{F}\}$. Moreover, we have that $null(A) = 2$ because $(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix})$ form a basis.

Example 2.9.9. Let $A = \begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 \\ 1 & 1 & -1 & -1 & 0 \end{bmatrix}$. We can compute the row echelon form of A to be $\begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 0 & 0 & -3 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. This is actually enough to tell us a basis of the image. A basis of the image is the columns containing the leading numbers in the original matrix A. A basis of the image is simply $\left(\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 3 \\ -1 \end{pmatrix}\right)$

Chapter 3

Determinants

3.1 Motivation

By default, we can draw a coordinate plane over $\mathbb{R}^2$ by letting our elementary basis vectors $(1, 0)$ and $(0, 1)$ form grid lines. Notice that each of the squares in this grid has area 1.

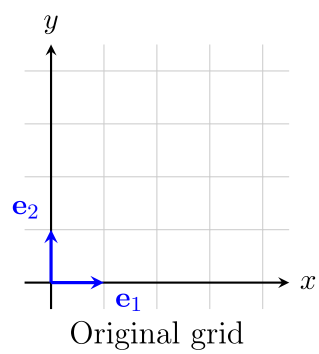
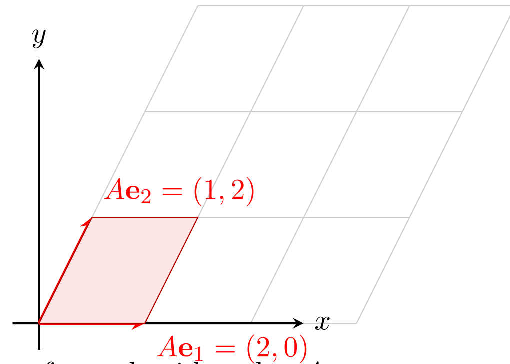


Figure 3.1: The standard basis and unit grid in $\mathbb{R}^2$.

Now consider a linear transformation on the standard basis by the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. We can see in the figure below what it does to elementary basis and how this shifts the grid lines. Now our squares in the grid are larger. With a little bit of geometry, you can see that the new area of each square is 4.



Transformed grid under A

Figure 3.2: The grid transformed by $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$.

The determinant of a linear transformation/matrix is a constant that is intended to capture this exact idea of area dilation. So for the example above, we would want the determinant of A to be 4 since the area grows four times. For this section, we will first talk about what properties we need the determinant to have in order to properly capture this dilation of area and then construct the determinant itself. We will see that if we define the properties in a certain way, the determinant is actually unique.

Remark 3.1.1. The notion of "area" sort of breaks down if we are in arbitrary vector spaces/fields for example $\mathbb{P}^n$. However, everything we will discuss in this chapter will still hold. It is just much harder to interpret in arbitrary vector spaces/fields.

3.2 Desired Properties

Let's begin discussing the desired properties. We will denote the determinant of a matrix A as $\det(A)$. First of all, this notion of changing area only makes sense if we stay in the same dimension. If I transform from a square in $\mathbb{R}^2$ to a parallelepiped in $\mathbb{R}^3$, I'm changing from area to volume which is a huge distinction. How would you describe a change from area 4 to volume 16 in terms of a singular constant? You can't or at the very least people don't seem to have figured out how. Because of this, we only define the determinant for square matrices in which the transformation does not change dimension.

Remark 3.2.1. I will be using the term "area" a lot, but you can interchange it with "volume" for higher dimension and everything will still hold.

If we want a notion of changing area, we need a baseline for what area is. A simple way to do this is to look at the unit square/cube in n dimensions. This always has area 1 so it is a good baseline. Moreover, we can represent the unit cube using the identity matrix.

Definition 3.2.2. Let $n \in \mathbb{N}$. We define $\det(I_n) = 1$. This is known as the *normalization* of the determinant.

Now let's analyze what happens to the area under operations like vector addition and scalar multiplication. We'll again begin with the matrix $A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, but now we will compare it to the matrix $A_2 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. This matrix results from adding $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to the transformation of $\vec{e}_1$.

