MATH 2090 - Linear Algebra 2

Course Notes, University of Manitoba

Tyrone Ghaswala

Summer 2018

Contents

1 Vector Spaces **4**

2 Subspaces **7**

3 Bases and Dimension **9**
3.1 Linear Indepence, Spanning Sets, and Bases . . . 9
3.2 Dimension . . . 11
3.3 Obtaining Bases . . . 13
3.4 Coordinates with Respect to a Basis . . . 13

4 Linear Maps **15**
4.1 Linear Maps as Matrices . . . 19

5 Isomorphisms of Vector Spaces **25**

6 Review - Determinants and matrix inverses **29**
6.1 Cofactor Expansion . . . 29
6.2 Elementary Row Operations and the Determinant . . . 30
6.3 Invertibility . . . 31
6.4 Determinants and Volume . . . 32

7 Eigenvectors and Diagonalization **32**
7.1 Finding Eigenvectors and Eigenvalues . . . 34
7.2 Diagonalisation . . . 38
7.3 Applications of Diagonalisation . . . 45

8 Inner Product Spaces **48**
8.1 Orthogonality . . . 53
8.2 Orthonormal Bases . . . 55
8.3 Projections . . . 57
8.4 Orthogonal Complements . . . 59
8.5 The Gram-Schmidt Orthogonalisation Procedure . . . 60
8.6 Application of projections: Method of Least Squares . . . 62

9 Hermitian, Unitary and Normal Matrices **65**
9.1 Spectral Theorem for Hermitian Matrices . . . 69
9.2 Inner Products and Hermitian Matrices . . . 74
9.3 Normal Matrices . . . 77

10 To infinity and beyond! **78**

These notes are the official text for the Summer 2018 offering of MATH 2090 - Linear Algebra 2. They will be updated as I go, and are definitely not free of typos and mistakes. If you find any, please let me know about it and I'll fix them as soon as possible.

For this course, $\mathbb{F}$ will denote the complex numbers $\mathbb{C}$ or the real numbers $\mathbb{R}$.

Lecture 1 - July 3

1 Vector Spaces

Linear algebra is the study of vector spaces. Before we formally define a vector space, let's introduce some examples of vector spaces. As you go through each example, pay close attention to the similarities between each example.

- The vector space $\mathbb{R}^n$ is given by

$$\mathbb{R}^n := \left\{ \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} : a_i \in \mathbb{R} \text{ for all } i \right\}.$$

  Addition and scalar multiplication are given by

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix} \quad \text{and} \quad \alpha \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \alpha a_1 \\ \vdots \\ \alpha a_n \end{pmatrix}.$$

  The intuitive picture that is helpful to have in mind are the cases of $\mathbb{R}^2$ and $\mathbb{R}^3$ that you are familiar with from previous courses. You can picture $\mathbb{R}^2$ as the Cartesian plane, and $\mathbb{R}^3$ as 3-dimensional space. In both of these vector spaces, you know how vector addition and scalar multiplication work, and intuitively, it's the same for $\mathbb{R}^n$. Although $\mathbb{R}^n$ is an n-dimensional vector space (we will define dimension later on in the course), it is usually helpful to think of $\mathbb{R}^2$ and $\mathbb{R}^3$.

- The vector space $\mathcal{P}_n(\mathbb{R})$ is the set of polynomials of degree at most n with real coefficients. That is

$$\mathcal{P}_n(\mathbb{R}) := \{a_n x^n + \cdots + a_1 x + a_0 : a_i \in \mathbb{R} \text{ for all } i\}$$

  with addition and scalar multiplication defined by

$$(a_n x^n + \cdots + a_0) + (b_n x^n + \cdots + b_0) = (a_n + b_n)x^n + \cdots + (a_0 + b_0)$$

  and

$$\alpha(a_n x^n + \cdots + a_0) = (\alpha a_n x^n + \cdots + \alpha a_0)$$

  respectively. So, for example, $1 + 2x - 3x^2 \in \mathcal{P}_2(\mathbb{R})$. Also,

$$(4 + 7x) + (1 + x^2) = 5 + 7x + x^2 \quad \text{and} \quad 25(1 + 2x^3) = 25 + 50x^3.$$

  You may be used to thinking of polynomials as functions. In the context of this course, don't! Although it is sometimes useful to evaluate a polynomial at a certain number, in this course, polynomials are not functions. They are simply objects which you can add together and multiply by scalars.

- The vector space of n by m matrices with real coefficients is given by

$$M_{n\times m}(\mathbb{R}) := \left\{ \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} : a_{ij} \in \mathbb{R} \text{ for all } i, j \right\}.$$

  Addition and scalar multiplication are given by matrix addition and scalar multiplication of matrices as usual. So, for example, in $M_{2\times 2}(\mathbb{R})$,

$$\begin{bmatrix} 2 & 5 \\ 7 & \pi \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 8 & 1+\pi \end{bmatrix} \quad \text{and} \quad \sqrt{2}\begin{bmatrix} 2 & 5 \\ 7 & \pi \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 5\sqrt{2} \\ 7\sqrt{2} & \pi\sqrt{2} \end{bmatrix}.$$

- The vector space of continuous functions is denoted by

$$\mathcal{C}([0,1]) := \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}.$$

  Addition and scalar multiplication are defined by

$$(f+g)(x) := f(x) + g(x) \quad \text{and} \quad (\alpha f)(x) := \alpha(f(x)).$$

  This one is a little trickier. Here, the vectors are continuous functions from $\mathbb{R}$ to $\mathbb{R}$. So when we add two functions, we get another function, and when we multiply a function by a scalar, we get another function. For example,

$$(\sin(x)) + (\cos(x)) = \sin x + \cos x \quad \text{and} \quad 3(\sin(x)) = 3\sin(x).$$

- Here's a slightly more interesting one. Let $\mathbb{V}$ be the set of all lines in $\mathbb{R}^2$ with slope 1. Each line has equation $y = x + d$. Addition and scalar multiplication in $\mathbb{V}$ is defined by

$$(y = x + d_1) + (y = x + d_2) := (y = x + (d_1 + d_2)) \quad \text{and} \quad \alpha(y = x + d) := (y = x + \alpha d).$$

- Vector spaces can also have the complex numbers as the scalars, instead of the real numbers. The complex 2-dimension vector space is given by

$$\mathbb{C}^2 := \left\{ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} : a_1, a_2 \in \mathbb{C} \right\}$$

  with addition and scalar multiplication exactly the same as in $\mathbb{R}^2$, except all the numbers in sight are in $\mathbb{C}$. We define $\mathbb{C}^n$ in the same way, and call this an n-dimensional complex vector space.

Lecture 2 - July 4

Now that we have seen a bunch of example of vector spaces, you might have a perfectly reasonable question in mind: What is a vector space? Well, before we give the formal definition, let's take a look at the similarities between all of these examples. They all come with a set of 'vectors' (even though sometimes these vectors can look a little wierd, like a straight line in $\mathbb{R}^2$ of slope 1) and some scalars ($\mathbb{R}$ or $\mathbb{C}$). Furthermore, there is a way to 'add' two vectors to get another vector, and to 'multiply' a vector by a scalar to get another vector. There is some other structure lurking in the background which is perhaps a little harder to notice just from these examples. Indeed, in each vector space there is a special vector (call it $\vec{0}$) with the property that $\vec{0} + \vec{v} = \vec{v}$ for all vectors $\vec{v}$ in the vector space.

Formally, we define a vector space as follows:

Definition. A **vector space** over a field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$ for this course) is a set $\mathbb{V}$ together with operations $+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ (addition) and $\cdot : \mathbb{F} \times \mathbb{V} \to \mathbb{V}$ (scalar multiplication) such that for all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{V}$ and $s, t \in \mathbb{F}$ the following hold.

1. $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$,
2. There exists a vector $\vec{0} \in \mathbb{V}$ such that $\vec{0} + \vec{x} = \vec{x} + \vec{0} = \vec{x}$,
3. For every $\vec{x} \in \mathbb{V}$, there exists a $-\vec{x} \in \mathbb{V}$ such that $\vec{x} + (-\vec{x}) = (-\vec{x}) + \vec{x} = \vec{0}$,
4. $\vec{x} + \vec{y} = \vec{y} + \vec{x}$,
5. $s(t\vec{x}) = (st)\vec{x}$,
6. $(s + t)\vec{x} = s\vec{x} + t\vec{x}$,
7. $s(\vec{x} + \vec{y}) = s\vec{x} + s\vec{y}$, and
8. $1\vec{x} = \vec{x}$.

Example. As an example, let's check that $\mathbb{R}^2$ with the addition and scalar multiplication defined above is a vector space.

To check axiom 1., let $\vec{x} = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$, $\vec{y} = \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)$, and $\vec{z} = \left(\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right)$ be three arbitrary vectors in $\mathbb{R}^2$. Then

$$(\vec{x} + \vec{y}) + \vec{z} = \begin{pmatrix} x_1 + y_2 \\ x_2 + y + 2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 + z_1 \\ x_2 + y_2 + z_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 + z_1 \\ y_2 + z_2 \end{pmatrix} = \vec{x} + (\vec{y} + \vec{z})$$

therefore 1 holds.

The vector $\vec{0} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ satisfies the properties of axiom 2. If $\vec{x} = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$, then $-\vec{x} = \left(\begin{smallmatrix} -x_1 \\ -x_2 \end{smallmatrix}\right)$ satisfies $\vec{x} + (-\vec{x}) = (-\vec{x}) + \vec{x} = \vec{0}$, so 3 holds. For 7, let $s \in \mathbb{R}$ be an arbitrary scalar. Then

$$\begin{aligned} s(\vec{x} + \vec{y}) &= s \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} s(x_1 + y_1) \\ s(x_2 + y_2) \end{pmatrix} \\ &= \begin{pmatrix} sx_1 + sy_1 \\ sx_2 + sy_2 \end{pmatrix} \\ &= \begin{pmatrix} sx_1 \\ sx_2 \end{pmatrix} + \begin{pmatrix} sy_1 \\ sy_2 \end{pmatrix} \\ &= s\vec{x} + s\vec{y}. \end{aligned}$$

We leave the rest of the axioms as an exercise for you to check.

Exercise. Prove that axioms 4, 5, 6, and 8 hold for $\mathbb{R}^2$, implying that $\mathbb{R}^2$ is indeed a vector space.

Here is an example of something that is *not* a vector space.

Example. Let $\mathbb{V}$ be the set of polynomials with coefficients in $\mathbb{C}$ of degree at most 1, with addition of vectors and scalar multiplication (by scalars in $\mathbb{C}$) given by

$$(a_1 + a_2x) + (b_1 + b_1x) = (a_1 + a_2) + (b_1 + b_2)x \quad \text{and} \quad \alpha \bullet (a_1 + a_2x) = \alpha a_2 + \alpha a_1 x$$

respectively. Then $\mathbb{V}$ is *not* a vector space since $1 \bullet (3 + ix) = i + 3x$ so axiom 8 fails.

Fact 1. *Let $\mathbb{V}$ be a vector space over a field $\mathbb{F}$. Then*

- $0\vec{x} = \vec{0}$ *for all* $\vec{x} \in \mathbb{V}$,
- $(-1)\vec{x} = \vec{-x}$ *for all* $\vec{x} \in \mathbb{V}$, *and*
- $t\vec{0} = \vec{0}$ *for all* $t \in \mathbb{F}$.

Exercise. Prove Fact 1.

Lecture 3 - July 5

2 Subspaces

We know that if we consider just the plane spanned by the x and y coordinates in $\mathbb{R}^3$, then we can think of this as $\mathbb{R}^2$ living inside $\mathbb{R}^3$. This is an example of a subspace of $\mathbb{R}^3$. To make this idea precise, we first formally define a subspace.

Definition. Let $\mathbb{V}$ be a vector space and $\mathbb{U} \subset \mathbb{V}$ a subset. We call $\mathbb{U}$ a **subspace** of $\mathbb{V}$ if $\mathbb{U}$, endowed with the addition and scalar multiplication from $\mathbb{V}$, is a vector space.

Example. Consider the subset $\mathbb{U} \subset \mathcal{P}_2(\mathbb{R})$ given by $\mathbb{U} = \{p \in \mathcal{P}_2(\mathbb{R}) : p(2) = 0\}$. First to get a feel for $\mathbb{U}$, note that $x^2 + x - 6 \in \mathbb{U}$ but $x^2 \notin \mathbb{U}$. This is a subspace of $\mathcal{P}_2(\mathbb{R})$, and let's check some of the axioms to convince ourselves.

First we have to check that the addition and scalar multiplication from $\mathcal{P}_2(\mathbb{R})$ makes sense as addition and scalar multiplication in $\mathbb{U}$. That is, we have to make sure that if we take two vectors in $\mathbb{U}$ and add them together, we get a vector in $\mathbb{U}$, and that every scalar multiple of a vector in $\mathbb{U}$ is in $\mathbb{U}$.

Suppose $p, q \in \mathbb{U}$ and $\alpha \in \mathbb{R}$. Then $(p+q)(2) = p(2) + q(2) = 0$ so $p + q \in \mathbb{U}$. Furthermore, $(\alpha p)(2) = \alpha p(2) = 0$ so $\alpha p \in \mathbb{U}$. Alas, addition and scalar multiplication make sense on $\mathbb{U}$.

Since the addition and scalar multiplication on $\mathbb{U}$ is simply that from $\mathcal{P}_2(\mathbb{R})$, and $\mathcal{P}_2(\mathbb{R})$ is a vector space, axioms 1, 4,5,6,7, and 8 hold for $\mathbb{U}$. We see that $\vec{0} = 0x^2 + 0x + 0 \in \mathbb{U}$ so axiom 2 is satisfied. Furthermore, by the second bullet point from Fact 1, $-p = (-1)p \in \mathbb{U}$, so axiom 3 is satisfied. We may finally conclude that $\mathbb{U}$ is a vector space.

Checking that addition and scalar multiplication make sense, followed by checking all 8 axioms is a little cumbersome. However, if you pay attention to what we checked, a lot of things came for free from the fact that $\mathcal{P}_2(\mathbb{R})$ was already a vector space. The next theorem allows us never to have to do that much work again, and simply check three things to check whether or not a subset of a vector space is a subspace or not.

Theorem 2 (The subspace test)**.** *Suppose $\mathbb{U}$ is a subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. The subset $\mathbb{U}$ is a subspace of $\mathbb{V}$ if and only if the following three conditions hold:*

- $\mathbb{U}$ *is non-empty,*
- *For all* $\vec{u}_1, \vec{u}_2 \in \mathbb{U}$, $\vec{u}_1 + \vec{u}_2 \in \mathbb{U}$, *and*
- *For all* $\alpha \in \mathbb{F}$ *and for all* $\vec{u} \in \mathbb{U}$, $\alpha\vec{u} \in \mathbb{U}$.

Proof. If $\mathbb{U}$ is a subspace, then 2 and 3 hold as part of being a definition of a subspace, and since all vector spaces have a zero vector, $\mathbb{U}$ must be non-empty. Conversely, suppose 1, 2, and 3 hold for a subset $\mathbb{U}$ of $\mathbb{V}$. Properties 2 and 3 imply that the addition and scalar multiplication from $\mathbb{V}$ restrict to addition and scalar multiplication on $\mathbb{U}$. Axioms 1,4,5,6,7, and 8 hold since $\mathbb{V}$ is a vector space. For axiom 2, since $\mathbb{U}$ is non-empty, choose a vector $\vec{u} \in \mathbb{U}$ and by Fact 1, $0\vec{u} = \vec{0}$ which is in $\mathbb{U}$ by property 3. Similarly, for axiom 3 let $\vec{u} \in \mathbb{U}$. Then by Fact 1 and property 3, $-\vec{u} = (-1)\vec{u} \in \mathbb{U}$, completing the proof. ∎

Example. Prove $\mathbb{U} = \{p \in \mathcal{P}_2(\mathbb{R}) : p(2) = 0\}$ is a subspace of $\mathcal{P}_2(\mathbb{R})$.

Proof. By the subspace test, we only need to check three things.

1. Since $\vec{0} = 0x^2 + 0x + 0 \in \mathbb{U}$, $\mathbb{U}$ is non-empty.
2. Let $p, q \in \mathbb{U}$. Then $(p+q)(2) = p(2) + q(2) = 0$, so $p + q \in \mathbb{U}$.
3. Let $p \in \mathbb{U}$ and $\alpha \in \mathbb{R}$. Then $(\alpha p)(2) = \alpha p(2) = 0$ so $\alpha p \in \mathbb{R}$.

Therefore by the subspace test, $\mathbb{U}$ is a subspace of $\mathcal{P}_2(\mathbb{R})$. ∎

It is natural to ask now what kind of things aren't subspaces. Here's an example.

Example. Consider the subset

$$\mathbb{L} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : a_1, a_1 \in \mathbb{Z} \right\}.$$

This is not a subspace of $\mathbb{R}^2$ since $\frac{1}{2}\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) \notin \mathbb{L}$ whereas $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) \in \mathbb{L}$

So we have a couple of examples of subspaces, an interesting question to think about is how subspaces can be created. One way is to take a bunch of vectors in your vector space, and then throw in everything else that needs to be there to make that subset a subspace! This is called taking the span of your initial set of vectors.

Definition. Let $\mathcal{B} = \{\vec{v_1}, \dots, \vec{v_k}\}$ be a subset of a vector space $\mathbb{V}$. Define the **span** of $\mathcal{B}$ by

$$\operatorname{Span}(\mathcal{B}) := \{t_1\vec{v_1} + \cdots + t_k\vec{v_k} : t_1, \dots, t_k \in \mathbb{F}\}.$$

You have have heard this terminology before, but a vector of the form $t_1\vec{v_1} + \cdots + t_k\vec{v_k}$ is called a **linear combination** of the vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$. So you can rephrase the definition of the span of a set of vectors to be the set of all linear combinations of the vectors.

So, for example, in $\mathbb{R}^3$, let $\mathcal{B} = \left\{ \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right) \right\}$. Then

$$\operatorname{Span}(\mathcal{B}) = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : z = 0 \right\}.$$

You should convince yourself that if $\mathcal{C} = \left\{ \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right) \right\}$, then $\operatorname{Span}(\mathcal{C}) = \mathbb{R}^3$.

Let's prove now that taking the span of some vectors does actually result in a subspace.

Proposition 3. *Let $\mathcal{B} = \{\vec{v_1}, \dots, \vec{v_k}\}$ be a subset of a vector space $\mathbb{V}$. Then $\operatorname{Span}(\mathcal{B})$ is a subspace of $\mathbb{V}$.*

Proof. Since $\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_k$, $\vec{0} \in \mathrm{Span}(\mathcal{B})$ so $\mathrm{Span}(\mathcal{B})$ is non-empty. Suppose $\vec{x}, \vec{y} \in \mathrm{Span}(\mathcal{B})$, and let $\vec{x} = t_1\vec{v}_1 + \cdots + t_k\vec{v}_k$ and $y = s_1\vec{v}_1 + \cdots + s_k\vec{v}_k$ for elements $t_1, \ldots, t_k, s_1, \ldots, s_k \in \mathbb{F}$. Then

$$\vec{x} + \vec{y} = (t_1 + s_1)\vec{v}_1 + \cdots + (t_k + s_k)\vec{v}_k$$

so $\vec{x} + \vec{y} \in \mathrm{Span}(\mathcal{B})$. Finally, let $\vec{x} \in \mathrm{Span}(\mathcal{B})$ be as above, and let $\alpha \in \mathbb{F}$. Then $\alpha\vec{x} = (\alpha t_1)\vec{v}_1 + \cdots + (\alpha t_k)\vec{v}_k$ and since $\alpha t_i \in \mathbb{F}$ for all i, $\alpha\vec{x} \in \mathrm{Span}(\mathcal{B})$. Therefore, by the subspace test, $\mathrm{Span}(\mathcal{B})$ is a subspace of $\mathbb{V}$. ■

Lecture 4 - July 6

3 Bases and Dimension

We now shift our focus to formalising the notion of dimension. Intuitively we know that $\mathbb{R}^2$ is a 2-dimensional space, because there are 2 different directions one can travel in, and no more. We may also have an idea that $\mathbb{R}^2$ is 2-dimensional since every vector is determined by 2 pieces of information (the x and y coordinate). Similarly, we may guess that $\mathbb{R}^n$ would be an n-dimensional vector space, and we would be correct! However, this geometric intuition fails us when thinking about other vector spaces. For example, what is the dimension of $\mathbb{C}^2$, or $\mathcal{P}_3(\mathbb{R})$, or $\mathcal{C}([0,1])$?

3.1 Linear Indepence, Spanning Sets, and Bases

In order to define dimension, we need to first define a basis.

Definition. A set of vectors $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_k\}$ in a vector space $\mathbb{V}$ is a **spanning set** for $\mathbb{V}$, and we say $\mathcal{B}$ **spans** $\mathbb{V}$, if $\mathrm{Span}(\mathcal{B}) = \mathbb{V}$.

Intuitively, a set of vectors span a vector space if every vector in that vector space can be obtained from those vectors. More precisely, every vector in the vector space is a linear combination of those from the spanning set.

A spanning set can sometimes have redundant information. For examples, the sets

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

are both spanning sets for $\mathbb{R}^2$, but the vector $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ in the first set is redundant. Somehow this is because in the second set, the two vectors point in different directions, but in the first, the three do not. To formalise this, we introduce the notion of linear independence.

Definition. A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_k\}$ in a vector space $\mathbb{V}$ is **linearly independent** if the only solution to the equation

$$t_1\vec{v}_1 + \cdots + t_k\vec{v}_k = \vec{0}$$

is $t_1 = \cdots = t_k = 0$. The set is **linearly dependent** otherwise.

Although this is the formal definition we are to work with, the intuition is that a linearly independent set is a set of vectors that all point in different directions.

Example. The set $\{1 + x, 1\}$ is linearly independent in $\mathcal{P}_1(\mathbb{C})$. To see this, set

$$0 = t_1(1 + x) + t_2(1) = (t_1 + t_2) + t_1 x.$$

Then equating the x coefficient gives us $t_1 = 0$, which implies $t_2 = 0$. Therefore the only solution is $t_1 = t_2 = 0$, so the set is linearly independent.

Example. Since in $\mathbb{R}^2$,

$$-1\begin{pmatrix}1\\0\end{pmatrix}+\frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}+\frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},$$

the set $\left\{\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right),\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right),\left(\begin{smallmatrix}1\\-1\end{smallmatrix}\right)\right\}$ is linearly dependent.

Sometimes it's not so easy to stare at a set of vectors and decide whether or not they are linearly independent. However, we do have some tools for solving simultaneous equations from a previous course to help us along the way!

Example. Is $\{x+x^2-2x^3, 2x-x^2+x^3, x+5x^2+3x^3\}$ linearly independent in $\mathcal{P}_3(\mathbb{R})$?

To check, we want to solve the equation

$$\alpha(x+x^2-2x^3)+\beta(2x-x^2+x^3)+\gamma(x+5x^2+3x^3)=0.$$

Equating coefficients gives us the system of simultaneous equations

$$\begin{aligned}\alpha+2\beta+\gamma&=0\\ \alpha-\beta+5\gamma&=0\\ -2\alpha+\beta+3\gamma&=0.\end{aligned}$$

To solve such a system of equations, we plug the coefficients into an augmented matrix and row reduce! We get

$$\left[\begin{array}{ccc|c}1&2&1&0\\1&-1&5&0\\-2&1&3&0\end{array}\right]\sim\left[\begin{array}{ccc|c}1&0&0&0\\0&1&0&0\\0&0&1&0\end{array}\right].$$

Therefore the system of equations has exactly one solution, and that solution is $\alpha=\beta=\gamma=0$. Therefore the set is linearly independent.

Now if we have a linearly independent spanning set, we have a spanning set which is not redudant. Such a set is a basis for the vector space.

Definition. A **basis** for a vector space $\mathbb{V}$ is a linearly independent subset that spans $\mathbb{V}$.

Fact 4. *Every vector space has a basis.*

We will not prove Fact 4. Here are some examples of bases.

- $\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$ is a basis for $\mathbb{R}^3$.
- $\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$ is a basis for $\mathbb{C}^3$.
- $\left\{\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix},\dots,\begin{pmatrix}0\\\vdots\\0\\0\\1\end{pmatrix}\right\}$ is the **standard basis** for $\mathbb{R}^n$ and $\mathbb{C}^n$.

- $\{1-x, 1+x\}$ is a basis for $\mathcal{P}_1(\mathbb{R})$.
- $\{1, x, x^2, \cdots, x^n\}$ is the **standard basis** for $\mathcal{P}_n(\mathbb{R})$ and $\mathcal{P}_n(\mathbb{C})$.
- $\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$ is a basis for $M_{2\times 2}(\mathbb{R})$.
- $\left\{\begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\3\end{pmatrix}\right\}$ is a basis for $\mathbb{C}^2$.

Lecture 5 - July 9

3.2 Dimension

Looking at these examples, it seems reasonable to define the dimension of a vector space to be the number of vectors needed to make a basis. For this to make sense we must first prove that all bases have the same size.