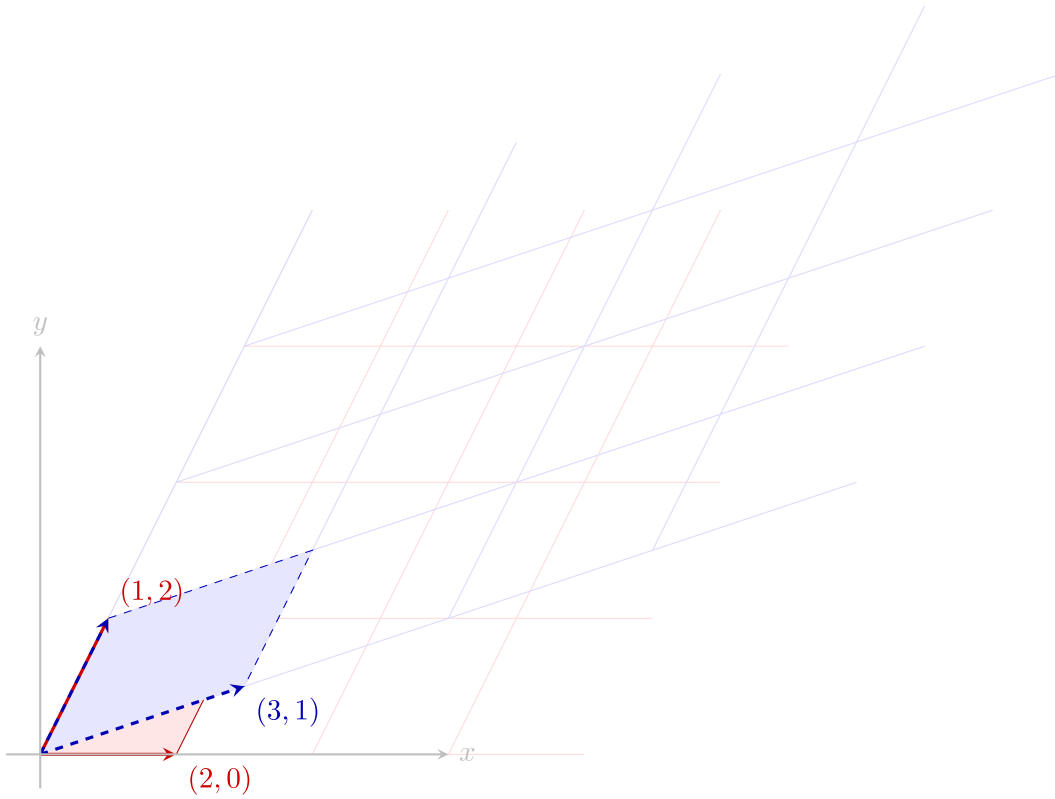


Figure 3.3: Comparing the new transformation

Definition 3.2.3. Let V be a vector space over a field $\mathbb{F}$. Let A be an $n \times n$ matrix where $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n]$ where $\vec{a}_i$ is the ith column of A. For any arbitrary $\vec{v} \in V$ and $\alpha \in \mathbb{F}$, we define $\det(\vec{a}_1, \ldots \alpha\vec{a}_i + \vec{v}, \ldots, \vec{a}_n) = \alpha \det(\vec{a}_1, \ldots, \vec{a}_{i-1}, \vec{a}_i, \ldots, \vec{a}_n) + \det(\vec{a}_1, \ldots, \vec{a}_{i-1}, \vec{v}, \ldots, \vec{a}_n)$. This is known as the *multi-linearity* property.

One can do all the geometry and check that this property does in fact hold for the above transformation. However, it is much more useful to think about it broadly. If I scale one of the vectors, my area clearly increases by the scalar. If I add to one of my vectors, I am adding a new parallelogram to my current one which results in us adding the determinants.

Now let's look at another simple property. We have been analyzing what happens to the grid squares formed by the basis vectors, but what happens when the vectors are not linearly independent? For example, what if my transformation is $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$? This transformation maps $(1, 0)$ and $(0, 1)$ to $(1, 0)$.

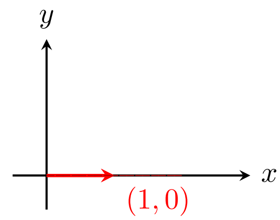


Figure 3.4: When both vectors are identical, the parallelogram collapses into a line, and the area is zero.

So in this case, we want the determinant to be 0. Briefly, let's talk about what this means in higher dimensions. If I have two vectors that are the same, I've essentially lost one dimension worth of information. This means my cube in $\mathbb{R}^3$ becomes a square in $\mathbb{R}^2$ which has volume 0.

Definition 3.2.4. Let A be an $n \times n$ matrix. If A has two columns that are the same, we define $\det(A) = 0$. This is known as the *alternating* property.

You may notice that this definition seems a little off. For example, the matrix A above does not have two columns that are the same, but it should have determinant zero. This is a result of both the *alternating* and *multilinearity* property.

Theorem 3.2.5. Let A be an $n \times n$ matrix. If the columns of A are linearly dependent, then $\det(A) = 0$.

Proof. Let $A = [\vec{v}_1, \ldots, \vec{v}_n]$. Since the columns are linearly dependent, there exists a $\vec{v}_i$ that can be written as a linear combination of the other vectors. Without loss of generality, let $\vec{v}_1 = \sum_{i=2}^{n} \alpha_i \vec{v}_i$.

$$\begin{aligned}
\det(A) &= \det(\vec{v}_1, \ldots, \vec{v}_n) \\
&= \det\Big(\sum_{i=2}^{n} \alpha_i \vec{v}_i, \vec{v}_2, \ldots, \vec{v}_n\Big) \\
&= \sum_{i=2}^{n} \alpha_i \det(\vec{v}_i, \vec{v}_2, \ldots, \vec{v}_n) && \text{(by Multilinearity)}
\end{aligned}$$

Now we can simply go through the terms one by one. Since i ranges from 2 to n, no matter what the value of i is we will have two copies of $\vec{v}_i$ inside the determinant. By the alternating property, this means the determinant is zero for all possible values of i. Thus, the sum is zero. $\square$

Theorem 3.2.6. Consider $[\vec{v}_1, \ldots, \vec{v}_n]$ where $\vec{v}_i \in \mathbb{F}^n$. Then

$$\det(\vec{v}_1, \ldots \vec{v}_i, \ldots, \vec{v}_j, \ldots, \vec{v}_n) = \det(\vec{v}_1, \ldots, \vec{v}_i + \vec{v}_j, \ldots, \vec{v}_j, \ldots, \vec{v}_n)$$

Proof.

$$\begin{aligned}
\det(\vec{v}_1, \dots \vec{v}_i + \vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n) &= \det(\vec{v}_1, \dots \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n) + \det(\vec{v}_1, \dots \vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n) \\
&\qquad \text{(by multi-linearity)} \\
&= \det(\vec{v}_1, \dots \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n) + 0 \qquad \text{(by alternating)} \\
&= \det(\vec{v}_1, \dots \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n)
\end{aligned}$$

□

Remark 3.2.7. There is another notion of what the final property of a determinant should be called *antisymmetry*. This one says that if you swap two columns of a matrix, the sign of the determinant should switch. However, the determinant is supposed to be the unique function from $\mathbb{F}^{n^2} \to \mathbb{F}$ and *antisymmetry* is not a sufficient condition for uniqueness in $\mathbb{F}_2$ where $1 = -1$. In fact, *antisymmetry* says literally nothing in $\mathbb{F}_2$. But in $\mathbb{R}^n$ *antisymmetry* will result in the same unique determinant as *alternating* will.

In light of this fact, let's prove that our current properties imply *antisymmetry*.

Theorem 3.2.8. Consider $[\vec{v}_1, \dots, \vec{v}_n]$ where $\vec{v}_i \in \mathbb{F}^n$. Then

$$\det(\vec{v}_1, \dots \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n) = -\det(\vec{v}_1, \dots \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n)$$

Proof. Almost every line of this proof makes use of Theorem 3.2.6 which says that you can add columns to one another without changing the determinant.

$$\begin{aligned}
\det(\vec{v}_1, \dots \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n) &= \det(\vec{v}_1, \dots, \vec{v}_i + \vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n) \\
&= \det(\vec{v}_1, \dots, \vec{v}_i + \vec{v}_j, \dots, \vec{v}_j - (\vec{v}_i + \vec{v}_j), \dots, \vec{v}_n) \\
&= \det(\vec{v}_1, \dots, \vec{v}_i + \vec{v}_j, \dots, -\vec{v}_i, \dots, \vec{v}_n) \\
&= \det(\vec{v}_1, \dots, \vec{v}_i + \vec{v}_j + (-\vec{v}_i), \dots, -\vec{v}_i, \dots, \vec{v}_n) \\
&= \det(\vec{v}_1, \dots, \vec{v}_j, \dots, -\vec{v}_i, \dots, \vec{v}_n) \\
&= -\det(\vec{v}_1, \dots, \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n) \qquad \text{(by multi-linearity)}
\end{aligned}$$

□

3.3 Defining the Determinant

We have already proved significant things about the determinant without even defining it explicitly. In this section, we will finally define the determinant and prove that it is in fact a unique function. Possibly to some people's surprise, we will not explicitly get the cofactor expansion formula. However, it will be equivalent.

Definition 3.3.1. The determinant of a linear transformation is the unique function satisfying *normalization*, *alternating*, and *multi-linearity.*

Example 3.3.2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. You may already know that the determinant of a 2×2 matrix is $ad - bc$, but let's actually prove it using these properties.

$$\begin{aligned}
\det(A) &= \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \\
&= \det\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) && \text{(by Multilinearity)} \\
&= \det\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} b \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ d \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ 0 \end{pmatrix}\right) + \det\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} b \\ 0 \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} 0 \\ d \end{pmatrix}\right) \\
& && \text{(by Multilinearity)} \\
&= \det\left(\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}\right) + \det\left(\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right) + \det\left(\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix}\right) + \det\left(\begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix}\right)
\end{aligned}$$

Let's figure what the determinants of each of these matrices are. We want to figure out how to reach each of them from the identity since we know the determinant of the identity and we know what specific operations do to the determinant.

For the first and fourth determinants, the vectors are clearly linearly dependent thus they are equal to 0. Now let's analyze the second determinant. Starting from I_2, I can scale the first column by a to get $\begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}$ which has determinant a by multi-linearity. I can now scale the second column by d to get $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$ with determinant ad again by multi-linearity. We can do the same for the third determinant except we have to do a column swap at the end so intead of having determinant bc we get determinant $-bc$ by antisymmetry. So by adding our results, we get that the determinant of an arbitrary 2×2 matrix is $ad - bc$.

Next, we will partially do the same for 3×3 matrices and then generalize it for $n \times n$ matrices.

Example 3.3.3. Let $A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$ and denote the columns $\vec{c}_1, \vec{c}_2, \vec{c}_3$ respectively. We let $\vec{e}_i$ denote the ith elementary basis vector.

$$\begin{aligned}
\det(A) &= \det(\vec{c}_1, \vec{c}_2, \vec{c}_3) \\
&= \det(\sum_{i=1}^{3} a_{1i}\vec{e}_i, \vec{c}_2, \vec{c}_3) \\
&= \sum_{i=1}^{3} a_{1i} \det(\vec{e}_i, \vec{c}_2, \vec{c}_3) && \text{(by Multilinearity)} \\
&= \sum_{i=1}^{3} a_{1i} \det(\vec{e}_i, \sum_{j=1}^{3} a_{2j}\vec{e}_j, \vec{c}_3) \\
&= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{1i} \det(\vec{e}_i, a_{2j}\vec{e}_j, \vec{c}_3) && \text{(by Multilinearity)} \\
&= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{2j} \det(\vec{e}_i, \vec{e}_j, \sum_{k=1}^{3} a_{3k}\vec{e_k}) \\
&= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} a_{1i}a_{2j}a_{3k} \det(\vec{e}_i, \vec{e}_j, \vec{e_k}) && \text{(by Multilinearity)}
\end{aligned}$$

Since our matrix is only 3×3, this is not too hard to brute force. However, we can use some of the facts we've proved so far to simplify this. Notice that if any of i, j, k are equal then $\det(\vec{e}_i, \vec{e}_j, \vec{e_k})$ will contain at least two columns that are the same. Since the determinant is *alternating*, this implies that the determinant is zero for the terms where any of i, j, k are equal.

If we ignore the cases where any of i, j, k are equal, we are actually left with the permutations of $[3] = \{1, 2, 3\}$. This is because the remaining combinations all have 3 distinct elements and there are only three choices possible namely $1, 2, 3$, so we end up only looking at permutations.

Now notice that if we are only looking at permutations of [3] then $[\vec{e}_i, \vec{e}_j, \vec{e_k}]$ is just I_3 but with swapped columns. So $\det(\vec{e}_i, \vec{e}_j, \vec{e_k}) = \pm 1$ by *antisymmetry* when we are only looking at permutations.