Lemma 5. *Let $\mathbb{V}$ be a vector space and* $\mathrm{Span}(\{\vec{v}_1, \dots, \vec{v}_n\}) = \mathbb{V}$. *If $\{\vec{u}_1, \dots, \vec{u}_k\}$ is a linearly independent set in $\mathbb{V}$, then $k \le n$.*

Proof. Since $\mathrm{Span}(\{\vec{v}_1, \dots, \vec{v}_n\}) = \mathbb{V}$, we have

$$\begin{aligned}\vec{u}_1 &= a_{11}\vec{v}_1 + \cdots + a_{1n}\vec{v}_n\\ &\vdots\\ \vec{u}_k &= a_{k1}\vec{v}_1 + \cdots + a_{kn}\vec{v}_n\end{aligned}$$

where $a_{ij} \in \mathbb{F}$ for all i and j. We will now aim to show that if $k > n$, then there is a solution to $t_1\vec{u}_1 + \cdots + t_k\vec{u}_k = \vec{0}$ where not all the t_i are 0. We have

$$\begin{aligned}&t_1\vec{u}_1 + \cdots + t_k\vec{u}_k\\ &\quad = t_1(a_{11}\vec{v}_1 + \cdots + a_{1n}\vec{v}_n) + \cdots + t_k(a_{k1}\vec{v}_1 + \cdots + a_{kn}\vec{v}_k)\\ &\quad = (a_{11}t_1 + a_{21}t_2 + \cdots + a_{k1}t_k)\vec{v}_1 + \cdots + (a_{1n}t_1 + \cdots + a_{kn}t_k)\vec{v}_n.\end{aligned}$$

Now, if $k > n$ the system of linear equations

$$\begin{aligned}a_{11}t_1 + \cdots + a_{k1}t_k &= 0\\ &\vdots\\ a_{1n}t_1 + \cdots + a_{kn}t_n &= 0\end{aligned}$$

has a solution where not all the t_i are 0. Consider such a solution. We then have

$$\begin{aligned}\vec{0} &= 0\vec{v}_1 + \cdots + 0\vec{v}_n\\ &= (a_{11}t_1 + \cdots + a_{k1}t_k)\vec{v}_1 + \cdots (a_{1n}t_1 + \cdots + a_{kn}t_n)\vec{v}_n\\ &= t_1\vec{u}_1 + \cdots + t_k\vec{u}_k\end{aligned}$$

contradicting the assumption that $\{\vec{u}_1, \dots, \vec{u}_k\}$ is linearly independent. Therefore $k \le n$. ■

Theorem 6. *Suppose $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\mathcal{C} = \{\vec{u}_1, \dots, \vec{u}_k\}$ are both bases of a vector space $\mathbb{V}$. Then $k = n$.*

Proof. Since $\mathcal{B}$ spans $\mathbb{V}$ and $\mathcal{C}$ is linearly independent, $n \leq k$. However, since $\mathcal{C}$ spans $\mathbb{V}$ and $\mathcal{B}$ is linearly independent, $k \leq n$. Alas $k = n$. ∎

Definition. The **dimension** of a vector space $\mathbb{V}$, denoted $\dim(\mathbb{V})$, is the size of any basis.

Note that Theorem 6 shows that this definition makes sense.

Remark. If there is no finite basis for a vector space $\mathbb{V}$, then we say $\mathbb{V}$ is infinite-dimensional. We set $\dim(\{\vec{0}\}) = 0$.

With the definition of dimension at our disposal, we can now talk about dimension with conviction!

- $\dim(\mathbb{F}^n) = n$ since the standard basis has size n.
- $\dim(\mathcal{P}_n(\mathbb{F})) = n+1$ since the standard basis has size $n+1$.
- $\dim(M_{n\times m}(\mathbb{F})) = nm$ since the standard basis has size nm.

Example. Let $\mathbb{U} = \left\{ \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in M_{2\times 2}(\mathbb{R}) : a+b+c+d = 0 \right\}$. It is an exercise for you to check that $\mathbb{U}$ is a subspace of $M_{2\times 2}(\mathbb{R})$. We will now compute the dimension of $\mathbb{U}$ by finding a basis for $\mathbb{U}$.

Note that every matrix in $\mathbb{U}$ is of the form $\left[\begin{smallmatrix} a & b \\ c & -a-b-c \end{smallmatrix}\right]$, so we can write every matrix in $\mathbb{U}$ as

$$a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$$

so $\left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & -1 \end{smallmatrix}\right] \right\}$ is a spanning set for $\mathbb{U}$. We now check to see whether the set is linearly independent. Consider

$$t_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + t_2 \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} + t_3 \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then by looking at the top left entry we have $t_1 = 0$, the top right gives $t_2 = 0$ and the bottom left gives $t_3 = 0$. Therefore $\left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & -1 \end{smallmatrix}\right] \right\}$ is linearly independent and a basis for $\mathbb{U}$. Behold, $\dim(\mathbb{U}) = 3$.

The next theorem is extremely useful in thinking about dimension. It formally proves things you already know in your heart to be true. Things like "You cannot have 4 linearly independent vectors in $\mathbb{R}^3$, there's just not enough space!" and "You can't span $M_{2\times 2}(\mathbb{C})$ with only 3 vectors, that's not enough because $\dim(M_{2\times 2}(\mathbb{C})) = 4$!" A sketch of the following proof is provided, and you should fill in the details as an exercise.

Theorem 7. *Let $\mathbb{V}$ be an n-dimensional vector space. Then*

1. *A set of more than n vectors must be linearly dependent.*
2. *A set of fewer than n vectors cannot span $\mathbb{V}$.*
3. *A set with n elements in $\mathbb{V}$ is a spanning set for $\mathbb{V}$ if and only if it is linearly independent.*

Proof. Statements 1 and 2 are restatements of Lemma 5. Statement 3 follows from the two paragraphs in Section 3.3. ∎

Lecture 6 - July 10

3.3 Obtaining Bases

There are many ways you could find a basis for a finite-dimensional vector space, here are a couple of important ways.

1. **Extending a linearly independent subset**. Suppose you have a linearly independent subset $\{\vec{v}_1, \dots, \vec{v}_k\}$ in a finite dimensional vector space $\mathbb{V}$. If it is a spanning set, then you have a basis. If not, choose a vector $\vec{v}_{k+1}$ not in the span of $\{\vec{v}_1, \dots, \vec{v}_k\}$. Then $\{\vec{v}_1, \dots, \vec{v}_{k+1}\}$ is must be linearly independent. If this new set spans, then it's a basis. If not, then repeat. This process must eventually stop since our vector space is finite-dimensional, and you will be left with a basis containing $\{\vec{v}_1, \dots, \vec{v}_k\}$.

2. **Reducing an arbitrary finite spanning set**. Suppose you have a finite spanning set $\{\vec{v}_1, \dots, \vec{v}_k\}$ for your vector space, and let's assume that it doesn't contain $\vec{0}$. If it is linearly independent, it is a basis! If not, you can write one of them, say v_i, as a linear combination of the others. Now $\operatorname{Span}(\{\vec{v}_1, \dots, \vec{v}_k\}) = \operatorname{Span}(\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k\})$, so $\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k\}$ spans our vector space. If this new set is linearly independent, then it is a basis! If not, repeat to remove another vector. This process must eventually stop since we started with finitely many vectors in our spanning set. The final product will be a basis made up entirely out of vectors from our original spanning set.

3.4 Coordinates with Respect to a Basis

In $\mathbb{R}^3$, you may have seen the vector $\left(\begin{smallmatrix}3\\2\\4\end{smallmatrix}\right)$ be written as $3\hat{\imath} + 2\hat{\jmath} + 4\hat{k}$. You may have seen this to mean that the vector $\left(\begin{smallmatrix}3\\2\\4\end{smallmatrix}\right)$ can be found 3-units in the x-direction, 2 in the y, and 4 in the z. Alternatively, if $\hat{\imath} = \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)$, $\hat{\jmath} = \left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right)$, and $\hat{k} = \left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right)$, that is, $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$ is the standard basis for $\mathbb{R}^3$, then we can write $\left(\begin{smallmatrix}3\\2\\4\end{smallmatrix}\right) = 3\hat{\imath} + 2\hat{\jmath} + 4\hat{k}$.

In fact, once we have a basis for a vector space, we can think of this as a choice of axes, and we can write every vector as a coordinate vector in much the same way as we think about vectors in $\mathbb{R}^3$.

Example. Consider the vector $\vec{v} = 3 + 5x - 2x^2 \in \mathcal{P}_2(\mathbb{R})$, and the bases $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \{1, 1+x, 1+x+x^2\}$ (as an exercise, prove $\mathcal{C}$ is a basis). Then $\vec{v} = 3(1) + 5(x) + (-2)(x^2)$ so we think of $\vec{v}$ as living at the coordinate $(3, 5, -2)$ with respect to the axes defined by $\mathcal{B}$. We also have $\vec{v} = -2(1) + 7(1+x) + (-2)(1+x+x^2)$ so, with respect to the axes determined by $\mathcal{C}$, we can think of $\vec{v}$ as living at the point $(-2, 7, -2)$. More formally, we can write the coordinate vectors of $\vec{v}$ with respect to $\mathcal{B}$ and $\mathcal{C}$ as

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} \quad \text{and} \quad [\vec{v}]_{\mathcal{C}} = \begin{bmatrix} -2 \\ 7 \\ -2 \end{bmatrix}$$

respectively. This gives us two different ways of looking at the same vector.

A natural question to ask is does it even make sense to talk about coordinate vectors like this. Is it possible that the same vector has two different coordinate vectors with respect to the same basis? The next theorem says the answer is no.

Theorem 8. *Let $\mathbb{V}$ be a vector space, let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a subset of $\mathbb{V}$, and let $\mathbb{U} = \operatorname{Span}(\mathcal{B})$. Then every vector in $\mathbb{U}$ can be expressed in a unique way as a linear combination of the vectors in $\mathcal{B}$ if and only if $\mathcal{B}$ is linearly independent.*

Proof. Suppose every vector in $\mathbb{U}$ is expressed uniquely as a linear combination of the vectors in $\mathcal{B}$. Then there is only one way to write

$$\vec{0} = t_1\vec{v}_1 + \cdots + t_k\vec{v}_k,$$

which is $t_1 = \cdots = t_k = 0$, so $\mathcal{B}$ is linearly independent. Conversely, suppose $\mathcal{B}$ is linearly independent and

$$t_1\vec{v}_1 + \cdots + t_k\vec{v}_k = s_1\vec{v}_1 + \cdots + s_k\vec{v}_k.$$

Rearranging we have $(t_1 - s_1)\vec{v}_1 + \cdots + (t_k - s_k)\vec{v}_k = \vec{0}$. Since $\mathcal{B}$ is linearly independent, this can only be true if $t_i = s_i$ for all i, completing the proof. ■

We can now make the following definition for the coordinate vector of a vector with respect to a given basis.

Definition. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for a vector space $\mathbb{V}$. If $\vec{x} \in \mathbb{V}$ with $\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$, then the **coordinate vector** of $\vec{x}$ with respect to $\mathcal{B}$ is

$$[\vec{x}]_{\mathcal{B}} := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Note that the order of our basis matters. Let $\vec{v} = 2 - i + 4x - ix^2 \in \mathcal{P}_2(\mathbb{C})$. If $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \{1, x^2, x\}$ are bases for $\mathcal{P}_2(\mathbb{C})$, then $[\vec{v}]_{\mathcal{B}} = \left[\begin{smallmatrix} 2-i \\ 4 \\ -i \end{smallmatrix}\right]$ whereas $[\vec{v}]_{\mathcal{C}} = \left[\begin{smallmatrix} 2-i \\ -i \\ 4 \end{smallmatrix}\right]$.

Lecture 7 - July 11

Example. Consider the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} \right\}$$

of $M_{2\times 2}(\mathbb{R})$. Let $\vec{x} = \left[\begin{smallmatrix} 1 & -1 \\ 0 & 3 \end{smallmatrix}\right]$. We wish to find $[\vec{x}]_{\mathcal{B}}$. Consider the equation

$$a\begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} + b\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + c\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix}.$$

To get the coordinate vector of $\vec{x}$ with respect to $\mathcal{B}$, we need to solve for a, b, c, d. Equating the entries of the matrices on the left and right hand side of the equals sign gives us the system of equations

$$\begin{aligned} 3a + b + c + d &= 1 \\ 2a + c + 4d &= -1 \\ 2a + b + c &= 0 \\ 2a + b + 3d &= 3. \end{aligned}$$

To solve this equation we create an augmented matrix and row reduce, giving

$$\left[\begin{array}{cccc|c} 3 & 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & 4 & -1 \\ 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 3 & 3 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

Therefore

$$\vec{x} = 1\begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} + 1\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} - 3\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + 0\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$$

and

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ -3 \\ 0 \end{bmatrix}.$$

Example. Earlier you may have noticed that there is some kind of similarity between $\mathbb{R}^3$ and $\mathcal{P}_2(\mathbb{R})$, and we can somehow identify the vectors

$$\vec{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \qquad \text{and} \qquad \vec{w} = ax^2 + bx + c.$$

Now we can get a glimpse as to how these two vectors may indeed be viewed as the same after picking bases for the two vector spaces. Choose the bases

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \qquad \text{and} \qquad \mathcal{C} = \{x^2, x, 1\}$$

for $\mathbb{R}^3$ and $\mathcal{P}_2(\mathbb{R})$ respectively. Then we see

$$[\vec{v}]_{\mathcal{B}} = [\vec{w}]_{\mathcal{C}} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Once we have chosen a basis for a vector space $\mathbb{V}$, every vector can now be represented as a column matrix (that is, a matrix with only one column). Matrices, as we know, come with an addition and scalar multiplication. A natural question to ask is whether or not the matrix addition and scalar multiplication agrees with the addition and scalar multiplication on $\mathbb{V}$. Since everything so far in this course has worked out so beautifully, it would be a huge surprise if this wasn't true! Indeed, it is true.

Theorem 9. *Let $\mathbb{V}$ be a vector space with basis $\mathcal{B}$. Then*

$$[\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}} = [\vec{x} + \vec{y}]_{\mathcal{B}} \qquad \textit{and} \qquad t[\vec{x}]_{\mathcal{B}} = [t\vec{x}]_{\mathcal{B}}$$

for all $\vec{x}, \vec{y} \in \mathbb{V}$ and all $t \in \mathbb{F}$.

Proof. Exercise. ■

4 Linear Maps

So far in the course we have studied vector spaces in isolation. That is, we've started with a single vector space and studied it, without looking at how it compares to other vector spaces. However, we have seen glimpses that there is something to be said about comparing vector spaces. For example, $\mathbb{R}^3$ and $\mathcal{P}_2(\mathbb{R})$ appear to be the same vector space in some sense, just wrapped up in a different package.

In mathematics in general, when we want to compare objects, we usually think about functions between them. However, when studying functions between two vector spaces, we don't want to just take any old function. We'd like to take into account that we're playing with vector spaces, and vector spaces come with addition and scalar multiplication.

Such a function will be called a linear map, and it's roughly a function that plays nicely with addition and scalar multiplication.

Definition. If $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over $\mathbb{F}$, a function $L : \mathbb{V} \to \mathbb{W}$ is a **linear map** if it satisfies the linearity properties:

1. $L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y})$, and
2. $L(t\vec{x}) = tL(\vec{x})$

for all $\vec{x}, \vec{y} \in \mathbb{V}$, $t \in \mathbb{F}$.

Said another way, it doesn't matter if you add two vectors before or after applying the linear map, and the same with scalar multiplication.

Example. Consider the map $L : \mathcal{P}_3(\mathbb{R}) \to \mathbb{R}$ defined by $L(p) := p(2)$. This is a linear map. In general, let $t \in \mathbb{F}$. Define the **evaluation map**

$$\mathrm{ev}_t : \mathcal{P}_n(\mathbb{F}) \longrightarrow \mathbb{F}$$

by $\mathrm{ev}_t(p) := p(t)$. This is a linear map, and the proof of this claim is left as an exercise.

Example. Let $\mathrm{tr} : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$. be the map defined by taking the trace of a matrix. Recall, if

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

then $\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$. We will prove that tr is a linear map.

Let $A, B \in M_{n\times n}(\mathbb{F})$ with $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$. Then

$$\begin{aligned} \mathrm{tr}(A+B) &= a_{11} + b_{11} + a_{22} + b_{22} + \cdots + a_{nn} + b_{nn} \\ &= a_{11} + a_{22} + \cdots + a_{nn} + b_{11} + b_{22} + \cdots + b_{nn} \\ &= \mathrm{tr}(A) + \mathrm{tr}(B). \end{aligned}$$

Let $t \in \mathbb{F}$, then

$$\begin{aligned} \mathrm{tr}(tA) &= ta_{11} + ta_{22} + \cdots + ta_{nn} \\ &= t(a_{11} + a_{22} + \cdots + a_{nn}) \\ &= t(\mathrm{tr}(A)) \end{aligned}$$

so tr is a linear map.

Lots of natural operations you are familiar with are linear maps. For example, integration and differentiation of polynomials are both linear maps. For example, the map

$$D : \mathcal{P}_3(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$$

given by $D(a_0+a_1x+a_2x^2+a_3x^3) = a_1+2a_2x+3a_3x^2$ can be shown to be a linear map. Similarly the map

$$I : \mathcal{P}_3(\mathbb{R}) \to \mathbb{R}$$

given by $I(p) = \int_{-1}^{1} p(x)dx$ is also a linear map. As an exercise, you should prove that both D and I defined here are linear maps.

Now, associated to every linear map are two subspaces. Roughly speaking, the kernel of a linear map $L : \mathbb{V} \to \mathbb{W}$ are all the vectors in $\mathbb{V}$ that are mapped to $\vec{0} \in \mathbb{W}$. The range of L is all the vectors in $\mathbb{W}$ that are hit by something in $\mathbb{V}$.

Definition. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map. The **range** of L is

$$\mathrm{Range}(L) := \{L(\vec{x}) \in \mathbb{W} : \vec{x} \in \mathbb{V}\}.$$

The **kernel** or **nullspace** of L is

$$\mathrm{Null}(L) := \{\vec{x} \in \mathbb{V} : L(\vec{x}) = \vec{0}\}.$$

Lecture 8 - July 12

We now state some basic properties about linear maps.

Theorem 10. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$, and let $L : \mathbb{V} \to \mathbb{W}$ be a linear map. Then*

1. $L(\vec{0}) = \vec{0}$,
2. $\mathrm{Range}(L)$ *is a subspace of* $\mathbb{W}$, *and*
3. $\mathrm{Null}(L)$ *is a subspace of* $\mathbb{V}$.

Proof. Properties 1 and 2 are left as exercises. For 3, by property 1 $\vec{0} \in \mathrm{Null}(L)$, so $\mathrm{Null}(L)$ is non-empty. Suppose $\vec{v}, \vec{w} \in \mathrm{Null}(L)$. Then $L(\vec{v}+\vec{w}) = L(\vec{v}) + L(\vec{w}) = \vec{0}+\vec{0} = \vec{0}$ so $\mathrm{Null}(L)$ is closed under addition. Let $t \in \mathbb{F}$. Then $L(t\vec{v}) = tL(\vec{v}) = \vec{0}$, so $\mathrm{Null}(L)$ is closed under scalar multiplication. Therefore by the subspace test, $\mathrm{Null}(L)$ is a subspace of $\mathbb{V}$. ■

Example. Consider the linear map map $L : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$L\left(\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Then $\mathrm{Null}(L) = \left\{ \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} \in \mathbb{R}^3 : c \in \mathbb{R} \right\}$ and $\mathrm{Range}(L) = \mathbb{R}^2$.

Example. Let $L : M_{2\times 2}(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be defined by

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = b + c + (c-d)x^2.$$

then

$$\mathrm{Null}(L) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}(\mathbb{R}) : b+c = c-d = 0 \right\} = \left\{ \begin{bmatrix} a & -c \\ c & c \end{bmatrix} : a, c \in \mathbb{R} \right\}.$$

It is clear that $\mathrm{Range}(L) \subset \mathrm{Span}(\{1, x^2\})$. Since $L\left(\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]\right) = 1$ and $L\left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & -1 \end{smallmatrix}\right]\right) = x^2$, we see $\mathrm{Range}(L) \supset \mathrm{Span}(\{1, x^2\})$. Therefore $\mathrm{Range}(L) = \mathrm{Span}(\{1, x^2\})$.

If you pay close attention to these examples, you notice something interesting about the dimensions of the vector spaces involved. In the first example, $\dim(\mathbb{R}^3) = 3$, $\dim(\text{Range}(L)) = 2$, $\dim(\text{Null}(L)) = 1$. In the second we have $\dim(M_{2\times 2}(\mathbb{R})) = 4$, $\dim(\text{Null}(L)) = 2$, and $\dim(\text{Range}(L)) = 2$. Something is clearly going on, so let's give these dimensions some names.

Definition. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. The **rank** of a linear map $L : \mathbb{V} \to \mathbb{W}$ is the dimension of the range of L. The **nullity** of L is the dimension of the nullspace of L. That is,

$$\text{rank}(L) = \dim(\text{Range}(L)) \quad \text{and} \quad \text{nullity}(L) = \dim(\text{Null}(L)).$$

It appears that the number of dimensions you start with is equal to the sum of the number of dimensions that are crushed (that is, the nullity of the linear map) and the number of dimensions that are remaining (the nullity). Let's see if we can formalise this a little more.

Theorem 11 (Rank-Nullity Theorem)**.** *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map. Then $\text{rank}(L) + \text{nullity}(L) = n$.*

The idea of the proof is as follows. We will start with a basis of $\text{Null}(L)$ with k-vectors and extend this to a basis of $\mathbb{V}$ with another n vectors (so $\dim(\mathbb{V}) = n + k$). Then we prove that the image of the new vectors under L give a basis for $\text{Range}(\mathbb{W})$, which will complete the proof.

Proof. Let $\{\vec{v}_1, \dots, \vec{v}_k\}$ be a basis for $\text{Null}(L)$ so $\text{nullity}(L) = k$. Extend this to a basis $\{\vec{v}_1, \dots, v_k, \vec{w}_1, \dots, \vec{w}_n\}$ for $\mathbb{V}$ so $\dim(\mathbb{V}) = n+k$. It suffices to show $\mathcal{B} = \{L(\vec{w}_1), \dots, L(\vec{w}_n)\}$ is a basis for $\text{Range}(L)$. We first show $\text{Span}(\mathcal{B}) = \text{Range}(L)$. Let $\vec{w} = L(\vec{v}) \in \text{Range}(L)$. Then $\vec{v} = t_1\vec{v}_1 + \cdots + t_k\vec{v}_k + s_1\vec{w}_1 + \cdots + s_n\vec{w}_n$ so

$$\begin{aligned}\vec{w} = L(\vec{v}) &= L(t_1\vec{v}_1 + \cdots + t_k\vec{v}_k + s_1\vec{w}_1 + \cdots + s_n\vec{w}_n) \\ &= t_1L(\vec{v}_1) + \cdots + t_kL(\vec{v}_k) + s_1L(\vec{w}_1) + \cdots + s_nL(\vec{w}_n) \\ &= s_1L(\vec{w}_1) + \cdots + s_nL(\vec{w}_n)\end{aligned}$$

so $\mathcal{B}$ is a spanning set for $\text{Range}(L)$. For linear independence, suppose

$$s_1L(\vec{w}_1) + \cdots + s_nL(\vec{w}_n) = \vec{0}.$$

Since L is linear, this implies $s_1\vec{w}_1 + \cdots + s_n\vec{w}_n \in \text{Null}(L)$. Therefore

$$s_1\vec{w}_1 + \cdots + s_n\vec{w}_n = t_1\vec{v}_1 + \cdots + t_k\vec{v}_k$$

for some $t_1, \dots, t_k \in \mathbb{F}$. However, $\{\vec{v}_1, \dots, \vec{v}_k, \vec{w}_1, \dots, \vec{w}_n\}$ is linearly independent, so we must conclude $s_1 = \cdots = s_n = t_1 = \cdots = t_k = 0$. Therefore $\mathcal{B}$ is a basis for $\text{Range}(L)$. Alas, $\text{nullity}(L) = k$, $\text{rank}(L) = n$, and $\dim(\mathbb{V}) = n + k$ completing the proof. ■

Lecture 9 - July 13

Here are some examples of the kinds of things you can conclude with the rank-nullity theorem in your back pocket.

Example. Let $L : \mathcal{P}_3(\mathbb{R}) \to \mathbb{R}^3$ be a linear map. Since $\dim(\mathbb{R}^3) = 3$, it must be that $\text{rank}(L) \leq 3$. Since $\dim(\mathcal{P}_3(\mathbb{R})) = 4$, the rank-nullity theorem implies $\text{nullity}(L) \geq 1$. Therefore without knowing anything about the linear map, we can conclude that there is at least one non-zero vector $\vec{v} \in \mathcal{P}_3(\mathbb{R})$ such that $L(\vec{v}) = \vec{0}$.

Example. Let $L : \mathbb{C}^4 \to M_{2\times 2}(\mathbb{C})$ be a linear map. Then $\text{Null}(L) = \{\vec{0}\}$ if and only if $\text{Range}(L) = M_{2\times 2}(\mathbb{C})$.

Proof. First note $\dim(\mathbb{C}^4) = \dim(M_{2\times 2}(\mathbb{C})) = 4$. If $\text{Null}(L) = \{\vec{0}\}$ then $\text{nullity}(L) = 0$ so the rank-nullity theorem says $\text{rank}(L) = 4$. Therefore $\text{Range}(L)$ is a 4-dimensional subspace of $M_{2\times 2}(\mathbb{C})$ so it must be that $\text{Range}(L) = M_{2\times 2}(\mathbb{C})$. Conversely, if $\text{Range}(L) = M_{2\times 2}(\mathbb{C})$, then $\text{rank}(L) = 4$. Therefore $\text{nullity}(L) = 0$ so $\text{Null}(L) = \{\vec{0}\}$. ■

4.1 Linear Maps as Matrices

Consider the linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ given by $L\left(\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} a+2b \\ a-2b \end{smallmatrix}\right)$. Then if we think of the vectors as 2×1 column matrices, we can actually find a matrix that does the linear map for us. Indeed,

$$\begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+2b \\ a-2b \end{bmatrix}.$$

The fact that a matrix even existed in this example was plausible because it's not much of a stretch of the imagination to view a vector in $\mathbb{R}^2$ as a column matrix. Wouldn't it be nice if we had a way to view every vector in every vector space as a column matrix? Wait, we do! Remember that once you fix a basis for a vector space, then every vector can be written as a column matrix by simply taking its coordinate vector.

So, now that we have this, it's reasonable to ask whether or not every linear map can be written as a matrix. Let's take a look at another example, and this time we'll turn our vectors into column matrices by taking coordinate vectors.