This is enough work with 3×3 matrices to motivate what we will do for $n \times n$ matrices. We first notice that the algebra performed above will still work for arbitrary $n \times n$ matrices. If you don't believe this, you can prove it by induction. But we immediately run into a syntactic issue. We do not want to write n different summations for each of the n coordinates so how can we clean up the notation?

Definition 3.3.4. Let $n \in \mathbb{N}$. A permutation on $[n] = \{1, 2, \dots, n\}$ is a bijective function $\sigma : [n] \to [n]$. We denote the set of all permutations on $[n]$ as S_n.

Remark 3.3.5. You are probably used to thinking about permutations as a rearrangement of the elements of $[n]$. This is just a more mathematical way of writing it. For example, the permutation $[1, 4, 3, 2, 5]$ is the function $\sigma : [5] \to [5]$ defined as:

$$\sigma(1) = 1$$
$$\sigma(2) = 4$$
$$\sigma(3) = 3$$
$$\sigma(4) = 2$$
$$\sigma(5) = 5$$

Using this notation and the same algebra we did in the 3×3 matrix case, we can write that for an arbitary $n \times n$ matrix A:

$$\det(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)} \det(\vec{e}_{\sigma(1)}, \vec{e}_{\sigma(2)}, \dots, \vec{e}_{\sigma(n)})$$

Although this formula now looks nicer, it is still not easy to evaluate. We still have this same issue where we don't know the sign of $\det(\vec{e}_{\sigma(1)}, \vec{e}_{\sigma(2)}, \dots, \vec{e}_{\sigma(n)})$. So our next quest is to figure out, given a permutation $\sigma : [n] \to [n]$, what is the sign of $\det(\vec{e}_{\sigma(1)}, \vec{e}_{\sigma(2)}, \dots, \vec{e}_{\sigma(n)})$?

Before we begin doing this, I think it is necessary to say that we are about to discuss permutations which appear highly unrelated to Linear Algebra. However, it is very important to be able to understand determinants not only as a "dilation" of a vector space but also as the sum across all possible permutations. So the next couple of pages will appear highly unrelated but we will tie it back to Linear Algebra when we get a cleaner formula for the determinant.

Definition 3.3.6. A permutation $\sigma : [n] \to [n]$ is a transposition if it swaps exactly two elements and fixes the rest.

Remark 3.3.5 gives an example of a transposition. The reason we define this special type of permutation is that all permutations can be generated by some number of transpositions. So if we let τ_{ij} represent the transposition that swaps entry i and entry j, we can write $\sigma = \tau_{i_1 j_1} \tau_{i_2 j_2} \dots \tau_{i_n j_n}$ for some number of transpositions. This should seem kind of believable but let's prove it just to be sure.

Theorem 3.3.7. For $n \geq 2$, every $\sigma \in S_n$ can be written as a sequence of transpositions.

Proof. We proceed by induction on n. If $n = 2$, the only possible permutations are already transpositions. Now fix $n \in \mathbb{N}$ and assume that for every $\sigma \in S_n$ we can write it as a sequence of transpositions. Now consider $\sigma \in S_{n+1}$. If $\sigma(n+1) = n+1$, then σ is a permutation of $[n]$ by ignoring $n+1$. So we can apply our Induction Hypothesis and be done. If $\sigma(n+1) \neq n+1$, then let $i \in [n]$ such that $\sigma(i) = n+1$. Then consider $\tau_{i(n+1)}\sigma$. This permutation has the property that $\tau_{i(n+1)}\sigma(n+1) = n+1$ so just as before apply the Induction Hypothesis and be done. $\square$

Remark 3.3.8. The previous theorem gives us a way of interpreting permutations as a sequence of transpositions. Notice that there are infinitely many ways to write a permutation as a sequence of transpositions. Say we have $\sigma \in S_n$ for $n \geq 2$ and we write $\sigma = \tau_1\tau_2 \dots \tau_n$. I can generate a new sequence of transpositions by adding $\tau_{12}\tau_{21}$ an arbitrary number of times.

Definition 3.3.9. The sign of a permutation σ, denoted $\mathrm{sgn}(\sigma)$, is 1 if it takes an even amount of transpositions to generate σ and -1 otherwise.

This definition appears nice and all but it turns out there is something we do need to check.

3.4 More Determinant Properties

Lemma 3.4.1. Let A be an $n \times n$ matrix and E be a $n \times n$ elementary matrix then $\det(AE) = \det(A)\det(E)$

Proof. There are three types of elementary matrices. We can simply check that this holds in every case. □

Theorem 3.4.2. Let A be an $n \times n$ matrix and $E_1, E_2, \ldots, E_k$ be $n \times n$ elementary matrices. Then $\det(AE_1E_2 \ldots E_k) = \det(A)\det(E_1)\det(E_2)\ldots\det(E_k)$.

Proof. Induction on previous Lemma. □

Theorem 3.4.3. Let A be a $n \times n$ matrix. A is invertible if and only if $\det(A) \neq 0$.

Proof. Since A is invertible, $A = E_kE_{k-1}\ldots E_1I$. We know the determinants of the elementary matrices and none of them are zero. Thus, $\det(A) = \det(E_kE_{k-1}\ldots E_1)$ and by the above lemma and induction we get that $\det(E_kE_{k-1}\ldots E_1) = \det(E_k)\det(E_{k-1})\ldots\det(E_1) \neq 0$.

Assume $\det(A) \neq 0$. □

Lemma 3.4.4. Let A and B be $n \times n$ matrices. If either A is not invertible or B is not invertible, then AB is not invertible.

Theorem 3.4.5. Let A and B be $n \times n$ matrices then $det(AB) = det(A)det(B)$.

Proof. If either A is not invertible or B is not invertible then by Lemma 3.1.8, AB is not invertible. Thus, by Theorem 3.1.7, $det(AB) = det(A)det(B) = 0$.