Example. Let $L : \mathcal{P}_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ be the linear map defined by $L(a + bx + cx^2) = \left[\begin{smallmatrix} a-2b & 4c \\ a+b+c & b-c \end{smallmatrix}\right]$. Fix the basis

$$\mathcal{B} = \{1, x, x^2\} \quad \text{and} \quad \mathcal{C} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

for $\mathcal{P}_2(\mathbb{R})$ and $M_{2\times 2}(\mathbb{R})$ respectively. Then if there is a matrix A which performs the linear map for us (by matrix multiplication of course), it must be such that

$$A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a-2b \\ 4c \\ a+b+c \\ b-c \end{bmatrix}.$$

We first note that if A is to exist, it must be a 3×4 matrix. With that in mind, if we stare at this really hard (we'll talk about how to do it without straining your eyes a little later on) we can see that

$$A = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 4 \\ 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

For some foreshadowing of notation, we let ${}_{\mathcal{C}}[L]_{\mathcal{B}} = A$.

At this point you could be forgiven for thinking that we can always find a matrix that performs the linear map for us. And you would be forgiven because you haven't thought anything incorrect! This is the content of the next theorem.

Before we state and prove it though, it is worth addressing why we'd care to do this. Matrices, while simply an array of numbers, come equipped with machinery to compute many things. You may have seen in previous courses methods to find bases for column spaces and nullspaces. It will turn out that once we turn our linear map into a matrix, we can use this machinery to learn about our linear map.

Theorem 12. *Let $\mathbb{V}$ be an n-dimensional vector space with basis $\mathcal{B}$. Let $\mathbb{W}$ be an m-dimensional vector space with basis $\mathcal{C}$. Then for every linear map $L : \mathbb{V} \to \mathbb{W}$, there exists an $m \times n$ matrix A such that $[L(\vec{v})]_{\mathcal{C}} = A[\vec{v}]_{\mathcal{B}}$ for all $\vec{v} \in \mathbb{V}$. Conversely, every $m \times n$ matrix A defines a linear map $L : \mathbb{V} \to \mathbb{W}$ by $[L(\vec{v})]_{\mathcal{C}} := A[\vec{v}]_{\mathcal{B}}$.*

Proof. Since matrix multiplication satisfies $A(B + C) = AB + AC$ and $t(AB) = A(tB)$ for all matrices A, B, C and all scalars $t \in \mathbb{F}$, A defines a linear map $L : \mathbb{V} \to \mathbb{W}$ by $A[\vec{v}]_{\mathcal{B}} = [L(\vec{v})]_{\mathcal{C}}$. For the forward direction, let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\mathcal{C} = \{\vec{w}_1, \dots, \vec{w}_m\}$. Let $\vec{v} \in \mathbb{V}$, then $\vec{v} = t_1\vec{v}_1 + \cdots + t_n\vec{v}_n$ and $L(\vec{v}) = s_1\vec{w}_1 + \cdots + s_m\vec{w}_m$. Since L is linear we have

$$L(\vec{v}) = t_1L(\vec{v}_1) + \cdots + t_nL(\vec{v}_n) = s_1\vec{w}_1 + \cdots + s_m\vec{w}_m.$$

For each $i \in \{1, \dots, n\}$, let $L(\vec{v}_i) = a_{1i}\vec{w}_1 + \cdots + a_{mi}\vec{w}_m$. Then

$$\begin{aligned} L(\vec{v}) = s_1\vec{w}_1 + \cdots + s_m\vec{w}_m &= t_1(a_{11}\vec{w}_1 + \cdots + a_{m1}\vec{w}_m) + \cdots + t_n(a_{1n}\vec{w}_1 + \cdots a_{mn}\vec{w}_m) \\ &= (a_{11}t_1 + a_{12}t_2 + \cdots + a_{1n}t_n)\vec{w}_1 + \cdots + (a_{m1}t_1 + \cdots + a_{mn}t_n)\vec{w}_m. \end{aligned}$$

Therefore we have $s_i = a_{i1}t_1 + \cdots + a_{in}t_n$ for all $i \in \{1, \dots, m\}$. This is of course how matrix multipication works, and we see

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}.$$

Since $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$ and $[L(\vec{v})]_{\mathcal{C}} = \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}$, the proof is completed. ■

Lecture 10 - July 16

Hidden in the proof is the fact that if $\vec{v}_i$ is the ith basis vector of $\mathcal{B}$, then $[L(\vec{v}_i)]_{\mathcal{C}}$ is simply the ith column of the desired matrix A. This gives us the following corollary.

Corollary 13. *Let $\mathbb{V}$ be a vector space with basis $\mathcal{B} = \{\vec{\beta}_1, \dots, \vec{\beta}_n\}$. Let $\mathbb{W}$ be a vector space with basis $\mathcal{C} = \{\vec{\gamma}_1, \dots, \vec{\gamma}_m\}$. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map. Then the $m \times n$ matrix A such that $[L(\vec{v})]_{\mathcal{C}} = A[\vec{v}]_{\mathcal{B}}$ for all $\vec{v} \in \mathbb{V}$, which we denote ${}_{\mathcal{C}}[L]_{\mathcal{B}}$, is given by*

$${}_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} [L(\vec{\beta}_1)]_{\mathcal{C}} & \cdots & [L(\vec{\beta}_n)]_{\mathcal{C}} \end{bmatrix}.$$

The fact that the matrix contains all the information of L, and is determined by the images of the basis vectors tells us something very interesting about linear maps: They are entirely determined by where they send a basis.

The matrix A for a linear map L is determined once you pick a basis for each vector space. We will give this matrix a name.

Definition. We call the matrix ${}_{\mathcal{C}}[L]_{\mathcal{B}}$ the **matrix of the linear map** L with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. If $L : \mathbb{V} \to \mathbb{V}$ and we are choosing the same basis $\mathcal{B}$ for both the domain and codomain of L, then we may write $[L]_{\mathcal{B}} = {}_{\mathcal{B}}[L]_{\mathcal{B}}$.

Example. Consider the differentiation map $D : \mathcal{P}_3(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$, and let both vector spaces be endowed with the standard bases $\mathcal{B}$ and $\mathcal{C}$ respectively. Then $D(1) = 0$, $D(x) = 1$, $D(x^2) = 2x$, and $D(x^3) = 3x^2$. Therefore

$${}_{\mathcal{C}}[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Let's just double check with a specific example. Let $\vec{v} = 4 + 2x + (-2)x^2 + ix^3$. Then $D(\vec{v}) = 2 - 4x + 3ix^2$ so $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 4 \\ 2 \\ -2 \\ i \end{bmatrix}$ and $[D(\vec{v})]_{\mathcal{C}} = \begin{bmatrix} 2 \\ -4 \\ 3i \end{bmatrix}$. Indeed, we can check that

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ -2 \\ i \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 3i \end{bmatrix}.$$

If you dwell on Theorem 12, it becomes apparent that the theorem only works because of the way matrix multiplication is defined. When you first came across matrix multiplication, the way it is defined may have been enough to put you off your food for the rest of the day. But it's defined that way so that Theorem 12 is true! Even better, the next fact is also true, although we will not prove it here. If you feel like a moderately difficult challenge, you should prove it! You definitely have the tools to do so at this point in the course, the proof is more a matter of careful bookkeeping.

Fact 14. *Let $\mathbb{V}$, $\mathbb{U}$, and $\mathbb{W}$ be vector spaces with bases $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ respectively. Let $L : \mathbb{V} \to \mathbb{U}$ and $M : \mathbb{U} \to \mathbb{W}$ be linear maps. Then ${}_{\mathcal{D}}[M \circ L]_{\mathcal{B}} = {}_{\mathcal{D}}[M]_{\mathcal{C}}\,{}_{\mathcal{C}}[L]_{\mathcal{B}}$.*

Now, let's see some of the computational power of matrices in action. First we recall the notions of a column space and nullspace for a matrix, and see how they relate to the range and nullspace of a linear map.

Definition. Let A be an $m \times n$ matrix with entries in $\mathbb{F}$. The **column space** of A is defined as the span of the columns of A and is denoted $\mathrm{Col}(A)$. Here the columns are veiwed as vectors in $\mathbb{F}^m$, so $\mathrm{Col}(A)$ is a subspace of $\mathbb{F}^m$. The **nullspace** of A, denoted $\mathrm{Null}(A)$, is all the $n \times 1$ matrices $\vec{v}$ (viewed as vectors in $\mathbb{F}^n$) such that $A\vec{v} = \mathbf{0}$. Here $\mathbf{0}$ is the $n \times 1$ matrix with all entries 0. Since the vectors are viewed as vectors in $\mathbb{F}^n$, $\mathrm{Null}(A)$ is a subspace of $\mathbb{F}^n$.

Given a matrix, let's recall how we find a basis for the column space and nullspace. For the column space, one row-reduces the matrix and chooses the original columns corresponding to the leading ones. For the nullspace, one solves the system of equations given by augmenting the matrix with a column of 0 and then taking the basic solutions. Let's see this in an example.

Example. Find a basis for $\mathrm{Col}(A)$ and $\mathrm{Null}(A)$ where $A = \begin{bmatrix} 1 & 2 & 5 & -3 & -8 \\ -2 & -4 & -11 & 2 & 4 \\ -1 & -2 & -6 & -1 & -4 \\ 1 & 2 & 5 & -2 & -5 \end{bmatrix}$.

Here we go! First, we put the matrix A into row reduced echelon form, which is given by

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

So, immediately, we have that a basis for $\text{Col}(A)$ is

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ -11 \\ -6 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ -1 \\ -2 \end{bmatrix} \right\}.$$

Finding a basis for $\text{Null}(A)$ is a little more involved. Finding a matrix $\vec{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}$ such that $A\vec{v} = \mathbf{0}$ is the same as solving the system of equations given by the augmented matrix $[A \mid \mathbf{0}]$. That is, we put the matrix A in an augmented matrix with 0's in the last column, and solve that system of equations. The row-reduced augmented matrix is given by

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

If we let the variables be $x_1, \dots, x_5$ for this system, we can write down an entire set of solutions as follows. For every column not corresponding to a leading 1, we let that variable be a free variable, and solve for the rest of them. In this example, the free variables are x_2 and x_5, so let $x_2 = s$ and $x_5 = t$. Then

$$\begin{aligned} x_1 &= -t - 2s \\ x_2 &= s \\ x_3 &= 0 \\ x_4 &= -3t \\ x_5 &= t \end{aligned}$$

so every vector in $\text{Null}(A)$ is of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = t \begin{bmatrix} -1 \\ 0 \\ 0 \\ -3 \\ 1 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Finally, we write down a basis for $\text{Null}(A)$ as

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 0 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

Now we draw our attention back to linear maps because after all, linear maps are matrices, and matrices are linear maps! The next proposition allows us to harness the computational power of matrices to learn about the range and nullspace of a linear map.

Proposition 15. *Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map and $\mathcal{B}$ and $\mathcal{C}$ bases for $\mathbb{V}$ and $\mathbb{W}$ respectively. Let $A = {}_{\mathcal{C}}[L]_{\mathcal{B}}$.*

- $\vec{v} \in \operatorname{Null}(L)$ *if and only if* $[\vec{v}]_{\mathcal{B}} \in \operatorname{Null}(A)$.
- $\vec{w} \in \operatorname{Range}(L)$ *if and only if* $[\vec{w}]_{\mathcal{C}} \in \operatorname{Col}(A)$.

Proof. Exercise. ■

This proposition tells us that if we want to find a basis for the nullspace and range of a linear map, we just need to pick some bases, find the matrix associated to the linear map and find bases for the nullspace and column space of the matrix.

Lecture 11 - July 17

Example. Consider the extremely contrived linear map $L : \mathcal{P}_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ given by $L(a + bx + cx^2) = \left[\begin{smallmatrix} a+b+c & a-b+3c \\ 3a+b+5c & 0 \end{smallmatrix}\right]$. We will now find a basis for $\operatorname{Null}(L)$ and $\operatorname{Range}(L)$.

Let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases for $\mathcal{P}_2(\mathbb{R})$ and $M_{2\times 2}(\mathbb{R})$ respectively. Since $L(1) = \left[\begin{smallmatrix} 1 & 1 \\ 3 & 0 \end{smallmatrix}\right]$, $L(x) = \left[\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right]$, and $L(x^2) = \left[\begin{smallmatrix} 1 & 3 \\ 5 & 0 \end{smallmatrix}\right]$ we have

$$ {}_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 3 \\ 3 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}. $$

Call this matrix A. We will now find bases for $\operatorname{Col}(A)$ and $\operatorname{Null}(A)$, and then convert this information back to find bases for $\operatorname{Range}(L)$ and $\operatorname{Null}(L)$. Row reducing A gives

$$ \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 3 \\ 3 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. $$

With a little work we compute bases for $\operatorname{Col}(A)$ and $\operatorname{Null}(A)$ to be

$$ \left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \right\} $$

respectively. Since these are coordinate vectors, we finally have that

$$ \left\{ \begin{bmatrix} 1 & 3 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \right\} \quad \text{and} \quad \{-2 + x + x^2\} $$

are bases for $\operatorname{Range}(L)$ and $\operatorname{Null}(L)$ respectively.

We may be faced with a situation where we want to switch bases for the same vector space, because a particular problem is computationally easier to solve in one bases. We do this all the time in physics when we choose a set of coordinates that is natural with respect to the problem at hand. So, if we are given bases $\mathcal{B}$ and $\mathcal{C}$ of a vector space $\mathbb{V}$, it would be great if we had a matrix that takes a coordinate vector with respect to $\mathcal{B}$ and spits out the coordinate vector with respect to $\mathcal{C}$. This can be achieved by simply finding the matrix of the linear map $\mathrm{id} : \mathbb{V} \to \mathbb{V}$, which is the linear map that sends every vector $\vec{v}$ to itself.

Example. Let $\mathcal{S} = \{1, x, x^2\}$ be the standard basis for $\mathcal{P}_2(\mathbb{R})$ and $\mathcal{B} = \{1, 1+x, 1+x+x^2\}$ another basis. We would like a matrix A such that $A[\vec{v}]_\mathcal{S} = [\vec{v}]_\mathcal{B}$. To find A, consider the linear map $\mathrm{id} : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ given by $\mathrm{id}(\vec{v}) = \vec{v}$ for all $\vec{v} \in \mathbb{V}$ and we will find ${}_\mathcal{B}[\mathrm{id}]_\mathcal{S}$. This should be our desired matrix since ${}_\mathcal{B}[\mathrm{id}]_\mathcal{S}[\vec{v}]_\mathcal{S} = [\mathrm{id}(\vec{v})]_\mathcal{B}$ but $\mathrm{id}(\vec{v}) = \vec{v}$. We will call this matrix ${}_\mathcal{B}\mathcal{P}_\mathcal{S}$.

We have

$$[\mathrm{id}(1)]_\mathcal{B} = [1]_\mathcal{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad [\mathrm{id}(x)]_\mathcal{B} = [x]_\mathcal{B} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad [\mathrm{id}(x^2)]_\mathcal{B} = [x^2]_\mathcal{B} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

We also have

$$[1]_\mathcal{S} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad [1+x]_\mathcal{S} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad [1+x+x^2]_\mathcal{S} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Therefore

$${}_\mathcal{B}\mathcal{P}_\mathcal{S} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad {}_\mathcal{S}\mathcal{P}_\mathcal{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

If these matrices do what we say they should, then we should be able to use them to switch coordinates between $\mathcal{S}$ and $\mathcal{B}$. Let's check. Consider the linear map $D : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ given by differentiation. Then $[D]_\mathcal{S} = \left[\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{smallmatrix}\right]$. If we want to find $[D]_\mathcal{B}$, we should be able to first change coordinates from $\mathcal{B}$ to $\mathcal{S}$, apply $[D]_\mathcal{S}$, and then switch back. That is, we should have $[D]_\mathcal{B} = {}_\mathcal{B}\mathcal{P}_\mathcal{S}[D]_\mathcal{S}{}_\mathcal{S}\mathcal{P}_\mathcal{B}$. Let's check!

$${}_\mathcal{B}\mathcal{P}_\mathcal{S}[D]_\mathcal{S}{}_\mathcal{S}\mathcal{P}_\mathcal{B} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Also, $[D(1)]_\mathcal{B} = \left[\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right]$, $[D(1+x)]_\mathcal{B} = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right]$, and $[D(1+x+x^2)]_\mathcal{B} = \left[\begin{smallmatrix} -1 \\ 2 \\ 0 \end{smallmatrix}\right]$. Therefore

$$[D]_\mathcal{B} = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

so we do indeed have $[D]_\mathcal{B} = {}_\mathcal{B}\mathcal{P}_\mathcal{S}[D]_\mathcal{S}{}_\mathcal{S}\mathcal{P}_\mathcal{B}$.

Definition. Let $\mathbb{V}$ be a finite dimensional vector space, and let $\mathcal{B}$ and $\mathcal{C}$ be two bases for $\mathbb{V}$. The **change of coordinate matrix** ${}_\mathcal{C}P_\mathcal{B}$ is the matrix of the linear mapping $\mathrm{id} : \mathbb{V} \to \mathbb{V}$ where $\mathrm{id}(\vec{v}) = \vec{v}$ for all $v \in \mathbb{V}$. This name makes sense since $[\vec{v}]_\mathcal{C} = {}_\mathcal{C}P_\mathcal{B}[\vec{v}]_\mathcal{B}$ for all $\vec{v} \in \mathbb{V}$.

We'll finish this section by addressing the following, perhaps natural, question: What is the relationship between ${}_{\mathcal{C}}\mathcal{P}_{\mathcal{B}}$ and ${}_{\mathcal{B}}\mathcal{P}_{\mathcal{C}}$? Notice

$$ {}_{\mathcal{C}}\mathcal{P}_{\mathcal{B}}\,{}_{\mathcal{B}}\mathcal{P}_{\mathcal{C}}[\vec{v}]_{\mathcal{C}} = [\vec{v}]_{\mathcal{C}} \quad \text{and} \quad {}_{\mathcal{B}}\mathcal{P}_{\mathcal{C}}\,{}_{\mathcal{C}}\mathcal{P}_{\mathcal{B}}[\vec{v}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{B}} $$

for all $\vec{v} \in \mathbb{V}$. With this in mind you are able to write up a proof of the next proposition.

Proposition 16. *Let $\mathbb{V}$ be a finite dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$. Then ${}_{\mathcal{C}}P_{\mathcal{B}} = ({}_{\mathcal{B}}P_{\mathcal{C}})^{-1}$.*

Proof. Exercise. ∎

Lecture 12 - July 18

5 Isomorphisms of Vector Spaces

Let's return to an observation that has come up a few times in this course: $\mathbb{R}^3$ and $\mathcal{P}_2(\mathbb{R})$ are the same! At least they feel the same. We could just rename the element $\left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right)$ by $a + bx + cx^2$ and everything would work exactly the same. Somehow it feels like these two elements are the same thing called by different names. We will soon see that these two vector spaces, while they have different names, have exactly the same structure. More precisely, we will see that $\mathbb{R}^3$ and $\mathcal{P}_2(\mathbb{R})$ are *isomorphic*.

An isomorphism between vector spaces (whatever that is) should be thought of kind of like a translator. It's a linear map that preserves information perfectly. No information is lost, and no information is missed. So far, admittedly, this doesn't make much sense. Let's look at a couple of examples to get a little more intuition.

Example. Consider the linear map $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^2$ given by $L(p) = \left(\begin{smallmatrix} p(0) \\ p(0) \end{smallmatrix}\right)$. This linear map is not an isomorphism because somehow it loses information. For example, $L(x + 2) = L(x^2 + 2) = L(2) = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right)$ so just by looking at the output of L, we can't tell the difference between $x + 2$ and 2 for example. Furthermore, L somehow misses information. For example, nothing maps to the vector $\left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right)$.

Example. On the other hand, the map $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ given by

$$ L(p) = \begin{pmatrix} p(-1) \\ p(0) \\ p(1) \end{pmatrix}. $$

There are two very interesting things about this map. Firstly, it turns out that if you know the value of a polynomial in $\mathcal{P}_2(\mathbb{R})$ evaluated and three distinct points, you are able to recover the polynomial. That is, if $L(p) = L(q)$ then $p = q$. Furthermore, for any three numbers $a, b, c \in \mathbb{R}$, there is a polynomial $p \in \mathcal{P}_2(\mathbb{R})$ such that $p(-1) = a$, $p(0) = b$, and $p(1) = c$. Therefore, $\text{Range}(L) = \mathbb{R}^3$. With these two pieces of information, we can see that L is a perfect dictionary between $\mathcal{P}_2(\mathbb{R})$ and $\mathbb{R}^3$ and both vector spaces contain the same information, just wrapped up in a different package.

Roughly, an isomorphism of vector spaces will be a linear map which is a perfect dictionary, that is, no information is lost, and no information is missed. More formally, it will be a linear map that is injective and surjective, which we will now define.

Definition. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map between vector spaces. We say L is **injective** (or one-to-one) if $L(\vec{v}_1) = L(\vec{v}_2)$ implies $\vec{v}_1 = \vec{v}_2$. We say L is **surjective** (or onto) if $\text{Range}(L) = \mathbb{W}$.

If we are handed a linear map and want to know whether or not it is an isomorphism, we just have to check that it's injective and surjective. Here is a little result that will make checking injectivity that much easier.

Lemma 17. *A linear map L is injective if and only if* $\text{Null}(L) = \{\vec{0}\}$.

Proof. Suppose $L : \mathbb{V} \to \mathbb{W}$ is injective and let $\vec{v} \in \text{Null}(L)$. Then $L(\vec{v}) = L(\vec{0}) = \vec{0}$ so $\vec{v} = \vec{0}$. Therefore $\text{Null}(L) = \{\vec{0}\}$. Conversely, suppose $\text{Null}(L) = \{0\}$ and let $L(\vec{v}) = L(\vec{w})$. Then $\vec{0} = L(\vec{v}) - L(\vec{w}) = L(\vec{v} - \vec{w})$ so $\vec{v} - \vec{w} \in \text{Null}(L)$. Since the only vector in $\text{Null}(L)$ is $\vec{0}$, we have $\vec{v} - \vec{w} = \vec{0}$ so $\vec{v} = \vec{w}$ completing the proof. ■

Definition. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear map. If L is injective and surjective, we say L is an **isomorphism**. If L is an isomorphism, we say the vector spaces $\mathbb{V}$ and $\mathbb{W}$ are **isomorphic** and write $\mathbb{V} \cong \mathbb{W}$.

Let's see some examples.

Example. Consider $L : M_{2\times 2}(\mathbb{C}) \to \mathbb{C}^3$ given by $L\left(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]\right) = \begin{pmatrix} a+b \\ b-2c \\ a+b+d \end{pmatrix}$. Then $\left[\begin{smallmatrix} 2 & -2 \\ -1 & 0 \end{smallmatrix}\right] \in \text{Null}(L)$ so L is not injective, and is therefore not an isomorphism.

Example. Let $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ be the linear map given by $L(a + bx + cx^2) = \left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right)$. You should check that $\text{Null}(L) = \{\vec{0}\}$. The rank-nullity theorem now implies that $\text{rank}(L) = 3$ so $\text{Range}(L)$ is a 3-dimensional subspace of $\mathbb{R}^3$, so $\text{Range}(L) = \mathbb{R}^3$. Alas, L is an isomorphism and $\mathcal{P}_2(\mathbb{R})$ is isomorphic to $\mathbb{R}^3$ (so we can write $\mathcal{P}_2(\mathbb{R}) \cong \mathbb{R}^3$).

There can be more than one isomorphism between isomorphic vector spaces.

Example. Consider again the linear map $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ given by

$$L(p) = \begin{pmatrix} p(-1) \\ p(0) \\ p(1) \end{pmatrix}.$$

Let's prove it is indeed an isomorphism, without assuming we know the fact that every degree at most 2 polynomial is uniquely determined by 3 points. Let's compute the nullspace and range of L by finding the matrix of the linear map with respect to the standard bases. Let $\mathcal{B}$ be the standard basis for $\mathcal{P}_2(\mathbb{R})$, and $\mathcal{C}$ the standard basis for $\mathbb{R}^3$. Since $L(1) = \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)$, $L(x) = \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right)$, and $L(x^2) = \left(\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right)$. Therefore

$$_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

which has row reduced echelon form $\left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]$. Since the identity matrix has 3 leading ones, $\text{rank}(L) = 3$ so L is surjective. Applying the rank-nullity theorem gives $\text{nullity}(L) = 0$. Therefore L is surjective and injective, and L is another isomorphism between $\mathcal{P}_2(\mathbb{R})$ and $\mathbb{R}^3$.

Suppose $\mathbb{V} \cong \mathbb{W}$. This does not imply that every linear map $L : \mathbb{V} \to \mathbb{W}$ is an isomorphism! Consider, for example, the linear map $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ given by $L(p) = \vec{0}$ for all $p \in \mathcal{P}_2(\mathbb{R})$. Then $\text{nullity}(L) = 3$ so L is not injective. However we have seen, twice, that $\mathcal{P}_2(\mathbb{R}) \cong \mathbb{R}^3$.

If the intuition that an isomorphism is a kind of translator, then there should be a way to do an isomorphism in reverse, just like you should be able to translate a word back into English, if you had already translated it into Spanish. This is indeed true, and the next proposition makes it precise.

Proposition 18. *A linear map $L : \mathbb{V} \to \mathbb{W}$ is an isomorphism if and only if there exists a linear map $L^{-1} : \mathbb{W} \to \mathbb{V}$ such that $L \circ L^{-1}(\vec{w}) = \vec{w}$ for all $\vec{w} \in \mathbb{W}$ and $L^{-1} \circ L(\vec{v}) = \vec{v}$ for all $\vec{v} \in \mathbb{V}$. In this case we call L^{-1} the* **inverse linear map** *to L.*

Proof sketch. Given an isomorphism $L : \mathbb{V} \to \mathbb{W}$, define $L^{-1} : \mathbb{W} \to \mathbb{V}$ by $L^{-1}(\vec{w}) = \vec{v}_{\vec{w}}$ where $\vec{v}_{\vec{w}} \in \mathbb{V}$ is the unique vector such that $L(\vec{v}_{\vec{w}}) = \vec{w}$. Such a unique vector exists since L is injective and surjective. It is left to you to prove that L^{-1} is a linear map satisfying the desired properties. For the converse direction, you should check that if such an inverse map exists, then L must necessarily be surjective and injective. ∎

Lecture 13 - July 19

Given an isomorphism, it is sometimes very easy to write down the inverse linear map, and sometimes not. For example, return to the isomorphism $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ given by $L(a+bx+cx^2) = \left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right)$. Then $L^{-1} : \mathbb{R}^3 \to \mathcal{P}_2(\mathbb{R})$ is given by $L^{-1}\left(\left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right)\right) = a + bx + cx^2$. Let's check this is indeed the inverse. We have

$$L \circ L^{-1}\left(\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = L(a + bx + cx^2) = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and

$$L^{-1} \circ L(a + bx + cx^2) = L^{-1}\left(\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = a + bx + cx^2$$

so this is the inverse.