Otherwise, A and B are both invertible. Let $A = E_k \ldots E_1I$ and $B = F_j \ldots F_1I$ where E and F are elementary matrices.

$$\begin{aligned}
det(AB) &= det(E_k \ldots E_1F_j \ldots F_1) \\
&= det(E_k \ldots E_1F_j \ldots F_2)det(F_1) && \text{(by Lemma 18.6)} \\
&= det(E_k \ldots E_1F_j \ldots F_3)det(F_2)det(F_1) && \text{(by Lemma 18.6)} \\
&\vdots \\
&= det(E_k)\ldots det(E_1)det(F_j)\ldots det(F_1) \\
&= det(E_k \ldots E_1)det(F_j \ldots F_1) \\
&= det(A)det(B)
\end{aligned}$$

□

3.5 Exercises 6

Exercise 3.5.1. Prove that the determinant of a lower/upper triangular matrix is the product of the entries along its main diagonal.

Exercise 3.5.2. Let A be a $n \times n$ invertible matrix. Prove that $\det(A) = \frac{1}{\det(A^{-1})}$

Exercise 3.5.3. Let A be an $n \times n$ matrix with real valued entries and $k \in \mathbb{R}$. Prove that $\det(kA) = k^n \det(A)$.

Exercise 3.5.4. Compute the determinant of $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

Exercise 3.5.5. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$. Find a general formula for the volume of the parallelepiped formed by them.

Chapter 4

Inner Products

4.1 Inner Products

Definition 4.1.1. Let V be a vector space over $\mathbb{R}$. A valid inner product of V satisfies the following four properties:

1. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$ for all $\vec{v}, \vec{u} \in V$ (Symmetry)
2. $\langle a\vec{x} + b\vec{y}, \vec{z} \rangle = a\langle \vec{x}, \vec{z} \rangle + b\langle \vec{y}, \vec{z} \rangle$ for all $a, b \in \mathbb{R}$ and $\vec{x}, \vec{y}, \vec{z} \in V$ (Linearity)
3. $\langle \vec{v}, \vec{v} \rangle \geq 0$ for all $\vec{v} \in V$ (Non-negativity)
4. $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$ (Non-degeneracy)

Definition 4.1.2. An inner product space is a vector space V along with an inner product that satisfies the aforementioned conditions.

Example 4.1.3. Consider $\mathbb{R}^n$. Let $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$. One example of an inner product is the dot product. This is often notated and defined as $\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \sum_{i=1}^n x_i y_i$.

Let's check the four conditions of inner products. Symmetry and Linearity follow from the fact that addition and multiplication are commutative. Let $\vec{v} \in V$. We have that $\vec{v} \cdot \vec{v} = \sum_{i=1}^n v_i v_i = \sum_{i=1}^n v_i^2$. This satisfies, non-negativity since squares are always non-negative. If $\vec{v} = \vec{0}$, each $v_i = 0$ so the whole summation equals 0. Assume $\langle \vec{v}, \vec{v} \rangle = \vec{v} \cdot \vec{v} = \sum_{i=1}^n v_i^2 = 0$. Since $v_i^2 \geq 0$, we must have that $v_i = 0$ for all i since otherwise we would have a positive sum. So we have non-degeneracy as well.

Definition 4.1.4. Let V be an inner product space and $\vec{v} \in V$. We define the norm of $\vec{v}$ to be $||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$.

Remark 4.1.5. There is a concept of normed spaces which are defined similarly to inner product spaces but with norms instead. In general, norms are a function from a vector to $\mathbb{R}$ that satisfy another set of properties similar to the inner product. However, for this context, we define the norm as above.

Example 4.1.6. Consider $\mathbb{R}^n$ with the dot product. Let $\vec{v} \in \mathbb{R}^n$. We have that $||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$.

Theorem 4.1.7. Let V be an inner product space and $\vec{v}, \vec{w} \in V$. The following inequality always holds:

$$|\langle \vec{v}, \vec{w} \rangle| \leq ||\vec{v}||||\vec{w}||$$

This is known as the Cauchy-Schwarz Inequality.

Theorem 4.1.8. Let V be an inner product space and $\vec{v}, \vec{w} \in V$. The following inequality always holds:

$$||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||$$

This is known as the Triangle Inequality.

Proof.

$$\begin{aligned}
||\vec{v} + \vec{w}||^2 &= (\sqrt{\langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle})^2 \\
&= \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle \\
&= \langle \vec{v}, \vec{v} \rangle + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle + \langle \vec{w}, \vec{w} \rangle && \text{(by Linearity of Inner Product)} \\
&= ||\vec{v}||^2 + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle + ||\vec{w}||^2 \\
&= ||\vec{v}||^2 + \langle \vec{v}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle + ||\vec{w}||^2 && \text{(by Symmetry of Inner Product)} \\
&\leq ||\vec{v}||^2 + |\langle \vec{v}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle| + ||\vec{w}||^2 \\
&\leq ||\vec{v}||^2 + |\langle \vec{v}, \vec{w} \rangle| + |\langle \vec{v}, \vec{w} \rangle| + ||\vec{w}||^2 \\
&\leq ||\vec{v}||^2 + ||\vec{v}||||\vec{w}|| + ||\vec{v}||||\vec{w}|| + ||\vec{w}||^2 && \text{(by Cauchy Schwarz Inequality)} \\
&\leq ||\vec{v}||^2 + 2||\vec{v}||||\vec{w}|| + ||\vec{w}||^2
\end{aligned}$$

Take the square root of both sides to get the Triangle Inequality. □

Definition 4.1.9. Let V be an inner product space and $\vec{v}, \vec{w} \in V$. We say that $\vec{v}$ and $\vec{w}$ are orthogonal, denoted $\vec{v} \perp \vec{w}$, if $\langle \vec{v}, \vec{w} \rangle = 0$.

Definition 4.1.10. Let $\mathcal{L} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ be a list of vectors. We say that $\mathcal{L}$ is an orthogonal list if $\vec{v}_i \perp \vec{v}_j$ for all $i, j \in [n]$.

Theorem 4.1.11. Let $\mathcal{L}$ be an orthogonal list such that $\vec{0} \notin \mathcal{L}$. Then $\mathcal{L}$ is linearly independent.

Definition 4.1.12. Let $\mathcal{L} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ be a list of vectors. We say that $\mathcal{L}$ is an orthonormal list if $\mathcal{L}$ is an orthogonal list and $||\vec{v}_i|| = 1$ for $1 \leq i \leq n$.

Definition 4.1.13. Let $\mathcal{L} = (\vec{v}_1, \ldots, \vec{v}_k)$ be an orthonormal list. We define the projection of a $\vec{u}$ onto $\mathcal{L}$ as $proj_{\mathcal{L}}(\vec{u}) = \sum_{i=1}^{k} \langle \vec{u}, \vec{v}_i \rangle \vec{v}_i$

Theorem 4.1.14. Let $\mathcal{L} = (\vec{v}_1, \ldots, \vec{v}_k)$ be an orthonormal list. Then $\vec{x} - proj_{\mathcal{L}}(\vec{x}) \perp \vec{v}_i$ for all $1 \leq i \leq k$.

Proof.

$$\begin{aligned}
\langle \vec{x} - proj_{\mathcal{L}}(\vec{x}), \vec{v}_i \rangle &= \langle \vec{x}, \vec{v}_i \rangle - \langle proj_{\mathcal{L}}(\vec{x}), \vec{v}_i \rangle && \text{(by Linearity)} \\
&= \langle \vec{x}, \vec{v}_i \rangle - \langle \sum_{j=1}^{k} \langle \vec{x}, \vec{v}_j \rangle \vec{v}_j, \vec{v}_i \rangle \\
&= \langle \vec{x}, \vec{v}_i \rangle - \sum_{j=1}^{k} \langle \vec{x}, \vec{v}_j \rangle \langle \vec{v}_j, \vec{v}_i \rangle && \text{(by Linearity)}
\end{aligned}$$

First, consider the case when $j = i$. We observe that $\langle \vec{v}_i, \vec{v}_i \rangle = ||\vec{v}_i||^2$ and since $\mathcal{L}$ is orthonormal we know that $||\vec{v}_i|| = 1$. Thus, $\langle \vec{v}_i, \vec{v}_i \rangle = 1$. When $i \neq j$, then $\langle \vec{v}_j, \vec{v}_i \rangle = 0$ since $\mathcal{L}$ is orthogonal.

$$\langle \vec{x}, \vec{v}_i \rangle - \sum_{j=1}^{k} \langle \vec{x}, \vec{v}_j \rangle \langle \vec{v}_j, \vec{v}_i \rangle = \langle \vec{x}, \vec{v}_i \rangle - \langle \vec{x}, \vec{v}_i \rangle = 0$$

□

Definition 4.1.15. The Gram-Schmidt Process is a way of finding an orthonormal list $\mathcal{O}$ given a list of vectors $\mathcal{L}$ such that $span(\mathcal{O}) = span(\mathcal{L})$.

Let $\mathcal{O}_i$ denote the list $\mathcal{O}$ at step i. Let $\mathcal{L}_i$ denote the list of the first i vectors of $\mathcal{L}$. We initialize $\mathcal{O}_0 = \emptyset$. Note that initially, $span(\mathcal{O}_0) = span(\mathcal{L}_0)$ and $\mathcal{O}_{i-1}$ is orthonormal.

At step i, look at the ith vector in $\mathcal{L}$ denoted $\vec{v}_i$. We assume that up to this point, we have constructed $\mathcal{O}_{i-1}$ such that $span(\mathcal{O}_{i-1}) = span(\mathcal{L}_{i-1})$.

Case 1: If $\vec{v}_i \in span(\mathcal{O}_{i-1})$, do nothing and move on to the $i+1$th vector.

Case 2: If $\vec{v}_i \notin span(\mathcal{O}_{i-1})$, then we want to add it to make $\mathcal{O}_{i-1}$ into an orthonormal list with a larger span. To ensure the new vector is orthogonal to $\mathcal{O}_{i-1}$, we will use the projection. We will add $\frac{\vec{v}_i - proj_{\mathcal{O}_{i-1}}(\vec{v}_i)}{||\vec{v}_i - proj_{\mathcal{O}_{i-1}}(\vec{v}_i)||}$. The numerator is simply a vector that is orthogonal to $\mathcal{O}_{i-1}$ as we proved with projections. We divide by its norm to make this vector into a unit vector.

Remark 4.1.16. The Gram-Schmidt Process works for arbitrary inner products. However, it is most commonly done with the dot product since this preserves the fact that orthogonal is equivalent to the vectors being perpendicular which is not true for arbitrary inner products.

4.2 Exercises 7

Exercise 4.2.1. Let $\vec{x}, \vec{y} \in \mathbb{R}^3$. Prove or disprove that $\langle \vec{x}, \vec{y} \rangle = 2x_1y_1 + 3x_2y_2 + 5x_3y_3$ where $\vec{x} = (x_1, x_2, x_3)$ and $\vec{y} = (y_1, y_2, y_3)$ is an inner product

Exercise 4.2.2. Prove that the Cauchy-Schwarz Inequality has equality if and only if one of the vectors is a scalar of the other.