Guessing the inverse linear map is not always so easy. For example, what is the inverse to the isomorphism $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ given by $L(p) = \left(\begin{smallmatrix} p(-1) \\ p(0) \\ p(1) \end{smallmatrix}\right)$? The next proposition, which is a consequence of Proposition 18, gives us a way to find inverses to isomorphisms.

Proposition 19. *Let $L : \mathbb{V} \to \mathbb{W}$ be an isomorphism. Let $\mathcal{B}$ be a basis for $\mathbb{V}$, and $\mathcal{C}$ a basis for $\mathbb{W}$. Then ${}_{\mathcal{C}}[L]_{\mathcal{B}}$ is an invertible matrix and ${}_{\mathcal{C}}[L]_{\mathcal{B}}^{-1} = {}_{\mathcal{B}}[L^{-1}]_{\mathcal{C}}$.*

Proof. Exercise. ∎

Let's see this in action!

Example. Let $L : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ be the isomorphism given by $L(p) = \left(\begin{smallmatrix} p(-1) \\ p(0) \\ p(1) \end{smallmatrix}\right)$. We have already seen that

$${}_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Using your favourite method of computing the inverse of a matrix, we have

$$_{\mathcal{B}}[L^{-1}]_{\mathcal{C}} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}.$$

Since

$$\begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ -\frac{1}{2}a + \frac{1}{2}c \\ \frac{1}{2}a - b + \frac{1}{2}c \end{bmatrix}$$

we have

$$L^{-1}\left(\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = b + \left(-\frac{1}{2}a + \frac{1}{2}c\right)x + \left(\frac{1}{2}a - b + \frac{1}{2}c\right)x^2.$$

Using only the power of linear algebra we have figured out how to write down a polynomial $p \in \mathcal{P}_2(\mathbb{R})$ given only $p(-1), p(0)$, and $p(1)$.

If we are to think of an isomorphism as simply a renaming of vectors, which we should, then we should expect two isomorphic vector spaces to have the same structure. At the very least, it wouldn't be unreasonable to expect two isomorphic vector spaces to have the same dimension. In fact, suppose $L : \mathbb{V} \to \mathbb{W}$ is a linear map. If $\dim(\mathbb{V}) < \dim(\mathbb{W})$, then the rank-nullity theorem says $\text{Range}(L)$ cannot be all of $\mathbb{W}$, so L cannot be surjective. If $\dim(\mathbb{V}) > \dim(\mathbb{W})$, the rank-nullity theorem says $\text{nullity}(L) \geq 1$, so L cannot be injective. So if $\mathbb{V} \cong \mathbb{W}$ we at least must have that $\dim(\mathbb{V}) = \dim(\mathbb{W})$. The natural thing to figure out now is whether or not we an have vector spaces of the same dimension that are not isomorphic.

Towards this, suppose $\mathbb{V}$ and $\mathbb{W}$ have the same dimension, and pick bases for both. Then the coordinate vectors for both vector spaces look exactly the same, they are column vectors with $\dim(\mathbb{V}) = \dim(\mathbb{W})$ rows. This perhaps suggests that if $\dim(\mathbb{V}) = \dim(\mathbb{W})$, then $\mathbb{V} \cong \mathbb{W}$.

Theorem 20. *Suppose $\mathbb{V}$ and $\mathbb{W}$ are finite dimensional vector spaces over the same field $\mathbb{F}$. Then $\mathbb{V}$ and $\mathbb{W}$ are isomorphic if and only if $\dim(\mathbb{V}) = \dim(\mathbb{W})$.*

Proof. Suppose $\mathbb{V} \cong \mathbb{W}$ via an isomorphism $L : \mathbb{V} \to \mathbb{W}$. Since L is injective, $\text{nullity}(L) = 0$ so the rank-nullity theorem implies $\dim(\mathbb{V}) = \text{rank}(L)$. Since L is surjective, $\text{rank}(L) = \dim(\mathbb{W})$ so $\dim(\mathbb{V}) = \dim(\mathbb{W})$.

Conversely, let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{V}$ and $\mathcal{C} = \{\vec{w}_1, \ldots, \vec{w}_n\}$ a basis for $\mathbb{W}$. Define a map $L : \mathbb{V} \to \mathbb{W}$ by

$$L(t_1\vec{v}_1 + \cdots + t_n\vec{v}_n) = t_1\vec{w}_1 + \cdots + t_n\vec{w}_n.$$

L is linear since

$$\begin{aligned} L((t_1\vec{v}_1 + \cdots + t_n\vec{v}_n) + (s_1\vec{v}_1 + \cdots + s_n\vec{v}_n)) &= L((t_1+s_1)\vec{v}_1 + \cdots + (t_n+s_n)\vec{v}_n) \\ &= (t_1+s_1)\vec{w}_1 + \cdots + (t_n+s_n)\vec{w}_n \\ &= (t_1\vec{w}_1 + \cdots + t_n\vec{w}_n) + (s_1\vec{w}_1 + \cdots + s_n\vec{w}_n) \\ &= L((t_1\vec{v}_1 + \cdots + t_n\vec{v}_n)) + L((s_1\vec{v}_1 + \cdots + s_n\vec{v}_n)) \end{aligned}$$

and

$$\begin{aligned} L(\alpha(t_1\vec{v}_1 + \cdots + t_n\vec{v}_n)) &= L(\alpha t_1\vec{v}_1 + \cdots + \alpha t_n\vec{v}_n) \\ &= \alpha t_1\vec{w}_1 + \cdots + \alpha t_n\vec{w}_n \\ &= \alpha(t_1\vec{w}_1 + \cdots + t_n\vec{w}_n) \\ &= \alpha L(t_1\vec{v}_1 + \cdots + t_n\vec{v}_n). \end{aligned}$$

To see L is injective, suppose $L(t_1\vec{v}_1 + \cdots + t_n\vec{v}_n) = t_1\vec{w}_1 + \cdots + t_n\vec{w}_n = \vec{0}$. Then since $\{\vec{w}_1, \ldots, \vec{w}_n\}$ is linearly independent, we must have $t_1 = \cdots = t_n = 0$ so $t_1\vec{v}_1 + \cdots + t_n\vec{v}_n = \vec{0}$ and $\text{Null}(L) = \{\vec{0}\}$. Finally, the rank-nullity theorem implies $\text{rank}(L) = \dim(\mathbb{V}) = \dim(\mathbb{W})$ so $\text{Range}(L) = \mathbb{W}$ and L is an isomorphism. ■

This is an incredibly powerful theorem. We immediately know that any two 7-dimensional vector spaces over $\mathbb{C}$, for example, are isomorphic. Furthermore, to find an isomorphism, we simply have to choose bases for both vector spaces and the map that appears in the proof of the theorem will be an isomorphism.

Lecture 14 - July 20

6 Review - Determinants and matrix inverses

The long term goal in this course is to understand linear maps. Since all linear maps are matrices (and vice versa), one way to gain insight into linear maps is to gain insight into matrices. One of those insights is being able to compute and interpret the determinant of a matrix.

While it is expected that you have seen the material in the section in a previous course, we will recall the appropriate definitions and results here without proof. It would be excellent review for you to prove all the facts in this section yourself. It only makes sense to talk about determinants and inverses of matrices when the matrices in question are square. Because of this, we will assume all matrices are square in this section.

6.1 Cofactor Expansion

There are many ways to define the determinant of a matrix. We will define it as the number obtained by cofactor expansion.

Definition. Let A be a $n \times n$ matrix. Let $A_{i,j}$ denote the $(n-1)\times(n-1)$ submatrix obtained from A by deleting the i-th row and j-th column. The **cofactor** of a_{ij} is defined to be $C_{ij} = (-1)^{i+j} \det A_{i,j}$.

Definition. The **determinant** of an $n \times n$ matrix A is defined to be

$$\det A := a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

and we define $\det[a] = a$. We sometimes write $\det A = |A|$.

Computing the determinant this way is called **cofactor expansion** along the first row.

Fact 21. *Suppose A is an $n \times n$ matrix. Then the determinant is given by cofactor expansion along any row or column. That is*

$$\begin{aligned} \det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}, \qquad \textit{and} \\ \det A &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}, \end{aligned}$$

for all $1 \leq i, j \leq n$.

Using this last fact to compute the determinant would be called **cofactor expansion** along the ith row or jth column. Each time we perform cofactor expansion we arrive at a bunch of matrices that are smaller. We continue this until we are left with a bunch of 1×1 matrices, at which point we compute the determinant by just taking the value of the entry in each matrix.

Example. Let $A = \left[\begin{smallmatrix} 2 & 1 & 5 \\ 4 & 3 & -1 \\ 0 & 1 & -2 \end{smallmatrix}\right]$. Then if we perform cofactor expansion across the first row we get

$$\begin{aligned} \det(A) &= 2\det\left(\begin{bmatrix} 3 & -1 \\ 1 & -2 \end{bmatrix}\right) - 4\det\left(\begin{bmatrix} 1 & 5 \\ 1 & -2 \end{bmatrix}\right) + 0\det\left(\begin{bmatrix} 1 & 5 \\ 3 & -1 \end{bmatrix}\right) \\ &= 2(-6+1) - 4(-2-5) \\ &= 18. \end{aligned}$$

The following facts are left unproven since you should have seen them in a previous course. However, we will rely on these facts heavily for the rest of the course.

Fact 22. *Let A and B be $n \times n$ matrices with entries in $\mathbb{F}$.*

1. *If one of A's rows or columns is 0, then $|A| = 0$. This is a special case of the next two facts.*
2. *$|A| = 0$ if and only if the columns of A are linearly dependent as vectors in $\mathbb{F}^n$.*
3. *$|A| = 0$ if and only if the rows of A are linearly dependent as vectors in $\mathbb{F}^n$.*
4. *If A is an upper or lower triangular matrix, then the determinant of A is the product of the diagonal entries.*
5. *$|A| = \left|A^T\right|$.*
6. *$|AB| = |A|\,|B|$.*

6.2 Elementary Row Operations and the Determinant

Every row operation has a matrix associated to it, called the elementary matrix. If E is the elementary matrix for a certain row operation, then this matrix has the property that if B is obtained from A by that particular row operation, then $EA = B$. Using this and the fact that $|AB| = |A|\,|B|$, we can figure out what happens to the determinant of a matrix when row operations are performed.

Proposition 23. *The elementary matrix for a row operation is obtained by applying the row operation to the identity matrix.*

For example, if we perform the elementary row operation $R_2 \to R_2 - 2R_2$ to the 3×3 identity matrix we get $\left[\begin{smallmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]$. Let's see that this matrix actually does the row operation for us. We have, for example,

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 7 \\ 2 & 1 & 2 \\ 5 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 7 \\ 0 & -5 & -12 \\ 5 & 1 & 4 \end{bmatrix}.$$

So, we can think of an elementary row operation simply as left-multiplying by some matrix. Using this, along with the fact that $|AB| = |A|\,|B|$ we can prove the following proposition.

Proposition 24. *Suppose A is an $n \times n$ matrix and B is obtained from A by an elementary row operation. Then*

- *If the row operation takes the form $R_k \to R_k + tR_j$ for any $t \in \mathbb{F}$ and for rows $k \neq j$, then $\det A = \det B$.*
- *If the row operation is multiplying a row by t, then $t\,|A| = |B|$.*
- *If the row operation is given by switching any two rows, then $|A| = -\,|B|$.*

Proof. Exercise. ∎

6.3 Invertibility

We have already seen matrix inverses pop up when talking about isomorphisms and change of coordinate matrices. Let's look at them some more from the perspective of matrices.

Definition. An $n \times n$ matrix A is **invertible** if there exists a matrix A^{-1} such that $A^{-1}A = AA^{-1} = I$, where I is the $n \times n$ identity matrix. We call A^{-1} the **inverse** of A.

Definition. For any matrix A, define the **rank** of A, denoted $\operatorname{rank}(A)$, as the number of leading 1s in its row reduced form.

Note that calling this the rank of the matrix makes sense, because if A is the matrix corresponding to some linear map L, then $\operatorname{rank}(A) = \operatorname{rank}(L)$.

We finish this quick review of invertability by recalling the following important fact.

Fact 25. *Let A be an $n \times n$ matrix. The following are equivalent.*

1. $\det A \neq 0$.
2. $\operatorname{rank}(A) = n$.
3. *A is invertible.*

Fact 26.
- *If A is invertible, A^{-1} is unique.*
- $\left|A^{-1}\right| = |A|^{-1}$.

There are a couple of ways to compute the inverse of a matrix. One trick to keep in mind is a formula for the inverse of a 2×2 matrix. If $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ and A is invertible, then $|A| = ad - bc \neq 0$. The inverse is given by

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

However, if the matrix is larger, it's not worth memorising a formula (which does exist). In this case, we can always obtain the inverse by row-reducing A augmented with the identity matrix. Since A is invertible, the first half of the matrix will end up as the identity, and you will be left with the inverse in the second half. That is

$$[A \mid I] \sim [I \mid A^{-1}]$$

where I is the $n \times n$ identity matrix. For example, suppose we want to find the inverse of $A = \left[\begin{smallmatrix} 2 & 1 & 5 \\ 4 & 3 & -1 \\ 0 & 1 & -2 \end{smallmatrix}\right]$ we perform the following row-reduction:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 5 & 1 & 0 & 0 \\ 4 & 3 & -1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{5}{18} & \frac{7}{18} & -\frac{8}{9} \\ 0 & 1 & 0 & \frac{4}{9} & -\frac{2}{9} & \frac{11}{9} \\ 0 & 0 & 1 & \frac{2}{9} & -\frac{1}{9} & \frac{1}{9} \end{array}\right].$$

Therefore we have

$$A^{-1} = \frac{1}{18}\begin{bmatrix} -5 & 7 & -16 \\ 8 & -4 & 22 \\ 4 & -2 & 2 \end{bmatrix}.$$

6.4 Determinants and Volume

So far, we only seem to care whether or not a determinant is 0. Here is a geometric interpretation of the actual value of the determinant.

Definition. Given n vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ in $\mathbb{R}^n$, we can talk about the volume of the **parallelotope** defined by these n vectors. This volume is the n-dimensional volume of the n-dimensional shape with vertices given by the sum of every subset of $\{\vec{v}_1, \dots, \vec{v}_n\}$. Denote this volume by $\mathrm{vol}\{\vec{v}_1, \dots, \vec{v}_n\}$.

For example, in $\mathbb{R}^2$, the $\mathrm{vol}(\{\vec{v}_1, \vec{v}_2\})$ is the volume (in this case area) of the parallelogram with vertices $\vec{0}$, $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_1 + \vec{v}_2$.

In $\mathbb{R}^3$, the $\mathrm{vol}(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\})$ is the volume of the parallelopiped with vertices $\{\vec{0}, \vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_3, \vec{v}_2 + \vec{v}_3, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$.

Fact 27. *Let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be n vectors in $\mathbb{R}^n$. Then* $\mathrm{vol}(\{\vec{v}_1, \dots, \vec{v}_n\}) = \left|\det \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}\right|$.

Fact 28. *Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map, and let A be the matrix of L with respect to the standard bases. Let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be n vectors in $\mathbb{R}^n$. Then* $\mathrm{vol}(\{L(\vec{v}_1), \dots, L(\vec{v}_n)\}) = |\det A| \, \mathrm{vol}(\{\vec{v}_1, \dots, \vec{v}_n\})$.

7 Eigenvectors and Diagonalization

As hinted to before, sometimes the standard basis is not the best basis with which to study a particular problem, or a linear map. For example, consider the linear map $L : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{2(x+y+z)}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

If $\mathcal{S}$ is the standard basis for $\mathbb{R}^3$, then you can check that

$$[L]_{\mathcal{S}} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Looking at this matrix, I do not really have any idea what this linear map is doing, geometrically or otherwise. However, if we look at the basis

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\},$$

then it can be checked that

$$[L]_{\mathcal{B}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Staring at this matrix, we can easily interpret what the linear map is doing. It is a reflection, negating the $\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)$ direction, and keeping the 2-dimensional subspace spanned by $\left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix}\right)$ unchanged.

This example shows us that sometimes looking at a particular problem with the right set of coordinates can prove enlightening. So, with this in mind, the following natural question arises:

Given a linear map from a vector space to itself, how can we find an "enlightening" basis with which to view the linear map?

It would be nice to find vectors which are not rotated, but simply scaled when the linear map is applied to it. That is, we'd like to find vectors $\vec{v}$ such that $L(\vec{v}) = \lambda\vec{v}$ for some $\lambda \in \mathbb{F}$. If we can find a basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ of $\mathbb{V}$ such that $L(\vec{v}_i) = \lambda_i\vec{v}_i$ for every i, then with respect to $\mathcal{B}$ we would have

$$[L]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

Unfortunately, as well will see, we cannot always find such a basis. Let's try anyway!

Definition. Let $L : \mathbb{V} \to \mathbb{V}$ be a linear map. A non-zero vector $\vec{v} \in \mathbb{V}$ such that $L(\vec{v}) = \lambda\vec{v}$ for some $\lambda \in \mathbb{F}$ is called an **eigenvector** of L. The number λ is called an **eigenvalue** of L.

Definition. Let $L : \mathbb{V} \to \mathbb{V}$ be a linear map, and let λ be an eigenvalue of L. Define the **eigenspace** of L corresponding to λ to be $\{\vec{v} \in \mathbb{V} : L(\vec{v}) = \lambda\vec{v}\}$.

Proposition 29. *Let $L : \mathbb{V} \to \mathbb{V}$ be a linear map, and let λ be an eigenvalue of L. The eigenspace corresponding to λ is a subspace of $\mathbb{V}$.*

Proof. Let $\mathbb{W}$ denote the eigenspace corresponding to λ. Then since $L(\vec{0}) = \vec{0} = \lambda\vec{0}$, $\vec{0} \in \mathbb{W}$. Let $\vec{v}, \vec{w} \in \mathbb{W}$. Then

$$L(\vec{v} + \vec{w}) = L(\vec{v}) + L(\vec{w}) = \lambda\vec{v} + \lambda\vec{w} = \lambda(\vec{v} + \vec{w})$$

so $\vec{v} + \vec{w} \in \mathbb{W}$. Finally, let $t \in \mathbb{F}$. Then

$$L(t\vec{v}) = tL(\vec{v}) = t\lambda\vec{v} = \lambda(t\vec{v})$$

so $t\vec{v} \in \mathbb{W}$. Since $\mathbb{W}$ is non-empty, closed under addition, and closed under scalar multiplication, $\mathbb{W}$ is a subspace of $\mathbb{V}$ by the subspace test. ∎

Whenever you see new abstract definitions like this, the best way to understand the definition is to apply it to a specific situation in your mind. This is the whole point of going through examples.

Example. Let $D : \mathcal{P}_4(\mathbb{C}) \to \mathcal{P}_4(\mathbb{C})$ be the differentiation map. Then 0 is an eigenvalue of D since $D(3) = 0 = 0(3)$, and 3 is not the zero vector in $\mathcal{P}_4(\mathbb{C})$. Furthermore, 0 is the only eigenvalue. You can see this by noticing that λp and p have the same degree if and only if $\lambda \neq 0$. So, since $D(p)$ and p never have the same degree (unless $p = 0$ of course), then the only way $D(p) = \lambda p$ can be true is if $\lambda = 0$.

Now let's work out what the eignespace corresponding to 0, which we will denote $\mathbb{W}$, looks like. By the definition of eigenspace we have

$$\mathbb{W} = \{p \in \mathcal{P}_4(\mathbb{C}) : D(p) = 0\}$$

so it is not too hard to convince yourself that $\mathbb{W} = \{p \in \mathcal{P}_4(\mathbb{C}) : p = k \text{ for some constant } k \in \mathbb{C}\}$.

Example. Consider the liner map $L : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right) = \begin{pmatrix} a \\ 2b \\ -c \end{pmatrix}.$$

Then $1, 2$, and -1 are eigenvalues of L since

$$L\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) = 1\begin{pmatrix}1\\0\\0\end{pmatrix}, \quad L\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right) = 2\begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \text{and} \quad L\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right) = -1\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

As an exercise, prove that the eigenspaces corresponding to $1, 2$, and -1 are $\operatorname{Span}\left(\left\{\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)\right\}\right)$, $\operatorname{Span}\left(\left\{\left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right)\right\}\right)$, and $\operatorname{Span}\left(\left\{\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right)\right\}\right)$ respectively.

Lecture 15 - July 23

So, it's all well and good to make definitions like this, and do examples where it's easy to stare at it to work out what the eigenvalues and eigenspaces are, but how can we actually find eigenvalues and eigenspaces in general? As is becoming a pattern, we pick a basis $\mathcal{B}$ of $\mathbb{V}$, turn our linear map into the matrix $[L]_{\mathcal{B}}$, and harness the computational power of matrices!

Once we've picked a basis, we can think of these definitions purely as definitions for matrices. In this case, we can think of a square matrix as a linear map from $\mathbb{F}^n$ to itself, and column matrices as vectors in $\mathbb{F}^n$.

Definition. Let A be an $n \times n$ matrix. A non-zero vector $\vec{v} \in \mathbb{F}^n$ is called an **eigenvector** of A if $A\vec{v} = \lambda\vec{v}$ for some $\lambda \in \mathbb{F}$. The scalar λ is called an **eigenvalue**.

Definition. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. Define the **eigenspace** of A corresponding to λ to be $\{\vec{v} \in \mathbb{F}^n : A\vec{v} = \lambda\vec{v}\}$.

Example. Consider the matrix

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

Then we can see that $2, i$, and -1 are eigenvalues for A. Furthermore, since

$$\begin{bmatrix}2 & 0 & 0\\0 & i & 0\\0 & 0 & -1\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}2a\\ib\\-c\end{bmatrix}$$

we see the only way $A\vec{v} = \lambda\vec{v}$ is if at most one of a, b, c are not zero, in which case λ must be $2, i$, or -1. Therefore the only eigenvalues of A are $2, i$, and -1.

The eigenspace corresponding to an eigenvalue λ is also the set of all eigenvectors corresponding to λ along with $\vec{0}$.

7.1 Finding Eigenvectors and Eigenvalues

To find eigenvectors and eigenvalues for a linear map, first pick a basis so you have an $n \times n$ matrix. Now the problem becomes finding eigenvalues and eigenvectors for a square matrix.

Let's give it a shot. To find an eigenvector, we're looking for a vector $\vec{v} \neq \vec{0}$ such that

$$A\vec{v} = \lambda\vec{v}$$

so if we rearrange this equation we get

$$A\vec{v} - \lambda\vec{v} = \vec{0}.$$

It would be tempting now to factor out the $\vec{v}$, which we will do, but we cannot as written. If we did, we would be left with a term $A - \lambda$, which makes no sense since A is a square matrix and λ is an element of $\mathbb{F}$. To get around this, we observe that $\lambda\vec{v} = \lambda I\vec{v}$ where I is the identity matrix of the appropriate size. Now our equation takes the form

$$(A - \lambda I)\vec{v} = \vec{0}.$$

If the matrix $A - \lambda I$ were invertible, then we could multiply both sides on the left by the inverse and get $\vec{v} = \vec{0}$. Since we're looking for non-zero vectors $\vec{v}$, this means we are looking for values of λ that make the matrix $A - \lambda I$ not invertible. Equivalently, we want values of λ such that $\det(A - \lambda I) = 0$. Furthermore, once we've found such a lambda, a corresponding eigenvector is any non-zero vector such that $(A - \lambda I)\vec{v} = \vec{0}$, which must exist because $\det(A - \lambda I) = 0$. Let's see this in practice.

Example. Consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}.$$

Then

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{vmatrix} \\ &= -\lambda(-3-\lambda) + 2 \\ &= (\lambda+1)(\lambda+2), \end{aligned}$$

therefore the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = -2$. Now we will find the eigenspaces corresponding to both λ_1 and λ_2.

$\lambda_1 = -1$: We want to find the nullspace of $A - (-1)I$. We have

$$A + I = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

So to find the nullspace we treat this as the coefficient matrix for a system of equations (a homogeneous one, meaning all equations are equal to 0) and solve. If we let x_1 and x_2 be the variables, we let $x_2 = t$ and then $x_1 = -t$. Therefore the nullspace is given by

$$\mathrm{Null}(A - (-1)I) = \left\{ \begin{bmatrix} -t \\ t \end{bmatrix} : t \in \mathbb{F} \right\}.$$

So $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$, for example, is an eigenvector corresponding to $\lambda_1 = -1$.

$\lambda_2 = -2$: Repeating the process we have

$$A + 2I = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix}.$$

Computing the nullspace gives the corresponding eigenspace as

$$\mathrm{Null}(A + 2I) = \left\{ t \begin{bmatrix} -1 \\ 2 \end{bmatrix} : t \in \mathbb{F} \right\}.$$

So we have found all the eigenvalues and the corresponding eigenspaces.

Remember, whenever we're finding the eigenvalues and eigenvectors of a matrix, the matrix is some matrix of a linear map with respect to some basis. You mustn't forget about what all your computations mean.

The determinant $\det(A - \lambda I)$ is a polynomial in λ, and it tells us a surprising amount about a matrix (and thus about the corresponding linear map). Because of this we give it a special name.

Definition. Let A be an $n \times n$ matrix. The **characteristic polynomial** of A is the polynomial in λ given by $\det(A - \lambda I)$.

So for example, the characteristic polynomial of $\left[\begin{smallmatrix} 0 & 1 \\ -2 & -1 \end{smallmatrix}\right]$ is $\lambda^2 + 3\lambda + 2$.

Let's formally prove that what we did above with the 2×2 matrix is legitimate.