Exercise 4.2.3. Perform the Gram-Schmidt Process on the following vectors using the dot product:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Exercise 4.2.4. Perform the Gram-Schmidt Process on the following vectors using the dot product:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad \vec{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Chapter 5

Elementary Spectral Theory

5.1 Eigenvalues and Eigenvectors

Here we will only consider linear transformations $T_A : V \to V$ via $T_A(\vec{x}) = A\vec{x}$ where V is an arbitrary vector space over some field $\mathbb{F}$. In general, it is most common for this vector space to be $\mathbb{C}^n$ and we'll discuss the algebraic motivation behind this shortly. The motivation behind this chapter is the fact that often times we need to repeatedly apply linear transformations. This means that we need to compute $A^n\vec{x}$, but this is an expensive computation since matrix multiplication is not trivial to compute. In this chapter, we will analyze when we have "nicer" ways to compute $A^n\vec{x}$ and to what extent we can determine when we are in these scenarios.

Definition 5.1.1. Let A be a $n \times n$ matrix. Assume that we have $A\vec{x} = \lambda\vec{x}$ for $\vec{x} \in V$ and $\lambda \in \mathbb{F}$. We call λ an eigenvalue of A and $\vec{x}$ a corresponding eigenvector. $\vec{0}$ is not considered an eigenvector.

Let's analyze when we have a solution for $A\vec{x} = \lambda\vec{x}$.

$$\begin{aligned} A\vec{x} &= \lambda\vec{x} \\ A\vec{x} - \lambda\vec{x} &= \vec{0} \\ A\vec{x} - \lambda I\vec{x} &= \vec{0} \\ (A - \lambda I)\vec{x} &= \vec{0} \end{aligned}$$

If $A - \lambda I$ is invertible, then this equation has exactly one solution: the zero vector. We require that eigenvectors are non-zero so we care about the solutions when $A - \lambda I$ is not invertible. As we know, this is equation has nontrivial solutions exactly when $\det(A - \lambda I) = 0$.

Example 5.1.2. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. We have that $A - \lambda I = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}$ and $\det(A - \lambda I) =$

$(2-\lambda)^2 - 1$.

$$\begin{aligned}
(2-\lambda)^2 - 1 &= 0 \\
4 - 4\lambda + \lambda^2 - 1 &= 0 \\
\lambda^2 - 4\lambda + 3 &= 0 \\
(\lambda - 3)(\lambda - 1) &= 0
\end{aligned}$$

This tells us that 1 and 3 are eigenvalues for A. Now, let's find their corresponding eigenvectors. First, let $\lambda = 1$.

$$\begin{aligned}
A\vec{x} &= \lambda\vec{x} \\
A\vec{x} &= \vec{x} \\
\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\end{aligned}$$

This gives us the equations $2x_1 + x_2 = x_1$ and $x_1 + 2x_2 = x_2$. These tell us that $x_1 = -x_2$. Moreover, any vector that satisfies $\begin{pmatrix} a \\ -a \end{pmatrix}$ is an eigenvector of A with eigenvalue 1.

Now let $\lambda = 3$. We perform a similar analysis.

$$\begin{aligned}
A\vec{x} &= \lambda\vec{x} \\
A\vec{x} &= 3\vec{x} \\
\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 3x_1 \\ 3x_2 \end{bmatrix}
\end{aligned}$$

This gives us the equations $2x_1 + x_2 = 3x_1$ and $x_1 + 2x_2 = 3x_2$. Solving this system gives us that $x_1 = x_2$. So the eigenvectors with eigenvalue 3 are any vector that satisfy $\begin{pmatrix} a \\ a \end{pmatrix}$.

The previous example actually tells us something more about eigenvectors. We notice that when we solved the system, we got eigenvectors of a certain form not just one specific vector. In fact, once you have one eigenvector, you can generate many more by scaling.

Example 5.1.3. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. We can compute that $\det(A - \lambda I) = \lambda^2(1-\lambda)$ which has roots 1 and 0. Let's consider the case where $\lambda = 0$. We want to solve $A\vec{x} = 0\vec{x} = \vec{0}$.

$$A\vec{x} = \vec{0}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} = \vec{0}$$

This implies that as long as the first coordinate is zero, the other coordinates can be anything. In particular, $(0, 1, 0)$ and $(0, 0, 1)$ are both eigenvectors and every vector in their span is also an eigenvector.

The case where $\lambda = 1$ is uninteresting so we omit it.

From the previous examples, we can see that the term eigenvector is not unique. A singular eigenvalue can have many, often infinitely many eigenvectors. This leads to the following definition:

Definition 5.1.4. Let λ be an eigenvalue of a matrix A. The eigenspace of λ is the space spanned by the corresponding eigenvectors. The dimension of the eigenspace for an eigenvalue λ is known as the geometric multiplicity of λ.

Remark 5.1.5. So in the previous example, the eigenspace of 0 would be $span((0, 1, 0), (1, 0, 0))$

Remark 5.1.6. We discussed above how eigenvalues are exactly the solutions to the equation $(A - \lambda I)\vec{x} = \vec{0}$. This implies that the eigenspace is equivalent to the kernel of the transformation $A - \lambda I$.

Definition 5.1.7. When finding eigenvalues, we call the equation $\det(A - \lambda I)$ the characteristic polynomial.

Example 5.1.8. Let $T : \mathbb{P}^n \to \mathbb{P}^n$ be the derivative operator. What are the eigenvalues of T?

First, we notice that if the input has degree $k \geq 1$, then the output has degree $k - 1$. There is no way to scale a degree $k - 1$ polynomial by a constant to get a degree k polynomial thus there are no eigenvalues for polynomials of degree $k \geq 1$.

This leaves the case when $k < 1$. When $k < 1$, our polynomial is simply a constant. The derivative of a constant is just 0. Thus, this tells us that T has eigenvalue 0 which has eigenvectors that are any constant.