Proposition 30. *Let A be an $n \times n$ matrix. The eigenvalues of A are the values of λ that are solutions to the equation* $\det(A - \lambda I) = 0$. *That is, they are the roots of the characteristic polynomial of A.*

Proof. Suppose λ is an eigenvalue of $\mathbb{F}$, that is, there is some $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \lambda\vec{v}$. Rearranging gives $(A - \lambda I)\vec{v} = \vec{0}$. Therefore $\dim(\mathrm{Null}(A - \lambda I))) \geq 1$, so we must have $\mathrm{rank}(A - \lambda I) < n$ and $\det(A - \lambda I) = 0$. Conversely, suppose $\det(A - \lambda I) = 0$. Then $\mathrm{rank}(A) < n$ so $\dim(\mathrm{Null}(A)) \geq 1$ so there is some non-zero $\vec{v} \in \mathbb{F}^n$ such that $(A - \lambda I)\vec{v} = \vec{0}$. Rearranging gives $A\vec{v} = \lambda\vec{v}$ so λ is an eignevalue of A. ■

A neat little consequence of the proposition is that the determinant of a matrix can be related to the eigenvectors. By thinking carefully about polynomials and how the roots relate to the coefficients, you can prove the following corollary.

Corollary 31. *The determinant of a matrix is the product of its eigenvalues.*

Proof. Exercise. ■

The previous proposition proves that our method fo finding the eigenvalues is correct, the next proves the method for finding eigenspaces is correct.

Proposition 32. *Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A. The eigenspace corresponding to λ is equal to* $\mathrm{Null}(A - \lambda I)$.

Proof. Notice the eigenspace corresponding to λ is

$$\{\vec{v} \in \mathbb{F}^n : A\vec{v} = \lambda\vec{v}\} = \left\{\vec{v} \in \mathbb{F}^n : A\vec{v} - \lambda I\vec{v} = \vec{0}\right\} = \left\{\vec{v} \in \mathbb{F}^n : (A - \lambda I)\vec{v} = \vec{0}\right\} = \mathrm{Null}(A - \lambda I)$$

completing the proof. ■

Example. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Let's find all the eigenvalues and bases for the corresponding eigenspaces. We first compute $\det(A-\lambda I)$. We have

$$\begin{aligned}
|A-\lambda I| &= \begin{vmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 3-\lambda & 3-\lambda & 3-\lambda \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} \\
&= (3-\lambda)\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} \\
&= (3-\lambda)\begin{vmatrix} 1 & 1 & 1 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} \\
&= (3-\lambda)\lambda^2.
\end{aligned}$$

During this manipulation, we performed various row operations and kept track of how that affected the determinant.

We now have that $\lambda_1 = 3$ and $\lambda_2 = 0$ are all the eigenvalues of A.

Lecture 16 - July 24

Now, to find bases for each eigenspace we must find bases for the nullspaces of $A - \lambda_1 I$ and $A - \lambda_2 I$.

For $\lambda_1 = 3$ we have

$$A - 3I = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

so a basis for the eigenspace corresponding to the eigenvalue 3 is

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

For $\lambda_2 = 0$ we have

$$A - 0I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so a basis for $\mathrm{Null}(A - 0I)$ is

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Let's shift our attention to two facts about characteristic polynomials of $n \times n$ matrices. After a couple of examples, the first fact should not be surprising, and can be proved by carefully unwrapping how cofactor expanion works. As a hint, the determinant of an $n \times n$ matrix is a polynomial in the entries of the matrices, and each term consist of the product of exactly n entries.

Fact 33. *If A is an $n \times n$ matrix, then the characteristic polynomial is a polynomial of degree n.*

Once we have established that the characteristic polynomial is degree n, then we can use the fundamental theorem of algebra to conclude that, when counted correctly, every $n \times n$ matrix has exactly n eigenvalues.

Fact 34 (Fundamental Theorem of Algebra)**.** *Every degree n polynomial factors as $a(x-c_1)^{k_1}(x-c_2)^{k_2}\cdots(x-c_m)^{k_m}$ where $c_1, \ldots, c_m \in \mathbb{C}$ are distinct, $k_1 + \cdots + k_m = n$ and a is some non-zero complex number. The c_i are the* **roots** *of the polynomial and k_i is the* **multiplicity** *of the root c_i.*

Another way to phrase this fact is to say that every degree n polynomial has n roots over the complex numbers, when they are counted with multiplicity.

7.2 Diagonalisation

This section will take place with $\mathbb{F} = \mathbb{C}$, and we will make a comment at the end about diagonalisation over $\mathbb{R}$.

Let's revisit the following example from assignment 2. Consider the linear map $L : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{2(x+y+z)}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

With respect to the standard basis $\mathcal{S} = \left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$ and the bases $\mathcal{B} = \left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right\}$ we have

$$[L]_{\mathcal{S}} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \quad \text{and} \quad [L]_{\mathcal{B}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Furthermore we note $({}_{\mathcal{S}}\mathcal{P}_{\mathcal{B}})^{-1}[L]_{\mathcal{S}}{}_{\mathcal{S}}\mathcal{P}_{\mathcal{B}} = [L]_{\mathcal{B}}$. So we see we can change bases to make $[L]_{\mathcal{S}}$ into the diagonal matrix $[L]_{\mathcal{B}}$.

With this in mind, we call a matrix diagonalisable if we can change the basis in question to obtain a diagonal matrix. Or, more precisely:

Definition. A square matrix A is **diagonalisable** if there exists an invertible matrix P such that $P^{-1}AP = D$ where D is a diagonal matrix.

Definition. If A and B are $n \times n$ matrices such that $P^{-1}AP = B$ for some invertible $n \times n$ matrix P, then we say A and B are **similar** matrices.

This definition is motivated by the fact that A and B are two matrix representatives of the same linear map, and P is the change of basis matrix.

Theorem 35. *If A and B are similar matrices, then they have the same determinant, same eigenvalues, same rank, and same trace.*

Proof. The fact that A and B have the same trace and same eigenvalues was proved in assignment 2. To see they have the same determinant we have

$$|A| = \left|P^{-1}BP\right| = |P|^{-1}|B||P| = |B||P|^{-1}|P| = |B|.$$

Since A and B are matrices of the same linear map L with respect to two different bases we have

$$\operatorname{rank}(L) = \operatorname{rank}(A) = \operatorname{rank}(B)$$

completing the proof. ∎

Lecture 17 - July 25

So the question now becomes, when is a matrix A diagonalisable? If we think of how the matrix with respect to a linear map works, then we wish to find a basis of eigenvectors.

Theorem 36. *An $n \times n$ matrix A is diagonalisable if and only if there exists a basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathbb{C}^n$ such that each $\vec{v}_i$ is an eigenvector for A. If such a basis exists, then $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ and $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$ where $\vec{v}_i$ is an eigenvector with eigenvalue λ_i.*

Proof. Suppose A is diagonalisable, that is $P^{-1}AP = D$ for some invertible P and diagonal $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. Let $P = [\, \vec{v}_1 \cdots \vec{v}_n \,]$. Then since $AP = PD$ we have

$$\begin{bmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \cdots & \lambda_n\vec{v}_n \end{bmatrix}.$$

Therefore the $\vec{v}_i$ are eigenvectors with eigenvalues λ_i. Furthermore, since P is invertible, $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a linearly independent subset of $\mathbb{C}^n$, and so it is a basis.

Conversely, if $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis of $\mathbb{C}^n$ such that $A\vec{v}_i = \lambda_i\vec{v}_i$ for all i then $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ is invertible and

$$\begin{bmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \cdots & \lambda_n\vec{v}_n \end{bmatrix}.$$

This implies $AP = PD$ so $P^{-1}AP = D$ completing the proof. ■

So, if we are to diagonalise a matrix, we need to find a basis for $\mathbb{C}^n$ consisting entirely of eigenvectors. Let's take a look at some examples that we've already explored.

Example. Let $A = \left[\begin{smallmatrix} 0 & 1 \\ -2 & -3 \end{smallmatrix}\right]$. Then we saw before that $\lambda_1 = -1$ and $\lambda_2 = -2$ are the eigenvectors, and bases for the eigenspaces are given by $\left\{\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]\right\}$ and $\left\{\left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]\right\}$ respectively. Since neither of these vectors are a scalar multiple of the other, we see that $\left\{\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]\right\}$ is a basis for $\mathbb{C}^2$.

Therefore, by Theorem 36 we know A is diagonalisable. In fact, we must have $P^{-1}AP = D$ where

$$P = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}.$$

In fact, we can just check this. We have

$$AP = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & -4 \end{bmatrix}$$

and

$$PD = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & -4 \end{bmatrix}$$

so $P^{-1}AP = D$.

Example. Consider the matrix $A = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$. Then we saw on assignment 2 that the basis $\left\{ \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix}\right) \right\}$ is a basis consisting of eigenvectors, with corresponding eigenvalues $-1, 1$, and 1 respectively. Therefore A is diagonalisable and $P^{-1}AP = D$ where

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that in this example, the characteristic polynomial of A is $-(\lambda+1)(\lambda-1)^2$, and the dimension of the eigenspace corresponding to 1 is 2. So even though we only have 2 eigenvalues, we still get three linearly independent eigenvectors, since $\lambda = 1$ contributes two of them.

Example. Consider the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, which has characteristic polynomial $-(\lambda - 3)\lambda^2$. We saw earlier that bases for the eigenspaces corresponding to 0 and 3 are

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

respectively. Since the first two vectors form a basis for the eigenspace corresponding to 0, they are linearly independent. We don't know that adding the third vector would give us a basis, but you can check that the collection of all three vectors is linearly independent, so

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is a basis of $\mathbb{C}^3$ consisting entirely of eigenvectors. Therefore by Theorem 36, A is diagonalisable and $P^{-1}AP = D$ where

$$P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Again we see that even though there are only two distinct eigenvalues, 0 appears with multiplicity 2 and contributes two basis vectors. Coincidence? Maybe.

It appears that as long as we have enough basis vectors in total from each eigenspace, then combining them gives a linearly independent subset of $\mathbb{C}^n$, so if there are n of them, we will get a basis. This will be true, but it needs proving.

Another observation to make is that although each eigenvalue contributes at least one basis vector (since $\dim(\text{Null}(A - \lambda I)) \geq 1$ for all eigenvalues λ), we may not have enough vectors to make up a basis, especially if there are not n distinct eigenvalues. However, it appears that in cases when we don't have n distinct eigenvalues, the eigenvalues with multiplicity greater than 1 contribute more than one vector to the basis, that is, the dimension of the eigenspace is greater than 1. In fact it appears that the dimension of the eigenspace is equal to the multiplicity of the eigenvalue as a root of the characteristic equation.

So, let's focus are attention to answering the following two questions:

1. Is it true that if I combine basis vectors from different eigenspaces, then the new set of vectors is linearly independent?

2. Is it true that the dimension of the eigenspace corresponding to an eigenvalue λ is equal to the multiplicity of λ as a root of the characteristic polynomial?

From the examples so far, it would appear that the answer to both these questions is 'yes'. However, while all this speculation is all well and good (and we'll see that some of it is wrong), we need to prove things!

Lecture 18 - July 30

Let's start with the first statement that combining bases for eigenspaces gives a linearly independent subset. We will prove this in a few steps.

Lemma 37. *Suppose $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of a square matrix A with corresponding eigenvectors $\vec{v}_1, \ldots, \vec{v}_k$. Then $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent.*

Proof. We will prove this by induction on k. For $k = 1$, since $\vec{v}_1$ is an eigenvector, it is non-zero and $\{\vec{v}_1\}$ is a linearly independent set.

For the induction, suppose that if we have $k-1$ eigenvectors corresponding to $k-1$ distinct eigenvalues, then the eigenvectors are linearly independent. Suppose now

$$t_1\vec{v}_1 + \cdots + t_k\vec{v}_k = \vec{0}. \tag{1}$$

We wish to show all the t_i are equal to 0. Multiplying equation (1) by λ_k gives

$$t_1\lambda_k\vec{v}_1 + \cdots + t_k\lambda_k\vec{v}_k = \vec{0}. \tag{2}$$

On the other hand, left multiplying equation (1) by A gives

$$\vec{0} = t_1A\vec{v}_1 + \cdots + t_kA\vec{v}_k = t_1\lambda_1\vec{v}_1 + \cdots + t_k\lambda_k\vec{v}_k. \tag{3}$$

Subtracting equation (3) from (2) gives

$$t_1(\lambda_k - \lambda_1)\vec{v}_1 + \cdots + t_{k-1}(\lambda_k - \lambda_{k-1})\vec{v}_{k-1} = \vec{0}.$$

The inductive hypothesis gives us that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is linearly independent. Therefore we must have $t_i(\lambda_k - \lambda_i) = 0$ for all $i = 1, \ldots, k-1$. Since $\lambda_k \neq \lambda_i$ for all $i \neq k$, we must have that $t_i = 0$ for all $i = 1, \ldots, k-1$. Substituting this back into equation (3) yields $t_k\vec{v}_k = \vec{0}$. Since $\vec{v}_k$ is an eigenvector, it is non-zero so $t_k = 0$. Alas, we have $t_1 = \cdots = t_k = 0$ so $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent.

Therefore, by the principal of mathematical induction, the lemma is true. ■

Lemma 37 doesn't quite give us what we want yet. For example, suppose we have a 3×3 matrix with eigenvalues 1 and 2. Suppose $\{\vec{v}_1, \vec{v}_2\}$ is a basis for the eigenspace corresponding to 1, and $\{\vec{v}_3\}$ is a basis for the eigenspace corresponding to 2. We are hoping that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set, and thus a basis for $\mathbb{C}^3$. However, Lemma 37 only tells us $\{\vec{v}_1, \vec{v}_3\}$ and $\{\vec{v}_2, \vec{v}_3\}$ are linearly independent sets. Even with the fact that $\{\vec{v}_1, \vec{v}_2\}$ is linearly independent, this is not enough to conclude that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent (even though it will turn out to be true). As an exercise, try to find three vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ in $\mathbb{C}^3$ such that $\{\vec{v}_1, \vec{v}_2\}$, $\{\vec{v}_1, \vec{v}_3\}$ and $\{\vec{v}_2, \vec{v}_3\}$ are linearly independent, but $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly dependent.

While Lemma 37 isn't quite the result we are after, it will be an important step. Also, it does give us the following useful theorem.

Theorem 38. *If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalisable.*

Proof. Suppose $\lambda_1, \ldots, \lambda_n$ are the distinct eigenvalues. Let $\vec{v}_1, \ldots, \vec{v}_n$ be eigenvectors corresponding to the eigenvalues. By Lemma 37, $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are linearly independent, and therefore a basis for $\mathbb{C}^n$. By Theorem 36, A is diagonalisable. ■

So, for example, suppose we have some 4×4 matrix with characteristic polynomial $(\lambda - i)(\lambda + i)(\lambda - 1)(\lambda - (3+i))$. Since this polynomial has 4 distinct roots, A has 4 distinct eigenvalues and it is diagonalisable. Furthermore, we know the diagonal matrix D to which A is similar is given by

$$D = \begin{bmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3+i \end{bmatrix}.$$

The next lemma is an analogue to Lemma 37, except it is extended to deal with the case that we have multiple linearly independent vectors in each eigenspace.

Lemma 39. *Suppose $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of an $n \times n$ matrix A, and let $\{\vec{v}_{i,1}, \vec{v}_{i,2}, \ldots, \vec{v}_{i,m_i}\}$ be a basis for the eigenspace corresponding to λ_i (so the dimension of the eigenspace corresponding to λ_i is m_i). Then*

$$\{\vec{v}_{1,1}, \vec{v}_{1,2}, \ldots, \vec{v}_{1,m_1}, \vec{v}_{2,1}, \ldots, \vec{v}_{2,m_2}, \ldots, \vec{v}_{k,1}, \ldots, \vec{v}_{k,m_k}\}$$

is a linearly independent subset of $\mathbb{C}^n$.

Proof. Consider the equation

$$\sum_{i=1}^{k} \sum_{j=1}^{m_i} t_{i,j} \vec{v}_{i,j} = \vec{0}. \tag{4}$$

where each $t_{i,j} \in \mathbb{C}$. We wish to show this can only happen when all the $t_{i,j}$ are 0. Let

$$\vec{w}_i = \sum_{j=1}^{m_i} t_{i,j} \vec{v}_{i,j}.$$

and note that $\vec{w}_i$ is in the eigenspace corresponding to λ_i. Since $\{\vec{v}_{i,1}, \vec{v}_{i,2}, \ldots, \vec{v}_{i,m_i}\}$ is a linearly independent set for all i, $\vec{w}_i$ is uniquely written as a linear combination of $\{\vec{v}_{i,1}, \vec{v}_{i,2}, \ldots, \vec{v}_{i,m_i}\}$. Therefore if $\vec{w}_i = \vec{0}$, we can conclude $t_{i,j} = 0$ for all j. Therefore it suffices to show $\vec{w}_i = \vec{0}$ for all i.

Substituting the $\vec{w}_i$ into equation (4) gives

$$\vec{w}_1 + \cdots + \vec{w}_k = \vec{0}.$$

Suppose that not all the $\vec{w}_i$ are equal to $\vec{0}$, so some are eigenvectors. Then the set consisting of the $\vec{w}_i$ that are eigenvectors is not linearly independent, contradicting Lemma 37. Therefore all the $\vec{w}_i$ are equal to $\vec{0}$, completing the proof. ■

Lecture 19 - July 31 (Review lecture)

Lecture 20 - August 1

Lemma 39 says the following. If fix bases for all the eigenspaces, and take the collection of all those vectors together, that set is linearly independent. So, the only thing standing between us and diagonalising a matrix is making sure that the sum of the dimensions of the eigenspaces is n.

Since each eigenspace is at least 1-dimensional, if we have n distinct eigenvalues we would have n distinct eigenspaces, giving n linearly independent eigenvectors and our matrix would be diagonalisable. This is the content of Theorem 38.

The only potential issues come when we have roots of the characteristic polynomial with multiplicity at least 2. In these cases, we would hope that if a root had multiplicity k, then the dimension of the corresponding eigenspace would be k. This would allow us to always find n linearly independent eigenvectors. Unfortunately, the next example shows that this is not always the case.

Example. Consider the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The characteristic polynomial is $(1-\lambda)^2$ so there is only one eigenvalue, $\lambda = 1$. To find a basis for the eigenspace we have

$$A - I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

so a basis for the eigenspace is given by $\{\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]\}$. Since this is 1-dimensional and it is the only eigenspace, we cannot find 2 linearly independent eigenvectors so A is not diagonalisable.

While we have just seen that the multiplicity of an eigenvalue as a root of the characteristic polynomial is not equal to the dimension of the corresponding eigenspace, there is a relationship between the two quantities. Let's first make some definitions.

Definition. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The **algebraic multiplicity** of λ is the multiplicity of λ as a root of the characteristic polynomial of A. The **geometric multiplicity** of λ is defined to be the dimension of $\mathrm{Null}(A - \lambda I)$.

We know the geometric multiplicity of any eigenvalue is at least 1, and not necessarily equal to the algebraic multiplicity. However, the algebraic multiplicity does act as an upper bound on the geometric multiplicity.

Proposition 40. *Let λ be an eigenvalue of an $n \times n$ matrix A. Then*

$$1 \leq \textit{geometric multiplicity of } \lambda \leq \textit{algebraic multiplicity of } \lambda.$$

Proof. The proof is an exercise that is walked through in the practice problems. ∎

So, if the geometric multiplicity is ever strictly less than the algebraic multiplicity for any eigenvalue, we immediately know we cannot find enough linearly independent eigenvectors to diagonalise the matrix. Conversely, if the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue, then we will be able to find enough linearly independent eigenvectors. We formalise this argument in the next theorem.

Theorem 41. *A matrix A is diagonalisable if and only if every eigenvalue of A has its geometric multiplicity equal to its algebraic multiplicity.*

Proof. Let A be an $n \times n$ matrix and suppose $a(\lambda - c_1)^{k_1} \cdots (\lambda - c_m)^{k_m}$ is the characteristic polynomial. If the algebraic multiplicity equals the geometric multiplicity for each eigenvalue, then the eigenspace corresponding to c_i has dimension k_i. Taking the basis vectors for each eigenspace gives $k_1 + \cdots + k_m = n$ linearly independent eigenvectors by Lemma 39, so A is diagonalisable by Theorem 36.

Conversely, suppose A is diagonalisable, and $P^{-1}AP = D$ where $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ and

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

By Theorem 36, $\vec{v}_i$ is an eigenvector with eigenvalue λ_i and $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly independent. Since A and D are similar, they have the same eigenvalues. Therefore we have that k_i of the diagonal entries of D are equal to c_i for all i. For each i, the k_i columns of P corresponding to c_i give k_i linearly independent eigenvectors in the eigenspace corresponding to c_i. Therefore the geometric multiplicity of c_i is at least equal to the algebraic multiplicity of c_i, so by Proposition 40, we have that they are equal. This completes the proof. ■

Remark. If you have a matrix A with real entries, you can ask whether or not there are matrices P and D with real entries such that $P^{-1}AP = D$. The complication in this case is that you may not even get n real eigenvectors. However, if you do, then all the results hold. If you don't, then there is no hope for diagonalisation!

Example. Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Let's try to figure out whether or not this matrix is diagonalisable. We start the only way we know how, by trying to find the eigenvalues. We have

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{vmatrix} \\ &= (1-\lambda)^2 + 1 \\ &= \lambda^2 - 2\lambda + 2. \end{aligned}$$

Staring at this polynomial, it's not clear what the roots are. However, it's a degree two polynomial so we have the quadratic formula at our disposal. The roots are

$$\begin{aligned} \lambda &= \frac{2 \pm \sqrt{4-8}}{2} \\ &= \frac{2 \pm 2i}{2} \\ &= 1 \pm i. \end{aligned}$$

Therefore the eigenvalues are $1+i$ and $1-i$. Since there are two distinct eigenvalues, and this is a 2×2 matrix, A is diagonalisable.

Note however that these eigenvalues are not in $\mathbb{R}$, so even though the entries of A are real, it is not diagonalisable over $\mathbb{R}$. That is, there is no real invertible matrix P and real diagonal matrix D such that $P^{-1}AP = D$.

Example. Consider the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. As a linear map from $\mathbb{R}^2$ to $\mathbb{R}^2$, this matrix corresponds to a 90 degree rotation counterclockwise with respect to the standard bases. Geometrically, we shouldn't expect this linear map to have any eigenvalues or eigenvectors since no vector in $\mathbb{R}^2$ is sent to a scalar multiplic of itself.

The characteristic polynomial is $\lambda^2 + 1$, so if we are working over $\mathbb{R}$, it is indeed the case that there are no eigenvalues or eigenvectors.

However, if we are working over $\mathbb{C}$, then $\lambda^2 + 1 = (\lambda - i)(\lambda + i)$ so there are eigenvalues and eigenvectors. In fact, since there are 2 distinct eigenvalues, A is diagonalisable over $\mathbb{C}$.

This is actually a really interesting example to think about, because multiplication by i (one of the eigenvalues) on $\mathbb{C}$ corresponds to rotation by 90 degrees counterclockwise.

Lecture 21 - August 2

7.3 Applications of Diagonalisation

We now know how to figure out whether or not a matrix is diagonalisable, and even better, how to find the invertible matrix P and the diagonal matrix D that actually perform the diagonalisation.

For this section, we'll take a look at two applications of diagonalising matrices.

Taking Powers of Matrices

Consider the matrix

$$A = \begin{bmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{bmatrix}.$$

If we are to compute A^2, that's easy enough, we just perform matrix multiplication. Same with A^3. After taking a few powers, we may start to get tired, and there's no way we're going to compute A^{100} by hand. We'll see now that if A is diagonalisable, then there's a neat little shortcut!

Suppose we have a diagonalisable matrix A, and we wish to take arbitrarily large powers of A. Then one way we can do this is by observing $A = PDP^{-1}$ for some diagonal matrix D so

$$A^n = (PDP^{-1})(PDP^{-1}) \cdots (PDP^{-1}) = PD(P^{-1}P)D(P^{-1}P) \cdots (P^{-1}P)DP^{-1} = PD^nP^{-1}.$$

We have reduced the problem to computing the nth power of a diagonal matrix, but this is much simpler than for an arbitrary matrix since if

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}$$

then

$$D^n = \begin{bmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_k^n \end{bmatrix}.$$

Let's see this in action.

Example. Suppose $A = \begin{bmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{bmatrix}$. Then $A = PDP^{-1}$ where

$$P = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then

$$\begin{aligned} A^{30} &= PD^{30}P^{-1} \\ &= \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2^{30} & 0 \\ 0 & 0 & 2^{30} \end{bmatrix} \begin{bmatrix} 3 & -3 & 1 \\ -3 & 4 & -1 \\ -1 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 3 - 2 \cdot 2^{30} & -3 + 3 \cdot 2^{30} & 1 - 2^{30} \\ 3 - 3 \cdot 2^{30} & -3 + 4 \cdot 2^{30} & 1 - 2^{30} \\ 3 + 3 \cdot 2^{30} & -3 - 3 \cdot 2^{30} & 1 \end{bmatrix} \end{aligned}$$

Taking powers of matrices like this arises when studying dynamical systems, and in an area of probability called Markov chains or the Markov process to name a couple. In particular the Google page rank system relies on taking large powers of matrices! If you're interested, check out the video "How does google google" on mathtube by Margot Gerritsen.

Decoupling Differential Equations

It often comes up when modelling nature, that you have a few differential equations where the variables from each equation appear in every other one. These can be extremely difficult to solve, but occasionally we can separate out the variables in a process called decoupling differential equations. It turns out that if your differential equations take on a very precise form, we can diagonalise a matrix and decouple the equations for us. Let's see this in an example.

Example. Xavier and Yvonne are in a zombie apocalypse, and they are continuously killing zombies, and the following things are true about the way they kill zombies.

- Both get better with practice (a reasonable assumption)
- Both slow down as the other kills zombies (they stop to congratulate the other person, and to give them a high-five if they are within slapping distance)
- For every zombie Xavier kills, his kill rate goes up by a factor of 5 and Evonne's goes down by a factor of 6.
- For every zombie Yvonne kills, her kill rate increases by a factor of 2 and Xavier's decreases by a factor of 6.