Before further discussion, we need to briefly discuss some algebra as solving the characteristic polynomial is the key component of finding eigenvalues. It is common to work in $\mathbb{C}^n$ when discussing eigenvalues because polynomials always have roots in the complex numbers. In particular, a polynomial of degree n always has n roots (counting multiplicities) over the complex numbers. Counting multiplicities means that if a polynomial is divisible by $(x - 3)^5$, we count 3 as a root 5 times.

Definition 5.1.9. The algebraic multiplicity of an eigenvalue λ for a matrix A is its degree in the characteristic polynomial of A.

This is important because when we analyze the characteristic polynomial of an $n \times n$ matrix A, n λ terms along the main diagonal of $A - \lambda I$. When we compute $\det(A - \lambda I)$, we will end up with a degree n polynomial whose roots are our eigenvalues.

Theorem 5.1.10. Every single variable polynomial with complex coefficients and degree at least one has at least one complex root. This is the *Fundamental Theorem of Algebra.*

Theorem 5.1.11. If $p(z)$ is a polynomial with complex coefficients and we have that $p(c) = 0$, then $(z - c)$ divides $p(z)$.

We will use these as facts without proof as their proofs are highly unrelated to linear algebra. These two facts alone give us a lot of important takeaways.

Corollary 5.1.12. Every square matrix A with complex values has at least one eigenvalue.

Proof. If A is an $n \times n$ matrix, $\det(A - \lambda I)$ will be a degree n polynomial. By the Fundamental Theorem of Algebra, this has at least one complex root. We know that this root is an eigenvalue so A has at least one eigenvalue. $\square$

Corollary 5.1.13. Every degree $n \geq 1$ polynomial has n roots counting multiplicities.

Proof. Let $p(z)$ be our polynomial of degree $n \geq 1$. We induct on n. If $n = 1$, then the Fundamental Theorem of Algebra gives us what we want. For $n > 1$, the Fundamental Theorem of Algebra gives us one root c. Theorem 5.1.8 tells us that we can divide $p(z)$ by this $z - c$ to get a degree $n - 1$ polynomial. Induction now concludes the proof. $\square$

5.2 Diagonalization

We previously discussed taking large powers of A. This can be very expensive especially if A is high dimensional. However, if we have the special case where A is diagonal, this computation is really easy since:

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} a^2 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & c^2 \end{bmatrix}$$

Moreover, by induction we can see that:

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}^n = \begin{bmatrix} a^n & 0 & 0 \\ 0 & b^n & 0 \\ 0 & 0 & c^n \end{bmatrix}$$

Definition 5.2.1. $T : V \to V$ is diagonalizable if there exists a basis B for V consisting of eigenvectors of T. Let $B = (\vec{v}_1, \ldots, \vec{v}_n)$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ then:

$$[T]_{BB} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

Let's consider where this matrix comes from. We know that the first column of the matrix is equal to $[T(\vec{v}_1)]_B$. We also know that since $\vec{v}_1$ is an eigenvector, that $T(\vec{v}_1) = \lambda_1\vec{v}_1$. This gives us that $T(\vec{v}_1) = \lambda_1\vec{v}_1 + 0\vec{v}_2 + 0\vec{v}_3 + \ldots + 0\vec{v}_n$. This gives us the coordinate vector shown in the first column of the above matrix. You can apply similar logic to the rest of the columns.

Now, let's talk about how do we find this diagonal matrix. Based on the definition, we could find every single eigenvector and then check if they form a basis. If not, then T is not diagonalizable. Otherwise, let's look at how we get $[T]_{BB}$. If we want, $[T]_{BB}$ we need to use change of basis matrices:

$$[T]_{BB} = [I]_{BE}[T]_{EE}[I]_{EB}$$

Note that E represents the elementary basis and as long as you know the transformation with respect to some basis you can do this. Initially, we only know $[T]_{EE}$.

$$\begin{aligned}
[T]_{BB} &= [I]_{BE}[T]_{EE}[I]_{EB} \\
[I]^{-1}_{BE}[T]_{BB} &= [I]^{-1}_{BE}[I]_{BE}[T]_{EE}[I]_{EB} \\
[I]^{-1}_{BE}[T]_{BB} &= [T]_{EE}[I]_{EB} \\
[I]^{-1}_{BE}[T]_{BB}[I]^{-1}_{EB} &= [T]_{EE}[I]_{EB}[I]^{-1}_{EB} \\
[I]^{-1}_{BE}[T]_{BB}[I]^{-1}_{EB} &= [T]_{EE} \\
[I]_{EB}[T]_{BB}[I]_{BE} &= [T]_{EE}
\end{aligned}$$

A common way this is notated is $P^{-1}DP = A$. Let's talk about why this achieves the efficient power multiplication that want.

$$\begin{aligned}
([T]_{EE})^2 &= ([I]_{EB}[T]_{BB}[I]_{BE})^2 \\
&= ([I]_{EB}[T]_{BB}[I]_{BE})([I]_{EB}[T]_{BB}[I]_{BE}) \\
&= ([I]_{EB}[T]_{BB}[T]_{BB}[I]_{BE}) \\
&= ([I]_{EB}[T]^2_{BB}[I]_{BE})
\end{aligned}$$

By induction, this gives us:

$$([T]_{EE})^n = [I]_{EB}[T]^n_{BB}[I]_{BE}$$

Since $[T]_{BB}$ is diagonal, computing $[T]^n_{BB}$ is simple. Afterwards, we only need to perform two matrix multiplications. This is way more efficient than trying to directly compute $[T]^n_{EE}$ which would require n difficult matrix multiplications.

Chapter 6

Applications

6.1 Markov Chains

6.2 Multidimensional Calculus

6.3 Machine Learning

6.4 Quantum Mechanics

References

- *Linear Algebra Done Wrong* by Sergei Treil
- *Linear Algebra Done Right* by Sheldon Axler
- Figures generated by ChatGPT-4o