So, if we let x be the number of zombies killed by Xavier, y the number killed by Yvonne and t the time since the apocalypse started, we can set up the following system of differential equations that models the situation at hand.

$$\begin{aligned} \frac{dx}{dt} &= 5x - 3y \\ \frac{dy}{dt} &= -6x + 2y. \end{aligned}$$

Seemingly out of nowhere, let's make the following substitutions.

$$u = -\frac{2}{3}x + \frac{1}{3}y \quad \text{and} \quad w = \frac{1}{3}x + \frac{1}{3}y.$$

We now have

$$\begin{aligned}
\frac{du}{dt} &= -\frac{2}{3}\frac{dx}{dt} + \frac{1}{3}\frac{dy}{dt} \\
&= -\frac{2}{3}(5x - 3y) + \frac{1}{3}(-6x + 2y) \\
&= -\frac{16}{3}x + \frac{8}{3}y \\
&= 8u
\end{aligned}$$

and

$$\begin{aligned}
\frac{dw}{dt} &= \frac{1}{3}(5x - 3y) + \frac{1}{3}(-6x + 2y) \\
&= -\frac{1}{3}x - \frac{1}{3}y \\
&= -w.
\end{aligned}$$

These differential equations are much easier to deal with, and can be easily solved, and then converted back to our x and y variables. That's not the important thing here. The question you should have burning in your mind is "how did we choose u and w"? To answer this, we do what we do best, bring matrices into the picture!

Let

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \frac{d\vec{x}}{dt} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} 5 & -3 \\ -6 & 2 \end{bmatrix}.$$

Then

$$\frac{d\vec{x}}{dt} = A\vec{x}.$$

Now, it turns out that A is diagonalisable, in fact $P^{-1}AP = D$ where

$$P = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix}$$

so

$$\frac{d\vec{x}}{dt} = PDP^{-1}\vec{x}.$$

Now, if we let

$$\vec{u} = \begin{bmatrix} u \\ w \end{bmatrix} = P^{-1}\vec{x} = \begin{bmatrix} -\frac{2}{3}x + \frac{1}{3}y \\ \frac{1}{3}x + \frac{1}{3}y \end{bmatrix}$$

we have $\vec{x} = P\vec{u}$. Putting this back into the differential equation gives

$$\frac{d(P\vec{u})}{dt} = P\frac{d\vec{u}}{dt} = PD\vec{u}.$$

Multiplying on the left by P^{-1} gives

$$\frac{d\vec{u}}{dt} = D\vec{u}$$

or

$$\begin{bmatrix} \frac{du}{dt} \\ \frac{dw}{dt} \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} 8u \\ -w \end{bmatrix}$$

giving us our decoupled differential equations.

The take home message here is that the matrix P told us what substitution to make. This shouldn't be too surprising because P can always be thought of as a change of coordinate matrix, that changes coordinates from the ones we started with, to a more natural set of coordinates depending on the problem at hand.

In general, suppose you have a system of differential equations of the form

$$\begin{aligned} \frac{dx}{dt} &= ax + by \\ \frac{dy}{dt} &= cx + dy. \end{aligned}$$

Here the variables x and y depend on each other, but it would be great if they didn't, since then you could solve two separate differential equations. If the matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is diagonalisable, then by a simple change of coordinates, we can decouple the differential equation.

Let $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\frac{d\vec{x}}{dt} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}$. Then the set of equations above can be written as $\frac{d\vec{x}}{dt} = A\vec{x}$.

Suppose $P^{-1}AP = D$ for some diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. Let $\vec{x} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ and let $\vec{u} = \left[\begin{smallmatrix} u \\ w \end{smallmatrix}\right] = P^{-1}\vec{x}$. Then

$$\frac{d\vec{x}}{dt} = A\vec{x} \quad \Longleftrightarrow \quad P\frac{d\vec{u}}{dt} = PDP^{-1}\vec{x} \quad \Longleftrightarrow \quad \frac{d\vec{u}}{dt} = D\vec{u}.$$

Rewriting this last equation as a pair of differential equations we get

$$\begin{aligned} \frac{du}{dt} &= \lambda_1 u \\ \frac{dw}{dt} &= \lambda_2 w. \end{aligned}$$

No we have two decoupled equations, each which can be solved independently of the other. This process of decoupling is easily generalised to n variables and n equations.

8 Inner Product Spaces

We have seen that if you are a finite dimensional real vector space, once you pick a basis, you may as well think of the vector space as $\mathbb{R}^n$ (and indeed, any n-dimensional real vector space is isomorphic to $\mathbb{R}^n$). This somehow gives us geometric intuition to go off, since $\mathbb{R}^2$ and $\mathbb{R}^3$ at least come with a well defined notion of distance and angle. Let's go over what that is.

In $\mathbb{R}^2$, we know that the length of a vector $\vec{v} = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$ is given by

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}.$$

This formula is given to us by none other than Pythagoras himself! To compute the angle between two vectors $\vec{v} = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$ and $\vec{w} = \left(\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right)$, we invoke the cosine rule. Suppose θ is the angle between $\vec{v}$

and $\vec{w}$. Applying the cosine rule gives

$$\begin{aligned}
\cos\theta &= \frac{\|\vec{v}\|^2 + \|\vec{w}\|^2 - \|\vec{v}-\vec{w}\|^2}{2\,\|\vec{v}\|\,\|\vec{w}\|} \\
&= \frac{v_1^2 + v_2^2 + w_1^2 + w_2^2 - (v_1 - w_1)^2 - (v_2 - w_2)^2}{2\,\|\vec{v}\|\,\|\vec{w}\|} \\
&= \frac{2v_1w_1 + 2v_2w_2}{2\,\|\vec{v}\|\,\|\vec{w}\|} \\
&= \frac{v_1w_1 + v_2w_2}{\|\vec{v}\|\,\|\vec{w}\|}.
\end{aligned}$$

In $\mathbb{R}^3$, a similar thing occurs. Let $\vec{v} = \left(\begin{smallmatrix} v_1 \\ v_2 \\ v_3 \end{smallmatrix}\right)$ and $\vec{w} = \left(\begin{smallmatrix} w_1 \\ w_2 \\ w_3 \end{smallmatrix}\right)$ be two vectors in $\mathbb{R}^3$. Then the length of $\vec{v}$ is

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

and the angle θ between $\vec{v}$ and $\vec{w}$ is given by

$$\cos\theta = \frac{v_1w_1 + v_2w_2 + v_3w_3}{\|\vec{v}\|\,\|\vec{w}\|}.$$

Just like electricity and magentism are actually two sides of the same coin, angles and lengths in $\mathbb{R}^2$ and $\mathbb{R}^3$ are just two sides of the same coin, and that coin is the dot product.

Recall the dot product on $\mathbb{R}^n$ is defined as

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1w_1 + \cdots + v_nw_n.$$

So it is a function that eats two vectors and spits out a real number. In fact, we see that in $\mathbb{R}^2$ and $\mathbb{R}^3$ we have

$$\|\vec{v}\| = \sqrt{\vec{v}\cdot\vec{v}} \quad \text{and} \quad \cos\theta = \frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\,\|\vec{w}\|}$$

where θ is the angle between $\vec{v}$ and $\vec{w}$. Since the dot product is defined on $\mathbb{R}^n$ (and not just $\mathbb{R}^2$ and $\mathbb{R}^3$), we can use this to define length and angle in $\mathbb{R}^n$!

Lecture 22 - August 3

If we are to generalise the dot product to other vector spaces, we would want it to satisfy certain properties. Whatever our generalisation is, it should be a function

$$\langle\ ,\ \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{F}$$

and we would like to define the length of a vector $\vec{v}$ as $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v}\rangle}$ and the angle θ between two vectors $\vec{v}, \vec{w} \in \mathbb{V}$ as $\cos\theta = \frac{\langle \vec{v}, \vec{w}\rangle}{\|\vec{v}\|\|\vec{w}\|}$.

Furthermore, since we're dealing with length and angles, we would some geometric properties from $\mathbb{R}^2$ to hold in general. For example, we would like the length of a vector to be a positive real number, that is $\|\vec{v}\| \geq 0$. It would also be an added bonus if the length of a vector were zero if and only if the vector was the zero vector. Since we want to define angle in a similar way to the dot product, we would also like

$$\frac{|\langle \vec{v}, \vec{w}\rangle|}{\|\vec{v}\|\,\|\vec{w}\|} \leq 1.$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$ we know the shortest path between two points is a straight line, which is reflected in the triangle inequality. That is, it would be great if

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

Finally, we would like the length of a vector to behave nicely under scalar multiplication. That is $\|\alpha\vec{v}\| = |\alpha|\,\|\vec{v}\|$.

With these desires in mind, let's define an inner product on an arbitrary vector space, and we'll see that all of our wishes will come true!

Definition. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. An **inner product** on $\mathbb{V}$ is a function

$$\langle\ ,\ \rangle : \mathbb{V}\times\mathbb{V}\to\mathbb{F}$$

such that for all $\vec{u},\vec{v},\vec{w}\in\mathbb{V}$ and $\alpha\in\mathbb{F}$,

1. $\langle\vec{v},\vec{w}\rangle = \overline{\langle\vec{w},\vec{v}\rangle}$.
2. $\langle\alpha\vec{v},\vec{w}\rangle = \alpha\,\langle\vec{v},\vec{w}\rangle$.
3. $\langle\vec{u}+\vec{v},\vec{w}\rangle = \langle\vec{u},\vec{w}\rangle + \langle\vec{v},\vec{w}\rangle$.
4. (a) $\langle\vec{v},\vec{v}\rangle \geq 0$.
   (b) If $\langle\vec{v},\vec{v}\rangle = 0$ then $\vec{v}=\vec{0}$.

A vector space $\mathbb{V}$ equipped with an inner product $\langle\ ,\ \rangle$ is called an **inner product space**.

Note that if $\mathbb{F}=\mathbb{R}$, then $\overline{a}=a$ for all $a\in\mathbb{R}$ so the first property becomes $\langle\vec{v},\vec{w}\rangle = \langle\vec{w},\vec{v}\rangle$.
Let's see some examples.

Example. The standard Hermitian inner product on $\mathbb{C}^n$ was defined in assignment 3. Let $\vec{v},\vec{w}\in\mathbb{C}$ where

$$\vec{v} = \begin{pmatrix} v_1\\ \vdots\\ v_n\end{pmatrix} \quad\text{and}\quad \vec{w} = \begin{pmatrix} w_1\\ \vdots\\ w_n\end{pmatrix}.$$

Then define the standard Hermitian inner product as

$$\langle\vec{v},\vec{w}\rangle = v_1\overline{w}_1 + \cdots + v_n\overline{w}_n.$$

Let's prove that the standard Hermitian inner product is indeed an inner product.

Proof. Let $\vec{v} = \begin{pmatrix} v_1\\ \vdots\\ v_n\end{pmatrix}$, $\vec{w} = \begin{pmatrix} w_1\\ \vdots\\ w_n\end{pmatrix}$, $\vec{u} = \begin{pmatrix} u_1\\ \vdots\\ u_n\end{pmatrix} \in\mathbb{V}$ and $\alpha\in\mathbb{C}$. Then

$$\begin{aligned}\langle\vec{v},\vec{w}\rangle &= v_1\overline{w}_1 + \cdots + v_n\overline{w}_n\\ &= \overline{w}_1 v_1 + \cdots + \overline{w}_n v_n\\ &= \overline{w_1\overline{v}_1 + \cdots + w_n\overline{v}_n}\\ &= \overline{\langle\vec{w},\vec{v}\rangle}\end{aligned}$$

so 1 holds. We have

$$\begin{aligned}\langle \alpha\vec{v}, \vec{w}\rangle &= \alpha v_1\overline{w}_1 + \cdots + \alpha v_n\overline{w}_n\\ &= \alpha(v_1\overline{w}_1 + \cdots + v_n\overline{w}_n)\\ &= \alpha\,\langle \vec{v}, \vec{w}\rangle\end{aligned}$$

and

$$\begin{aligned}\langle \vec{u}+\vec{v}, \vec{w}\rangle &= (u_1+v_1)\overline{w}_1 + \cdots + (u_n+v_n)\overline{w}_n\\ &= u_1\overline{w}_1 + v_1\overline{w}_1 + \cdots + u_n\overline{w}_n + v_n\overline{w}_n\\ &= (u_1\overline{w}_1 + \cdots + u_n\overline{w}_n) + (v_1\overline{w}_1 + \cdots + v_n\overline{w}_n)\\ &= \langle \vec{u}, \vec{w}\rangle + \langle \vec{v}, \vec{w}\rangle\end{aligned}$$

so 2 and 3 hold. For 4a we have

$$\langle \vec{v}, \vec{v}\rangle = v_1\overline{v}_1 + \cdots + v_n\overline{v}_n = |v_1|^2 + \cdots + |v_n|^2 \geq 0.$$

Finally, suppose $\langle \vec{v}, \vec{v}\rangle = |v_1|^2 + \cdots + |v_n|^2 = 0$. Since each $|v_i|^2$ is a postive real number, the only way this can be true is if $|v_1| = \cdots = |v_n| = 0$. This in turn implies $v_1 = \cdots = v_n = 0$ so $\vec{v} = \vec{0}$.

Alas, the standard Hermitian inner product is indeed an inner product. ■

Example. Define the following function on $\mathcal{P}_3(\mathbb{R})$.

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx.$$

This is an inner product on $\mathcal{P}_3(\mathbb{R})$, and it is left as an exercise to verify this claim.

Since the inner product of a vector with itself is always a positive real number, it now makes sense to define the length of a vector.

Definition. For any vector $\vec{v}$ in an inner product space, define the **norm** as $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v}\rangle}$.

As a warm up exercise, prove that in any inner product space, $\left\|\vec{0}\right\| = 0$.

Example. On $\mathcal{P}_2(\mathbb{C})$ consider

$$\langle p, q\rangle = p(i)\overline{q(i)} + p(-i)\overline{q(-i)} + p(1)\overline{q(1)}.$$

This is an inner product. Indeed, it is left as an exercise to check properties 1, 2, and 3. For 4a,

$$\langle p, p\rangle = |p(i)|^2 + |p(-i)|^2 + |p(1)|^2 \geq 0.$$

For 4b, suppose $\langle p, p\rangle = 0$. Then

$$|p(i)|^2 + |p(-i)|^2 + |p(1)|^2 = 0$$

so $p(i) = p(-i) = p(1) = 0$. Since p is a polynomial of degree at most 2 that evaluates to 0 at three distinct points, we must have $p = 0$. Alternatively, suppose $p = ax^2 + bx + c$. Then $p(i) = p(-i) = p(1) = 0$ gives the system of equations

$$\begin{aligned}-a + bi + c &= 0\\ -a - bi + c &= 0\\ a + b + c &= 0.\end{aligned}$$

Plugging this into an augmented matrix and row-reducing gives

$$\left[\begin{array}{ccc|c} -1 & i & 1 & 0 \\ -1 & -i & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

so $a = b = c = 0$ and $p = 0$.

So this is an inner product. With respect to this inner product, let's compute the norm of some vectors. We have

$$\|1\| = \sqrt{\langle 1, 1\rangle} = \sqrt{3} \quad \text{and} \quad \|x\| = \sqrt{\langle x, x\rangle} = \sqrt{|i|^2 + |-i|^2 + |1|^2} = \sqrt{3}.$$

Example. Consider the vector space $\mathbb{C}^2$ and Let $\vec{v} = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$, $\vec{w} = \left(\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right)$. Define

$$\langle \vec{v}, \vec{w}\rangle = 2\vec{v}_1\overline{\vec{w}_1} - \vec{v}_2\overline{\vec{w}_2}.$$

As above, properties 1, 2, and 3 are left as an exercise to verify. However, for property 4 we have

$$\langle \vec{v}, \vec{v}\rangle = 2\left|v_1\right|^2 - \left|v_2\right|^2.$$

So in fact

$$\left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = -1$$

so this is not an inner product on $\mathbb{C}^2$.

Lecture 23 - August 7

Okay, let's take a break from all these examples. It may be natural to ask at this point whether or not every vector space can be turned into an inner product space. Let's answer that now.

Proposition 42. *Every finite-dimensional vector space admits an inner product.*

Proof. Let $\mathbb{V}$ be a vector space with basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$. Let $\vec{v} = t_1\vec{v}_1 + \cdots + t_n\vec{v}_n$ and $\vec{w} = s_1\vec{v}_1 + \cdots + s_n\vec{v}_n$. Then it is left as an exercise to check

$$\langle \vec{v}, \vec{w}\rangle = t_1\overline{s_1} + \cdots + t_n\overline{s_n}$$

is an inner product on $\mathbb{V}$. ■

It is true that every infinite-dimensional vector space also admits an inner product, but we won't be proving that here.

To finish off this introduction to inner products, we'll state a few important properties that are left to you to prove.

Proposition 43. *Let $\mathbb{V}$ be an inner product space. For all $\vec{v}, \vec{u}, \vec{w} \in \mathbb{V}$ and $\alpha \in \mathbb{F}$ the following is true.*

- $\langle \vec{0}, \vec{v}\rangle = 0$.
- $\langle \vec{v}, \alpha\vec{w}\rangle = \overline{\alpha}\,\langle \vec{v}, \vec{w}\rangle$.
- $\langle \vec{v}, \vec{u} + \vec{w}\rangle = \langle \vec{v}, \vec{u}\rangle + \langle \vec{v}, \vec{w}\rangle$.

Proof. Exercise. ■

8.1 Orthogonality

Recall in $\mathbb{R}^n$ that the angle θ between two vectors $\vec{v}$ and $\vec{w}$ is given by $\cos\theta = \frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\|\vec{w}\|}$. Therefore if the dot product of two vectors is 0, we know they are perpendicular. The notion of being perpendicular (or orthogonal as we will call it) turns out to be an extremely usful notion in inner product spaces.

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

Example. Consider $\mathcal{P}_2(\mathbb{R})$ with the inner product

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx.$$

Then

$$\langle 1, x \rangle = \int_{-1}^{1} xdx = 0$$

so 1 and x are orthogonal. However,

$$\left\langle 1, x^2 \right\rangle = \int_{-1}^{1} x^2 dx = \frac{2}{3}$$

so 1 and x^2 are not orthogonal.

One thing we know from geometry is that if you have a right-angled triangle, then Pythagoras' theorem holds. So, if we are to believe that being orthogonal really means that two vectors are at right angles to each other, we should expect the Pythagorean theorem to hold. Indeed it does!

Proposition 44. *Let $\mathbb{V}$ be an inner product space. Suppose $\vec{v} \perp \vec{w}$. Then $\|\vec{v}+\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$.*

Proof. We have

$$\begin{aligned}
\|\vec{v}+\vec{w}\|^2 &= \langle \vec{v}+\vec{w}, \vec{v}+\vec{w} \rangle \\
&= \langle \vec{v}, \vec{v}+\vec{w} \rangle + \langle \vec{w}, \vec{v}+\vec{w} \rangle \\
&= \langle \vec{v}, \vec{v} \rangle + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle + \langle \vec{w}, \vec{w} \rangle \\
&= \langle \vec{v}, \vec{v} \rangle + \langle \vec{w}, \vec{w} \rangle \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2
\end{aligned}$$

completing the proof. ∎

Our short term goal is to prove the Cauchy-Schwarz inequality (Theorem 46 below), which allows us to define the angle between two vectors. To do that we first need the following technical lemma

Lemma 45. *Let $\mathbb{V}$ be an inner product space and let $\vec{v}, \vec{w} \in \mathbb{V}$ such that $\vec{w} \neq \vec{0}$. Then $\vec{w}$ is perpendicular to $\vec{v} - \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle}\vec{w}$.*

Proof. We simply need to take the inner product between these two vectors and show they are zero. We have

$$\begin{aligned}\left\langle \vec{w}, \vec{v} - \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\vec{w}\right\rangle &= \langle \vec{w}, \vec{v}\rangle - \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\langle \vec{w}, \vec{w}\rangle \\ &= \langle \vec{w}, \vec{v}\rangle - \langle \vec{w}, \vec{v}\rangle \frac{1}{\|\vec{w}\|^2}\|\vec{w}\|^2 \\ &= 0\end{aligned}$$

completing the proof. ■

You may have come across the orthogonal projection of a vector $\vec{v}$ onto $\vec{w}$ in $\mathbb{R}^n$ before, and the vector is given by $\frac{\vec{v}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\vec{w}$. This term appeared in the previous proof, and we will revisit it later. Phrased in this language, Lemma 45 proves that the perpendicular component of $\vec{v}$ when projected onto $\vec{w}$ is indeed perpendicular to $\vec{w}$.

Theorem 46 (Cauchy-Schwarz Inequality)**.** *Let $\mathbb{V}$ be an inner product space. Then $|\langle \vec{v}, \vec{w}\rangle| \leq \|\vec{v}\|\,\|\vec{w}\|$.*

Proof. If $\vec{w} = \vec{0}$ we have $|\langle \vec{v}, \vec{w}\rangle| = \|\vec{v}\|\,\|\vec{w}\| = 0$ so the statement is true. Assume now that $\vec{w} \neq \vec{0}$. We have

$$\begin{aligned}\|\vec{v}\|^2 &= \left\|\vec{v} - \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\vec{w} + \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\vec{w}\right\|^2 \\ &= \left\|\vec{v} - \frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\vec{w}\right\|^2 + \left\|\frac{\langle \vec{v}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\vec{w}\right\|^2 \\ &\geq \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{\|\vec{w}\|^4}\|\vec{w}\|^2 \\ &= \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{\|\vec{w}\|^2}.\end{aligned}$$

Since $\|\vec{w}\|^2$ is positive, this implies

$$\|\vec{v}\|^2\,\|\vec{w}\|^2 \geq |\langle \vec{v}, \vec{w}\rangle|^2 .$$

Since norms are always positive real numbers we can conclude

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

completing the proof. ■

There is actually a geometric interpretation of the Cauchy-Schwarz inequality, at least in $\mathbb{R}^2$. It turns out that the inequality is a rephrasing of the following fact from geometry: If you have a parallelogram with side lengths x and y, then the area of that parallelogram is maximised exactly when the parallelogram is a rectangle. It's a fun exercise to try and see how this fact relates to the Cauchy-Schwarz inequality for the dot product on $\mathbb{R}^2$!

Because of the Cauchy-Schwarz inequality, we can now sensibly define the angle between two vectors, at least when our vector space is over the field $\mathbb{R}$.

Definition. Let $\mathbb{V}$ be a real inner product space. Define the angle θ between two vectors $\vec{v}$ and $\vec{w}$ by $\cos(\theta) = \frac{\langle \vec{v}, \vec{w}\rangle}{\|\vec{v}\|\|\vec{w}\|}$.

There are various ways to define the angle between vectors in a complex inner product space, each serving a different purpose. We won't be talking about the angle between vectors in a complex vector space in this course, except for the case when vectors are orthogonal.

Lecture 24 - August 8

We will finish this section by returning to our motivation for defining an inner product. The next proposition shows us that our notion of length is indeed a sensible one.

Proposition 47. *Let $\mathbb{V}$ be an inner product space. Then the norm satisfies the following conditions.*

1. $\|\alpha\vec{v}\| = |\alpha|\,\|\vec{v}\|$,
2. $\|\vec{v}+\vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$, *(this is called the* **triangle inequality***)*
3. *If* $\|\vec{v}\| = 0$*, then* $\vec{v} = 0$.

Proof. Properties 1 and 3 are left as an exercise. For the triangle inequality we will make use of the Cauchy-Schwarz inequality. We have

$$\begin{aligned}
\|\vec{v}+\vec{w}\|^2 &= \langle \vec{v}+\vec{w}, \vec{v}+\vec{w}\rangle \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2 + \langle \vec{v}, \vec{w}\rangle + \overline{\langle \vec{v}, \vec{w}\rangle} \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\,\mathrm{Re}(\langle \vec{v}, \vec{w}\rangle) \\
&\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\,|\mathrm{Re}(\langle \vec{v}, \vec{w}\rangle)| \\
&\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\,|\langle \vec{v}, \vec{w}\rangle| \\
&\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\,\|\vec{v}\|\,\|\vec{w}\| \\
&= (\|\vec{v}\| + \|\vec{w}\|)^2.
\end{aligned}$$

Since both sides are positive we have $\|\vec{v}+\vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ completing the proof. ■

There is a notion of a norm on a vector space that is not an inner product space, and it's a function that satisfies the 3 properties of Proposition 47. So we have actually proved that what we are calling the norm is in fact a norm on a vector space $\mathbb{V}$. It is true that there are norms that do not arise this way (being built from an inner product), but we will not be concered with such norms in this course.

8.2 Orthonormal Bases

Consider the standard basis in $\mathbb{R}^n$ equipped with the dot product. Then each vector in the basis has length 1, and even better, any two vectors in the basis are orthogonal. This will be our gold standard to head towards.

Definition. Let $\mathbb{V}$ be an inner product space. A set $\{\vec{v}_1, \ldots, \vec{v}_k\} \subset \mathbb{V}$ is called **orthogonal** if $\langle \vec{v}_i, \vec{v}_j\rangle = 0$ whenever $i \neq j$.

Definition. A vector $\vec{v}$ in an inner product space is a **unit vector** if $\|\vec{v}\| = 1$.

Definition. A set $\{\vec{v}_1, \ldots, \vec{v}_k\}$ in an inner product space is an **orthonormal set** if it is an orthogonal set and each vector is a unit vector.

Example. Consider $M_{2\times 2}(\mathbb{R})$ with inner product $\langle A, B\rangle = \operatorname{tr}(A^T B)$. It is left as an exercise to show it is in an inner product.

Define the matrices

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then

$$\begin{aligned} \langle A, B\rangle &= \operatorname{tr}(A^T B) = \operatorname{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = 0 \\ \langle A, A\rangle &= \operatorname{tr}(A^T A) = \operatorname{tr}\left(\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}\right) = 1 \\ \langle B, B\rangle &= 1. \end{aligned}$$

Since $\langle A, C\rangle = \langle B, C\rangle = 0$ we have $\{A, B, C\}$ is an orthogonal set, but it is not orthonormal (since $\|C\| = 0$). However, $\{A, B\}$ is an orthonormal set.

Example. The set $\left\{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\right\}$ is orthonormal in $\mathbb{R}^2$ with respect to the dot product.

Example (Legendre polynomials)**.** Recall from Question 1 on Assignment 2 the first 4 Legendre polynomials $\left\{1, x, \frac{3}{2}x^2 - \frac{1}{2}, \frac{5}{2}x^3 - \frac{3}{2}x\right\}$. You can check that this is an orthogonal set in $\mathcal{P}_3(\mathbb{R})$ with respect to the inner product

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx.$$

However, it is not orthonormal since

$$\begin{aligned} \|1\| &= \sqrt{\int_{-1}^{1} 1dx} = \sqrt{2} \\ \|x\| &= \sqrt{\int_{-1}^{1} x^2 dx} = \sqrt{\frac{2}{3}} \\ \left\|\frac{3}{2}x^2 - \frac{1}{2}\right\| &= \sqrt{\frac{2}{5}} \\ \left\|\frac{5}{2}x^3 - \frac{3}{2}x\right\| &= \sqrt{\frac{2}{7}}. \end{aligned}$$

However,

$$\left\{\frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}}x, \sqrt{\frac{5}{2}}\left(\frac{3}{2}x^2 - \frac{1}{2}\right), \sqrt{\frac{7}{2}}\left(\frac{5}{2}x^3 - \frac{3}{2}x\right)\right\}$$

is an orthonormal set. This follows from the fact that $\|\alpha\vec{v}\| = |\alpha|\,\|\vec{v}\|$ in any inner product space.

If our model for an orthogonal set is the standard basis in $\mathbb{R}^n$, we should expect an orthogonal set to be linearly independent. Indeed that is the case.

Proposition 48. *Suppose $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is orthogonal and $\vec{v}_i \neq \vec{0}$ for all i. Then $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent.*

Proof. Suppose $t_1\vec{v}_1 + \cdots + t_k\vec{v}_k = \vec{0}$. Fix $i \in \{1, \ldots, k\}$. Then

$$\begin{aligned} 0 &= \langle \vec{v}_i, t_1\vec{v}_1 + \cdots + t_k\vec{v}_k \rangle \\ &= t_1 \langle \vec{v}_i, \vec{v}_1 \rangle + \cdots + t_i \langle \vec{v}_i, \vec{v}_i \rangle + \cdots + t_k \langle \vec{v}_i, \vec{v}_k \rangle \\ &= t_i \|\vec{v}_i\|^2 . \end{aligned}$$

Since $\vec{v}_i \neq \vec{0}$, we must have $t_i = 0$. Since this is true for all i, we conclude $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent. ■

We now make a special definition for the case that we have an orthonormal set which forms a basis for an inner product space.

Definition. A set $\{\vec{v}_1, \ldots, \vec{v}_n\} \subset \mathbb{V}$ is an **orthonormal basis** if it is an orthonormal set and it is a basis.

Lecture 25 - August 9

Example.

- The standard basis of $\mathbb{C}^n$ with respect to the standard Hermitian inner product is an orthonormal basis.
- In $\mathbb{C}^2$, if
$$\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right\rangle = 2a\bar{c} + b\bar{d}$$
then $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ is an orthonormal basis for $\mathbb{C}^2$.
- The set
$$\left\{ \frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}}x, \sqrt{\frac{5}{2}}\left(\frac{3}{2}x^2 - \frac{1}{2}\right), \sqrt{\frac{7}{2}}\left(\frac{5}{2}x^3 - \frac{3}{2}x\right) \right\}$$
is an orthonormal basis for $\mathcal{P}_3(\mathbb{R})$ with respect to the inner product
$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx.$$

As a fun exercise, see if you can prove that given a basis of a finite dimensional vector space, there is an inner product such that the basis is an orthonormal basis!

8.3 Projections

You may have seen the projection of a vector onto another vector in a previous course, and we have seen hints of it in the proof of the Cauchy-Schwarz inequality. We now shift our attention to fleshing out the details in full.

Let's think about what projection looks like in $\mathbb{R}^2$. Suppose $\vec{v}$ and $\vec{w}$ are two non-zero vectors in $\mathbb{R}^2$, and we wish to project $\vec{v}$ onto $\vec{w}$. We can think of this as shining a light perpendicular to $\vec{w}$, and drawing a vector representing the shadow of $\vec{v}$.

Suppose θ is the angle between $\vec{v}$ and $\vec{w}$. Then by drawing out a right triangle, we see the length of the projection must be $\|\vec{v}\| \cos\theta$. The direction we wish the vector to go in is the direction $\vec{w}$ is pointing, so we can obtain the projection by scalar multiplying the unit vector in the direction of $\vec{w}$ by $\|\vec{v}\| \cos\theta$. This gives the projection as

$$\|\vec{v}\| \cos\theta \frac{1}{\|\vec{w}\|}\vec{w} = \frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}.$$

We will use this as motivation for defining projections in inner product spaces in general.

Definition. Let $\mathbb{V}$ be an inner product space, let $\vec{w}, \vec{v} \in \mathbb{V}$ with $\vec{w} \neq \vec{0}$. Define the **projection** of $\vec{v}$ onto $\vec{w}$ as

$$\operatorname{proj}_{\vec{w}}(\vec{v}) := \frac{\langle \vec{v}, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w}.$$

Define the **perpendicular** vector of $\vec{v}$ with respect to $\vec{w}$ by

$$\operatorname{perp}_{\vec{w}}(\vec{v}) := \vec{v} - \frac{\langle \vec{v}, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w}.$$

We already saw in Lemma 45 that $\operatorname{perp}_{\vec{w}}(\vec{v})$ is orthogonal to $\vec{w}$, which is what we'd expect to be true if these definitions imitate the situation in $\mathbb{R}^2$.

Example. Let $\vec{v} = \begin{pmatrix} 4 \\ 1+i \\ 2 \end{pmatrix} \in \mathbb{C}^3$. Let $\vec{w}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ and $\vec{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. With respect to the standard Hermitian inner product we have

$$\operatorname{proj}_{\vec{w}_1}(\vec{v}) = \frac{2}{1}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

$$\operatorname{proj}_{\vec{w}_2}(\vec{v}) = \frac{i+1}{1}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ i+1 \\ 0 \end{pmatrix}$$

These computations are hopefully what we'd expect if we are to project onto the z and y axes of $\mathbb{C}^3$.

Another way to think about a projection is as finding the closest vector on a subspace to a given vector. More specifically, the projection of $\vec{v}$ onto $\vec{w}$ is the closest vector in the subspace spanned by $\{\vec{w}\}$ to $\vec{v}$. Remember, since we're in an inner product space, we have a notion of length so asking for the closest vector makes sense.

With this in mind, what we're really doing when we're projecting onto a vector is we're projecting onto the one-dimensional subspace spanned by that vector. It's natural to now ask how we can project onto a general subspace. Let's look at an example.

Example. Consider $\mathbb{R}^3$ with the dot product. Let $\vec{v} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$, $\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, and $\vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Consider the subspace $\mathbb{W} = \operatorname{Span}(\{\vec{e}_1, \vec{e}_2\})$, so $\mathbb{W}$ is the x-y plane in $\mathbb{R}^3$. Then the projection of $\vec{v}$ onto $\mathbb{W}$ should be the vector $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$. This vector can be obtained by summing the projections onto $\vec{e}_1$ and $\vec{e}_2$. That is,

$$\operatorname{proj}_{\mathbb{W}}(\vec{v}) = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \operatorname{proj}_{\vec{e}_1}(\vec{v}) + \operatorname{proj}_{\vec{e}_2}(\vec{v}).$$

From this example, it may be tempting to guess the following: That the projection of a vector $\vec{v}$ onto a subspace $\mathbb{W}$ is simply obtained by choosing a basis, projecting $\vec{v}$ onto each basis vector, and summing up the resulting vectors. Unfortunately this does not work all the time, but it does work when we have an orthogonal basis (and only when the basis is orthogonal, see the exercise following the next definition). Because of this, we make the following definition.

Definition. Let $\mathbb{V}$ be an inner product space and let $\mathbb{W} \subset \mathbb{V}$ be a subspace. Let $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be an orthogonal basis for $\mathbb{W}$. Let $\vec{v} \in \mathbb{V}$. Define the **projection** of $\vec{v}$ onto $\mathbb{W}$, and the **perpendicular** vector of $\vec{v}$ with respect to $\mathbb{W}$ by

$$\mathrm{proj}_{\mathbb{W}}(\vec{v}) := \mathrm{proj}_{\vec{w}_1}(\vec{v}) + \cdots + \mathrm{proj}_{\vec{w}_k}(\vec{v}) \quad \text{and} \quad \mathrm{perp}_{\mathbb{W}}(\vec{v}) := \vec{v} - \mathrm{proj}_{\mathbb{W}}(\vec{v}).$$

respectively.

The next exercise shows that the projection of $\vec{v}$ onto a subspace $\mathbb{W}$ gives the unique vector in $\mathbb{W}$ closest to $\vec{v}$. Furthermore, it shows that we really do need an orthogonal basis for projection to give us the closest vector.

Exercise. Let $\mathbb{V}$ be an inner product space, $\mathbb{W}$ a subspace of $\mathbb{V}$ and $\vec{v} \in \mathbb{V}$.

- Prove that for all $\vec{w} \in \mathbb{W}$, $\|\vec{v} - \mathrm{proj}_{\mathbb{W}}(\vec{v})\| \leq \|\vec{v} - \vec{w}\|$.
- Prove that if $\|\vec{v} - \mathrm{proj}_{\mathbb{W}}(\vec{v})\| = \|\vec{v} - \vec{w}\|$, then $\vec{w} = \mathrm{proj}_{\mathbb{W}}(\vec{v})$.
- Let $\mathcal{C} = \{\vec{c}_1, \ldots, \vec{c}_k\}$ be a basis for $\mathbb{W}$ that is *not* orthogonal. Show that there exists a vector $\vec{v} \in \mathbb{V}$ such that $\mathrm{proj}_{\vec{c}_1}(\vec{v}) + \cdots + \mathrm{proj}_{\vec{c}_k}(\vec{v}) \neq \mathrm{proj}_{\mathbb{W}}(\vec{v})$.

Example. Consider the vector space $\mathcal{P}_3(\mathbb{R})$ with the inner product $\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx$. We know from an earlier example that $1 \perp x$, so $\{1, x\}$ is an orthogonal basis for $\mathbb{W} = \mathrm{Span}(\{1, x\})$. Let's find the projection of x^2 onto $\mathbb{W}$.

We have

$$\mathrm{proj}_{\mathbb{W}}(x^2) = \frac{\left\langle x^2, 1\right\rangle}{\|1\|^2}1 + \frac{\left\langle x^2, x\right\rangle}{\|x\|^2}x.$$

It can be checked that $\left\langle x^2, 1\right\rangle = \frac{2}{3}$, $\left\langle x^2, x\right\rangle = 0$, and $\|1\|^2 = 2$. Therefore

$$\mathrm{proj}_{\mathbb{W}}(x^2) = \frac{1}{2}\frac{2}{3}1 = \frac{1}{3}$$

and

$$\mathrm{perp}_{\mathbb{W}}(x^2) = x^2 - \frac{1}{3}.$$

This tells us that the closest vector in $\mathbb{W}$ to x^2 is the vector $\frac{1}{3}$. Go figure!

Lecture 26 - August 10

8.4 Orthogonal Complements

Suppose $\mathbb{V}$ is an inner product space. With every subspace $\mathbb{W}$ comes another subspace called the orthogonal complement for free! This subspace consists of all the vectors in $\mathbb{V}$ that are orthogonal to every vector in $\mathbb{W}$.

Definition. Let $\mathbb{V}$ be an inner product space and let $\mathbb{W} \subset \mathbb{V}$ be a subspace. Define the **orthogonal complement** of $\mathbb{W}$ by

$$\mathbb{W}^\perp := \{\vec{v} \in \mathbb{V} : \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in \mathbb{W}\}.$$

Example. Consider $\mathbb{R}^3$ with the dot product and let $\mathbb{W} = \text{Span}\left(\left\{\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right)\right\}\right)$. We would expect $\mathbb{U} = \text{Span}\left(\left\{\left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right)\right\}\right)$ to be equal to $\mathbb{W}^\perp$.

Proof. We want to show $\mathbb{U} = \mathbb{W}^\perp$. Let $\vec{u} \in \mathbb{U}$. Then $\vec{u} = \left(\begin{smallmatrix}0\\0\\t\end{smallmatrix}\right)$ for some $t \in \mathbb{R}$. Let $\vec{w}$ be an arbitrary vector in $\mathbb{W}$, so $\vec{w} = \left(\begin{smallmatrix}a\\b\\0\end{smallmatrix}\right)$ for some $a, b \in \mathbb{R}$. Then $\vec{u} \cdot \vec{w} = 0$ so $\vec{u} \in \mathbb{W}^\perp$ and $\mathbb{U} \subset \mathbb{W}^\perp$. Conversely, suppose $\vec{u} = \left(\begin{smallmatrix}u_1\\u_2\\u_3\end{smallmatrix}\right) \in \mathbb{W}^\perp$. Then since $\vec{u} \cdot \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right) = 0$, we must have $u_1 = 0$. Similarly we must have $u_2 = 0$. Therefore $\vec{u} \in \mathbb{U}$ and $\mathbb{W}^\perp \subset \mathbb{U}$. We can now conclude $\mathbb{W}^\perp = \mathbb{U}$, completing the proof. ∎

Example. We saw earlier that

$$\left\{1, x, \frac{3}{2}x^2 - \frac{1}{2}, \frac{5}{2}x^3 - \frac{3}{2}x\right\}$$

is an orthogonal basis for $\mathcal{P}_3(\mathbb{R})$ with respect to the inner product $\langle p, q \rangle = \int_{-1}^1 p(x)q(x)dx$. Therefore if $\mathbb{W} = \text{Span}\left(\left\{1, \frac{3}{2}x^2 - \frac{1}{2}\right\}\right)$ then $\mathbb{W}^\perp = \text{Span}\left(\left\{x, \frac{5}{2}x^3 - \frac{3}{2}x\right\}\right)$.

Exercise. Let $\mathbb{V}$ be an inner product space and $\mathbb{W} \subset \mathbb{V}$ a subspace. Prove the following statements.

- $\mathbb{W}^\perp$ is a subspace of $\mathbb{V}$.
- $\mathbb{W} \cap \mathbb{W}^\perp = \{\vec{0}\}$.
- If $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a basis for $\mathbb{W}$ and $\{\vec{w}_1, \ldots, \vec{w}_m\}$ is a basis for $\mathbb{W}^\perp$, then $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_1, \ldots, \vec{w}_m\}$ is a basis for $\mathbb{V}$.
- $\dim(\mathbb{W}) + \dim(\mathbb{W}^\perp) = \dim(\mathbb{V})$.

8.5 The Gram-Schmidt Orthogonalisation Procedure

In the previous section, we defined projection onto a subspace by starting with an orthogonal basis. This raises the question: does an orthogonal basis always exist?

Let's look at an example where we start with a basis and create an orthonormal basis from it.

Example. In $\mathbb{R}^3$ equipped with the dot product, suppose we have

$$\vec{v}_1 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \text{and} \quad \begin{pmatrix}0\\1\\1\end{pmatrix}.$$

It turns out that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis for $\mathbb{R}^3$. However, $\vec{v}_1 \cdot \vec{v}_2 = 1$ so it is not an orthogonal set (or an orthonormal one for that matter). We will now create an orthogonal basis starting form this one, and then scale the vectors to obtain an orthonormal set.

Let's first create an orthogonal set $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. We may as well start with

$$\vec{w}_1 = \vec{v}_1.$$

Now, whatever we make $\vec{w}_2$, it better be orthogonal to $\vec{w}_1$. We have proved earlier that $\text{perp}_{\vec{w}_1}(\vec{v}_2)$ is orthogonal to $\vec{w}_1$, so let's use that. We have

$$\begin{aligned}
\text{perp}_{\vec{w}_1}(\vec{v}_2) &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 \\
&= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}.
\end{aligned}$$

So that we don't have to deal with fractions, set

$$\vec{w}_2 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

No we have two perpendicular vectors $\vec{w}_1$ and $\vec{w}_2$. To create $\vec{w}_3$, we do the same thing except we will want to take $\text{perp}_{\mathbb{W}}(\vec{v}_3)$ where $\mathbb{W} = \text{Span}(\{\vec{w}_1, \vec{w}_2\})$, which will give us a vector orthogonal to both $\vec{w}_1$ and $\vec{w}_2$. Let's do it!

$$\begin{aligned}
\vec{w}_3 &= \text{perp}_{\text{Span}(\{\vec{w}_1, \vec{w}_2\})}(\vec{v}_3) \\
&= \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2 \\
&= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{6}\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}.
\end{aligned}$$

Again, to make our lives easier, set

$$\vec{w}_3 = \begin{pmatrix} -2 \\ 2 \\ 2 \end{pmatrix}.$$

Now we have three vectors that are all orthogonal, so $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is an orthogonal set. Since $\vec{0}$ is not part of this set, it is linearly independent and thus a basis for $\mathbb{R}^3$.

To create an orthonormal basis for $\mathbb{R}^3$ we simply scale the vectors so they have norm 1. Therefore

$$\left\{ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

is an orthonormal basis of $\mathbb{R}^3$ with respect to the dot product.

The method we illustrated in this example works in general, and it is called the **Gram-Schmidt orthogonalisation procedure**. In general, here's how it goes.

Let $\mathbb{V}$ be an inner product space with basis $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$. To obtain an orthogonal basis $\mathcal{C} = \{\vec{w}_1, \dots, \vec{w}_n\}$ proceed as follows.

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 \\
\vec{w}_2 &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 \\
\vec{w}_3 &= \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2 \\
&\vdots \\
\vec{w}_n &= \vec{v}_n - \frac{\langle \vec{v}_n, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 - \cdots - \frac{\langle \vec{v}_n, \vec{w}_{n-1} \rangle}{\|\vec{w}_{n-1}\|^2}\vec{w}_{n-1}.
\end{aligned}$$

Then $\mathcal{C} = \{\vec{w}_1, \dots, \vec{w}_n\}$ is an orthogonal basis for $\mathbb{V}$. To obtain an orthonormal basis $\mathcal{D} = \{\vec{u}_1, \dots, \vec{u}_n\}$ let $\vec{u}_i = \frac{1}{\|\vec{w}_i\|}\vec{w}_i$.

To prove that this procedure actually works, we need to ensure that at each step we actually get a non-zero vector. This amounts to proving the following statement.

Exercise. Let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for an inner product space $\mathbb{V}$. Let $\{\vec{w}_1, \dots, \vec{w}_i\}$ be the first i vectors obtained from the Gram-Schmidt orthogonalisation procedure. Prove $\operatorname{Span}(\{\vec{v}_1, \dots, \vec{v}_i\}) = \operatorname{Span}(\{\vec{w}_1, \dots, \vec{w}_i\})$. Use this to prove the resulting basis $\{\vec{w}_1, \dots, \vec{w}_n\}$ is an orthogonal basis for $\mathbb{V}$.

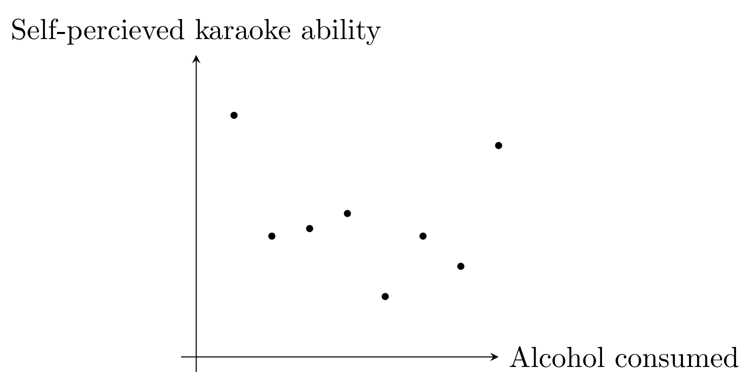
As a consequence of the Gram-Schmidt process, we get the following corollary, which ensures that we can always project onto a subspace (among other things).

Corollary 49. *Every inner product space admits an orthonormal basis.*

Lecture 27 - August 13

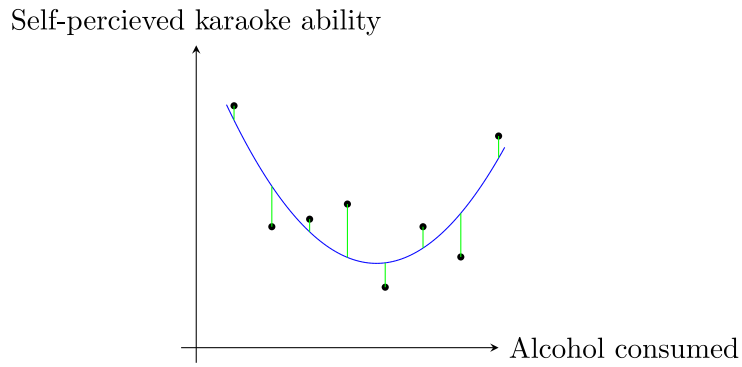
8.6 Application of projections: Method of Least Squares

Suppose we're doing a super-serious study, and we've collected data which is looking for some kind of relationship between "self-percieved karaoke ability" and "alcohol consumed." The data we've collected looks like this when plotted:



Our goal is to model this data by some quadratic equation $y = a + bx + cx^2$ where y is the perceived karaoke ability and x is the alcohol consumed. After all, we would expect this to occur in reality: a person while sober thinks they're quite good, after a couple of drinks is aware they will be slurring a little, but after drinking more will begin to think they are god's gift to vocal performance!

So, we would like to find a quadratic that looks something like the blue curve:



Furthermore, we would like such a quadratic to make the lengths of the vertical green lines as small as possible, since the vertical green lines represent the error between our model and the experimental data.

So let's say we had the data points $(x_1, y_1), \ldots, (x_n, y_n)$ which we want to approximate by $y = a + bx + cx^2$. Then we want to minimise the vertical green bars, or equivalently

$$(y_1 - (a + bx_1 + cx_1^2))^2 + \cdots + (y_n - (a + bx_1 + cx_1^2))^2.$$

This looks an awful lot like a norm in $\mathbb{R}^n$ with respect to the dot product. In fact, if we let

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \vec{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{x^2} = \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix}$$

be vectors in $\mathbb{R}^n$. Then minimizing the length of the errors (the vertical green bars) is the same as minimising

$$\left\| \vec{y} - (a\vec{1} + b\vec{x} + c\vec{x^2}) \right\|^2$$

with respect to the dot product. In other words, to find a, b, and c, we need to find the vector on the subspace $\text{Span}(\{\vec{1}, \vec{x}, \vec{x^2}\})$ closest to the vector $\vec{y}$. We know how to do this! Putting all of these observations together, we find a, b, and c by setting

$$a\vec{1} + b\vec{x} + c\vec{x^2} = \text{proj}_{\mathbb{W}}(\vec{y}).$$

Now, it is tempting to think we can just project using the basis $\{\vec{1}, \vec{x}, \vec{x^2}\}$, but this basis is not guaranteed to be orthogonal. So, in order to project onto the desired subspace, we will use the following two properties, which are left as exercises for you to prove.

Proposition 50. *Let $\mathbb{V}$ be an inner product space and $\mathbb{W}$ a subspace. Let $\vec{v} \in \mathbb{V}$.*

- *$\vec{w} = proj_{\mathbb{W}}(\vec{v})$ if and only if $\vec{v} - \vec{w} \in \mathbb{W}^{\perp}$.*
- *Let $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be a basis (not necessarily orthogonal) for $\mathbb{W}$. Then $\vec{v} \in \mathbb{W}^{\perp}$ if and only if $\langle \vec{v}, \vec{w}_i \rangle = 0$ for all i.*

Proof. Exercise. ∎

So putting this together, we need to find a, b, c such that

$$(\vec{y} - (a\vec{1} + b\vec{x} + c\vec{x^2})) \cdot \vec{1} = 0$$
$$(\vec{y} - (a\vec{1} + b\vec{x} + c\vec{x^2})) \cdot \vec{x} = 0$$
$$(\vec{y} - (a\vec{1} + b\vec{x} + c\vec{x^2})) \cdot \vec{x^2} = 0.$$

We can organise this information as follows. First note that if we view vectors in $\mathbb{R}^n$ as column matrices, then $\vec{v} \cdot \vec{w}$ is given by the entry in the 1×1 matrix $\vec{w}^T \vec{v}$. With this in mind, let

$$\vec{a} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} \vec{1} & \vec{x} & \vec{x^2} \end{bmatrix}.$$

Then the three equations above can be rephrased by the matrix equation

$$X^T(\vec{y} - X\vec{a}) = \vec{0}.$$

Rearranging this gives

$$\vec{a} = (X^T X)^{-1} X^T \vec{y}.$$

Let's see this in action.

Example. Suppose we have the following data:

x	-1	0	1	2
y	4	1	1	-1

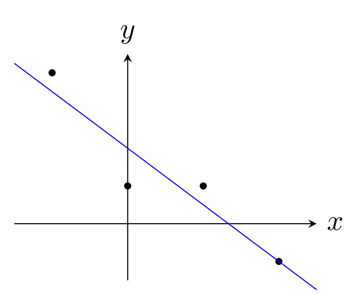
We wish to approiximate this data set by a linear equation $y = a + bx$. So we let

$$\vec{a} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \text{and} \quad \vec{y} = \begin{bmatrix} 4 \\ 1 \\ 1 \\ -1 \end{bmatrix}.$$

Then

$$\vec{a} = (X^T X)^{-1} X^T \vec{y} = \left(\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \right)^{-1} \begin{bmatrix} 5 \\ -5 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ -5 \end{bmatrix} = \begin{bmatrix} 2 \\ -\frac{3}{2} \end{bmatrix}.$$

Therefore $y = 2 - \frac{3}{2}x$ is the line of best fit to the given data. Let's see what this line looks like.



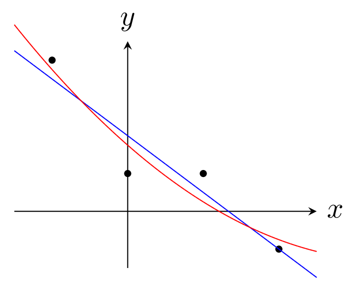
While this is good, maybe it's not as good as we'd like! Let's see if we can do better approximating the data by the equation $y = a + bx + cx^2$. This time we have

$$\vec{a} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad X = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \quad \text{and} \quad \vec{y} = \begin{bmatrix} 4 \\ 1 \\ 1 \\ -1 \end{bmatrix}.$$

Then computing $\vec{a}$ in the same way as above gives

$$\vec{a} = (X^T X)^{-1} X^T \vec{y} = \begin{bmatrix} \frac{7}{4} \\ -\frac{7}{4} \\ \frac{1}{4} \end{bmatrix}.$$

Therefore the quadratic of best fit is $y = \frac{7}{4} - \frac{7}{4}x + \frac{1}{4}x^2$. Plotting this (in red) looks like this:



That's a little better!

In general, suppose we have some data points

x	x_1	$\cdots$	x_n
y	y_1	$\cdots$	y_n

and we want to find the equation $y = a_0 + a_1 x + \cdots + a_k x^k$ of best fit to this data. Let

$$\vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \vec{x^2} = \begin{bmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{bmatrix}, \ldots, \vec{x^k} = \begin{bmatrix} x_1^k \\ \vdots \\ x_n^k \end{bmatrix}, \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{and} \quad \vec{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_k \end{bmatrix}.$$

Let $X = \begin{bmatrix} \vec{1} & \vec{x} & \cdots & \vec{x^k} \end{bmatrix}$, then $\vec{a} = (X^T X)^{-1} X^T \vec{y}$ gives the equation of best fit.

It's an interesting exercise to think about what happens if $X^T X$ is not invertible and we can't do this!

Lecture 28 - August 14

9 Hermitian, Unitary and Normal Matrices

Suppose we have an inner product $\langle\ ,\ \rangle$ on $\mathcal{P}_1(\mathbb{C})$ such that

$$\langle 1, 1 \rangle = 2, \quad \langle x, x \rangle = 3, \quad \text{and} \quad \langle 1, x \rangle = 1 + i.$$

Is it possible to recover the entire inner product just from this information? After all, we now know how the basis vectors play with the inner product. Let's see if we can recover the inner product. We have

$$\begin{aligned}\langle a+bx, c+dx\rangle &= \langle a, c\rangle + \langle bx, c\rangle + \langle a, dx\rangle + \langle bx, dx\rangle \\ &= a\overline{c}\,\langle 1,1\rangle + b\overline{c}\,\langle x, 1, +\rangle\, a\overline{d}\,\langle 1, x\rangle + b\overline{d}\,\langle x, x\rangle \\ &= 2a\overline{c} + (1-i)(b\overline{c}) + (1+i)(a\overline{d}) + 3b\overline{d}.\end{aligned}$$

This defines our inner product! So, just like linear maps are determined by what they do to basis vectors, it appears, at least in this example, that inner products may be determined by what they do to basis vectors.

Recall that the fact that linear maps are determined by the images of the basis vectors is intimately tied to the fact that every linear map can be represented by a matrix (once we've chosen bases for our vector spaces of course). So a natural question arises: can we represent an inner product by a matrix, and does every matrix represent an inner product?

In the example we've just seen, let $\mathcal{B}$ be the standard basis for $\mathcal{P}_1(\mathbb{C})$, and let $p = a + bx$, $q = c + dx$. Then $[p]_{\mathcal{B}} = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $[q]_{\mathcal{B}} = \left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$. Then

$$\begin{bmatrix} \overline{c} & \overline{d} \end{bmatrix} \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a\overline{c} + (1-i)b\overline{c} + (1+i)a\overline{d} + 3b\overline{d} \end{bmatrix}.$$

What do you know?! If we identity a 1×1 matrix with an element of $\mathbb{C}$ we have $\langle p, q\rangle = \overline{[q]_{\mathcal{B}}}^T A[p]_{\mathcal{B}}$ for some 2×2 matrix A.

Let's phrase our motivating question a little better.

Question. Let $\mathbb{V}$ be an n-dimensional vector space with basis $\mathcal{B}$.

1. Let $\langle\ ,\ \rangle$ be an inner product on $\mathbb{V}$. Is there an $n\times n$ matrix A such that $\langle \vec{v}, \vec{w}\rangle = \overline{[\vec{w}]_{\mathcal{B}}}^T A[\vec{v}]_{\mathcal{B}}$?
2. Let A be an $n\times n$ matrix A. Does

$$\langle \vec{v}, \vec{w}\rangle = \overline{[\vec{w}]_{\mathcal{B}}}^T A[\vec{v}]_{\mathcal{B}}$$

   define an inner product on $\mathbb{V}$?

In order to answer this question, we'll need a little terminology.

Definition. Let A be a matrix. Define the **adjoint** of A, denoted A^* by the conjugate transpose. That is, $A^* = \overline{A}^T$.

So for example

$$\begin{bmatrix} 2 & 4 \\ -i & 2+i \end{bmatrix}^* = \begin{bmatrix} 2 & i \\ 4 & 2-i \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}^* = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}.$$

Here are some basic properties about adjoints, which are left as an exercise to verify.

Proposition 51. *Let A and B be matrices. Then*

- $(AB)^* = B^*A^*$,
- $(A^*)^* = A$,

- *If A is a real matrix, then $A^* = A^T$,*
- $(\alpha A)^* = \overline{\alpha} A^*$.

Proof. Exercise. ■

Now, let's return to the motivating question. To try and get some headway into the problem, let's assume $\mathbb{V}$ is an inner product space with $\dim(\mathbb{V}) = 2$ and $\mathcal{B} = \{\vec{\beta}_1, \vec{\beta}_2\}$. Let $\vec{v} = a\vec{\beta}_1 + b\vec{\beta}_2$ and $\vec{w} = c\vec{\beta}_1 + d\vec{\beta}_2$, so $[\vec{v}]_{\mathcal{B}} = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $[\vec{w}]_{\mathcal{B}} = \left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$.

Suppose there is a 2×2 matrix $A = \left[\begin{smallmatrix} w & x \\ y & z \end{smallmatrix}\right]$ such that $\langle \vec{v}, \vec{w} \rangle = [\vec{w}]^*_{\mathcal{B}} A [\vec{v}]_{\mathcal{B}}$. Then

$$\langle \vec{v}, \vec{w} \rangle = \begin{bmatrix} \overline{c} & \overline{d} \end{bmatrix} \begin{bmatrix} w & x \\ y & z \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\overline{c}w + a\overline{d}y + b\overline{c}x + b\overline{d}z \end{bmatrix}.$$

Since

$$\langle \vec{v}, \vec{w} \rangle = a\overline{c} \left\langle \vec{\beta}_1, \vec{\beta}_1 \right\rangle + a\overline{d} \left\langle \vec{\beta}_1, \vec{\beta}_2 \right\rangle + b\overline{c} \left\langle \vec{\beta}_2, \vec{\beta}_1 \right\rangle + b\overline{d} \left\langle \vec{\beta}_2, \vec{\beta}_2 \right\rangle$$

we must have

$$\begin{bmatrix} w & x \\ y & z \end{bmatrix} = \begin{bmatrix} \left\langle \vec{\beta}_1, \vec{\beta}_1 \right\rangle & \left\langle \vec{\beta}_2, \vec{\beta}_1 \right\rangle \\ \left\langle \vec{\beta}_1, \vec{\beta}_2 \right\rangle & \left\langle \vec{\beta}_2, \vec{\beta}_2 \right\rangle \end{bmatrix}.$$

For the general case of an n-dimensional inner product space with basis $\mathcal{B} = \{\vec{\beta}_1, \ldots, \vec{\beta}_n\}$, if there is to be an $n \times n$ matrix $A = [A_{ij}]$ such that $\langle \vec{v}, \vec{w} \rangle = [\vec{w}]^*_{\mathcal{B}} A [\vec{v}]_{\mathcal{B}}$ then we must have $A_{ij} = \left\langle \vec{\beta}_j, \vec{\beta}_i \right\rangle$. Furthermore, since $\langle \vec{v}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{v} \rangle}$ we must have $\overline{A_{ij}} = A_{ji}$, or equivalently, $A^* = A$.

Definition. A matrix A is **Hermitian** if $A = A^*$. If A is real, this is the same as $A = A^T$ and we say A is **symmetric**.

Example. The matrices

$$\begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 4+i \\ 2 & 0 & 2-i \\ 4-i & 2+i & 0 \end{bmatrix}$$

are all Hermitian. The second one is also symmetric (since it is real).

So, if $\langle \vec{v}, \vec{w} \rangle = [\vec{w}]^*_{\mathcal{B}} A [\vec{v}]_{\mathcal{B}}$ has any hope of being an inner product, we must have $A = A^*$. Let's see what else we need.

For the first axiom of inner products we have

$$\begin{aligned} [\langle \vec{v}, \vec{w} \rangle]^* &= ([\vec{w}]^*_{\mathcal{B}} A [\vec{v}]_{\mathcal{B}})^* \\ &= [\vec{v}]^*_{\mathcal{B}} A^* [\vec{w}]_{\mathcal{B}} \\ &= [\vec{v}]^*_{\mathcal{B}} A [\vec{w}]_{\mathcal{B}} \\ &= [\langle \vec{w}, \vec{v} \rangle] \end{aligned}$$

which, since the transpose of a 1×1 matrix is just the matrix itself, implies $\overline{\langle \vec{v}, \vec{w} \rangle} = \langle \vec{w}, \vec{v} \rangle$.

For the other axioms we will abuse notation and identify the 1×1 matrix $[\langle \vec{v}, \vec{w} \rangle]$ with the number $\langle \vec{v}, \vec{w} \rangle$. We have

$$\begin{aligned} \langle \vec{v} + \vec{u}, \vec{w} \rangle &= [\vec{w}]^*_{\mathcal{B}} A [\vec{v} + \vec{u}]_{\mathcal{B}} \\ &= [\vec{w}]^*_{\mathcal{B}} A ([\vec{v}]_{\mathcal{B}} + [\vec{u}]_{\mathcal{B}}) \\ &= [\vec{w}]^*_{\mathcal{B}} A [\vec{v}]_{\mathcal{B}} + [\vec{w}]^*_{\mathcal{B}} A [\vec{u}]_{\mathcal{B}} \\ &= \langle \vec{v}, \vec{w} \rangle + \langle \vec{u}, \vec{w} \rangle. \end{aligned}$$

Similarly,

$$\begin{aligned}\langle t\vec{v}, \vec{w}\rangle &= [\vec{w}]_{\mathcal{B}}^* A[t\vec{v}]_{\mathcal{B}} \\ &= \alpha[\vec{w}]_{\mathcal{B}}^* A[\vec{v}]_{\mathcal{B}} \\ &= \alpha \langle \vec{v}, \vec{w}\rangle .\end{aligned}$$

As usual, the interesting situation is axiom 4. It's not clear what conditions we need on A so that $[\vec{v}]_{\mathcal{B}}^* A[\vec{v}]_{\mathcal{B}} \geq 0$. Let's keep this in the back of our minds. We will return to it later, but in order to do so, we need to take a foray into the world of unitary diagonalisation.

Lecture 29 - August 15

Roughly, a unitary matrix is a change of basis matrix, but not just any ordinary one. It's a change of basis matrix that keeps length and angle fixed.

Definition. A square matrix U is called **unitary** if $U^* = U^{-1}$. If a matrix O is real and $O^T = O^{-1}$, then we say O is **orthogonal**.

Example.

- The $n \times n$ identity matrix is unitary.
- Let $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}}i \end{bmatrix}$. Then U is unitary since

$$UU^* = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

  and

$$U^*U = \frac{1}{2}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- Let $\mathcal{B}$ be the standard basis in $\mathbb{R}^2$ and $R : \mathbb{R}^2 \to \mathbb{R}^2$ the linear map given by rotation about the origin by an angle of θ. Then $[R]_{\mathcal{B}}$ is a unitary matrix (or an orthogonal matrix since it will be real). It's an exercise for you to justify this fact.

Proposition 52. *Let U be an $n \times n$ matrix. Equip $\mathbb{C}^n$ with the standard Hermitian inner product. The following are equivalent.*

- *U is unitary.*
- *The rows form an orthonormal basis of $\mathbb{C}^n$.*
- *The columns form an orthonormal basis of $\mathbb{C}^n$.*

Proof sketch. The main observation is the following. Let $U = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$. Then

$$U^*U = \begin{bmatrix} \overline{\vec{v}_1} \\ \vdots \\ \overline{\vec{v}_n} \end{bmatrix}\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \langle \vec{v}_1, \vec{v}_1\rangle & \cdots & \langle \vec{v}_n, \vec{v}_1\rangle \\ \vdots & \ddots & \vdots \\ \langle \vec{v}_1, \vec{v}_n\rangle & \cdots & \langle \vec{v}_n, \vec{v}_n\rangle \end{bmatrix}.$$

So if $U^*U = I$ we must have

$$\langle \vec{v}_i, \vec{v}_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

implying $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal set. The rest of the details are left as an exercise. ■

We'll finish off this introductory section by showing that unitary matrices, when veiwed as linear maps with respect to the standard basis in $\mathbb{C}^n$, do not affect the standard Hermitian inner product. That is, unitary matrices preserve length and angle!

Proposition 53. *Let U be a unitary matrix and consider $\mathbb{C}^n$ with the standard Hermitian inner product. Then*

$$\langle U\vec{v}, U\vec{w}\rangle = \langle \vec{v}, \vec{w}\rangle$$

for all $\vec{v}, \vec{w} \in \mathbb{C}^n$.

Proof. Note that if we view vectors in $\mathbb{C}^n$ as column matrices, we have $\langle \vec{v}, \vec{w}\rangle = \vec{w}^*\vec{v}$. Since U is unitary we have

$$\langle U\vec{v}, U\vec{w}\rangle = (U\vec{w})^*U\vec{v} = \vec{w}^*U^*U\vec{v} = \vec{w}^*\vec{v} = \langle \vec{v}, \vec{w}\rangle$$

completing the proof. ■

9.1 Spectral Theorem for Hermitian Matrices

We have just seen that a unitary matrix can be viewed as a special kind of change of basis matrix, one that perserves length and angles. While it is not true that ever matrix is diagonalisable, we will now see that every matrix is upper-triangularisable, that is, for every matrix we can find a basis with respect to which the matrix is upper-triangular. Even better, we can choose this basis to be an orthonormal basis (with respect to the standard Hermitian inner product of course).

Theorem 54 (Schur's Triangularisation Theorem)**.** *Let A be an $n \times n$ matrix. There is a unitary matrix U and an upper triangular matrix T such that $U^*AU = T$.*

Proof. We will proceed by induction on n. For 1×1 matrices, this is clearly true since every matrix is upper triangular.

Now suppose A is an $n \times n$ matrix, and assume that the theorem is true for all $(n-1) \times (n-1)$ matrices. Let $\vec{v}_1$ be a unit eigenvector of A with eigenvalues λ. Extend $\vec{v}_1$ to a basis for $\mathbb{C}^n$ and perform the Gram-Schmidt procedure to obtain an orthonormal basis $\{\vec{v}_1, \vec{w}_2, \ldots, \vec{w}_n\}$.

Let $V_1 = \begin{bmatrix}\vec{v}_1 & \vec{w}_2 & \cdots & \vec{w}_n\end{bmatrix}$. Since the columns are orthonormal, V_1 is a unitary matrix. We now have

$$V_1^* A V_1 = \left[\begin{array}{c} \overline{\vec{v}_1} \\ \overline{\vec{w}_2} \\ \ddots \\ \overline{\vec{w}_n} \end{array}\right] A \begin{bmatrix} \vec{v}_1 & \vec{w}_2 & \cdots & \vec{w}_n \end{bmatrix}$$

$$= \left[\begin{array}{c} \overline{\vec{v}_1} \\ \overline{\vec{w}_2} \\ \ddots \\ \overline{\vec{w}_n} \end{array}\right] \begin{bmatrix} A\vec{v}_1 & A\vec{w}_2 & \cdots & A\vec{w}_n \end{bmatrix}$$

$$= \left[\begin{array}{c} \overline{\vec{v}_1} \\ \overline{\vec{w}_2} \\ \ddots \\ \overline{\vec{w}_n} \end{array}\right] \begin{bmatrix} \lambda\vec{v}_1 & A\vec{w}_2 & \cdots & A\vec{w}_n \end{bmatrix}$$

$$= \begin{bmatrix} \langle \lambda\vec{v}_1, \vec{v}_1 \rangle & \langle A\vec{w}_2, \vec{v}_2 \rangle & \cdots & \langle A\vec{w}_n, \vec{v}_1 \rangle \\ \langle \lambda\vec{v}_1, \vec{w}_2 \rangle & \langle A\vec{w}_2, \vec{w}_2 \rangle & \cdots & \langle A\vec{w}_n, \vec{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \lambda\vec{v}_1, \vec{w}_n \rangle & \langle A\vec{w}_2, \vec{w}_n \rangle & \cdots & \langle A\vec{w}_n, \vec{w}_n \rangle \end{bmatrix}$$

$$= \left[\begin{array}{c|ccc} \lambda & * & * & * \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right].$$

B is an $(n-1) \times (n-1)$ matrix, so by the inductive hypothesis there is a unitary matrix V_2 such that $V_2^* B V_2 = T$ where T is an upper-triangular matrix. Let

$$U = V_1 \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right].$$

Then

$$U^* A U = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right]^* V_1^* A V_1 \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right]$$

$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2^* \end{array}\right] \left[\begin{array}{c|c} \lambda & * \\ \hline 0 & B \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V_2 \end{array}\right]$$

$$= \left[\begin{array}{c|c} \lambda & * \\ \hline 0 & V_2^* B V_2 \end{array}\right]$$

$$= \left[\begin{array}{c|c} \lambda & * \\ \hline 0 & T \end{array}\right].$$

Since this last matrix is upper triangular, the result has been proven by the principle of mathematical induction. ■

Lecture 30 - August 16

We now have the missing piece to prove in general what you proved on an assignment.

Corollary 55. *For any matrix A, the determinant is the product of its eigenvalues and the trace is the sum of its eigenvalues.*

Proof. By Schur's triangularisation theorem we know there is a unitary matrix U and an upper triangular matrix

$$T = \begin{bmatrix} t_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & t_n \end{bmatrix}$$

such that $U^*AU = T$. The characteristic polynomial for T is

$$(t_1 - \lambda) \cdots (t_n - \lambda)$$

so the t_i are the eigenvalues of T. Furthermore we have

$$\operatorname{tr}(T) = t_1 + \cdots + t_n \quad \text{and} \quad |T| = t_1 \cdots t_n.$$

Since A and T are similar they have the same eigenvalues, determinant and trace, completing the proof. ■

Now that we have this, we can make a very interesting observation. It turns out that the characteristic polynomial of a matrix has the trace and determinant hidden in the coefficients!

Corollary 56. *Let A be an $n \times n$ matrix and let the characteristic polynomial be*

$$\det(A - \lambda I) = a_n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

Then $a_n = (-1)^n$, $a_{n-1} = (-1)^{n-1}\operatorname{tr}(A)$, and $a_0 = |A|$.

Proof. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Then the characteristic polynomial is given by

$$\det(A - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda) = a_n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

expanding and equating coefficients gives us

$$a_n = (-1)^n, \quad a_{n-1} = (-1)^{n-1}(\lambda_1 + \cdots + \lambda_n), \quad \text{and} \quad a_0 = \lambda_1 \cdots \lambda_n.$$

By the previous corollary, the trace is the sum of the eigenvalues and the determinant is the product, so $a_{n-1} = (-1)^{n-1}\operatorname{tr}(A)$ and $a_0 = |A|$, completing the proof. ■

The take home message here is that every matrix is similar to an upper-triangular one! That's pretty great.

Let's now return back to Hermitian matrices. A curious thing happens when we upper-triangularise a Hermitian matrix.

Lemma 57. *Let A be a Hermitian matrix. Then A is unitarily diagonalisable (that is there is a unitary U and a diagonal D such that $U^*AU = D$) and all its eigenvalues are real.*

Proof. By Schur's triangularisation theorem, we know $U^*AU = T$ for some upper triangular matrix T. However,

$$T^* = (U^*AU)^* = U^*A^*U = U^*AU = T$$

so T is Hermitian. Since T is upper triangular and Hermitian, it must be diagonal so A is unitarily diagonalisable. Furthermore, the entries on the diagonal of any Hermitian matrix are real, so all the eigenvalues of A are real. ■

Lemma 58. *Let A be a matrix such that there exists a unitary U and a diagonal matrix D with real entries such that $U^*AU = D$. Then A is Hermitian.*

Proof. Exercise. ■

So putting these two lemmas together we get the spectral theorem for Hermitian matrices, and in the real case, the spectral theorem for orthogonal matrices.

Theorem 59 (Spectral Theorem for Hermitian Matrices)**.** *A square matrix is Hermitian if and only if it is unitarily diagonalisable with real eigenvalues.*

Theorem 60 (Spectral Theorem for Symmetric Matrices)**.** *A real square matrix is symmetric if and only if it is orthogonally diagonalisable over $\mathbb{R}$.*

It is important to note that in the spectral theorem for Hermitian matrices the requirement that the eigenvalues are all real is definitely required, as shown in the next example.

Example. Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Diagonalising this matrix, we find the eigenvalues are i and $-i$ with bases for the eigenspaces given by

$$\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\}$$

respectively. Notice that with the Hermitian inner product we have

$$\left\langle \begin{bmatrix} i \\ 1 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\rangle = i^2 + 1 = 0.$$

Therefore we can make both these vectors unit vectors and

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}}i & -\frac{1}{\sqrt{2}}i \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

is a unitary matrix. Furthermore, since both the columns are eigenvectors we have $U^*AU = \left[\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right]$.

So we see here that A is unitarily diagonalisable, but the eigenvalues are not real, and A is not Hermitian.

Great, so now we know that if we have a Hermitian matrix, then it is diagonalisable. Unitarily. With real eigenvalues. Unfortunately, the results so far don't really give us any indication as to how we find U. It is tempting to simply find a basis of eigenvectors as usual (that is, diagonalise as usual) and perform Gram-Schmidt on that basis to obtain an orthonormal basis. There's only one problem with this idea: there's a chance that while performing Gram-Schmidt, you are no longer left with eigenvectors!

Lecture 31 - August 17

The previous example gives us a hint as to how to get around this. Notice that the two eigenvectors were already orthogonal, by magic! Even though the matrix in the example was not Hermitian, it turns out that for Hermitian matrices, the magic always happens!

Proposition 61. *Let A be an $n \times n$ Hermitian matrix, and suppose $A\vec{v} = \lambda\vec{v}$ and $A\vec{w} = \mu\vec{w}$ with $\lambda \neq \mu$. Then $\langle \vec{v}, \vec{w} \rangle = 0$ with respect to the standard Hermitian inner product on $\mathbb{C}^n$.*

Proof. Let $A\vec{v} = \lambda\vec{v}$ and $A\vec{w} = \mu\vec{w}$ with $\lambda \neq \mu$. Then

$$\begin{aligned}\lambda \langle \vec{v}, \vec{w} \rangle &= \langle \lambda\vec{v}, \vec{w} \rangle \\ &= \langle A\vec{v}, \vec{w} \rangle \\ &= \vec{w}^* A\vec{v} \\ &= (A\vec{w})^* \vec{v} \\ &= \langle \vec{v}, A\vec{w} \rangle \\ &= \langle \vec{v}, \mu\vec{w} \rangle \\ &= \overline{\mu} \langle \vec{v}, \vec{w} \rangle .\end{aligned}$$

Since A is Hermitian we know all the eigenvalues are real so $\overline{\mu} = \mu$. Rearranging this we get

$$(\lambda - \mu) \langle \vec{v}, \vec{w} \rangle = 0.$$

Since $\lambda \neq \mu$ we must have $\langle \vec{v}, \vec{w} \rangle = 0$, completing the proof. ■

Proposition 61 should ring a bell to Lemma 37, which says that eigenvectors corresponding to distinct eigenvalues are linearly independent. Here we have that when A is Hermitian, eigenvectors corresponding to distinct eigenvalues are orthogonal.

So, with this in mind, we now know that we don't have to perform Gram-Schmidt on the entire basis of eigenvectors, just on the basis for each eigenspace! This leads to the following algorithm to unitarily diagonalise a Hermitian matrix. Suppose we want to find a unitary U and a diagonal D such that $U^*AU = D$ for a unitarily diagonalisable matrix A.

1. Diagonalise as usual, obtaining $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ and a basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ of eigenvectors for $\mathbb{C}^n$.

2. Perform Gram-Schmidt on each eigenspace to obtain $\{\vec{w}_1, \dots, \vec{w}_n\}$, which is an orthonormal basis of eigenvectors for $\mathbb{C}^n$.

3. Let $U = \begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_n \end{bmatrix}$.

In particular, suppose an $n \times n$ Hermitian matrix A has n distinct eigenvalues. Since it has n distinct eigenvalues, we know it's diagonalisable. Since it's hermitian, we know the eigenvalues are all real. Furthermore, since it's Hermitian, any basis of eigenvectors for $\mathbb{C}^n$ will consist of one eigenvector for each eigenvalue, so the basis will be orthogonal. In this case, to obtain an orthonormal basis (and therefore a unitary U such that $U^*AU = D$ for some diagonal D) we simply need to choose each eigenvector to be a unit vector.

Let's see this algorithm in all its glory.

Example. Let's unitarily diagonalise the matrix

$$A = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}.$$

First note that A is Hermitian. Computing $|A - \lambda I|$ using your favourite method gives the characteristic polynomial as $\lambda(9 - \lambda)^2$. We'll now find bases for the eigenspaces corresponding to $\lambda = 0$

and $\lambda = 9$. Note that since A is Hermitian, we know it's diagonalisable so we should have the geometric multiplicity of 0 be 1, and the geometric multiplicity of 9 be 2.

We have

$$A - 0I = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore a basis for the eigenspace corresponding to 0 is $\left\{ \left[\begin{smallmatrix} 2 \\ 2 \\ 1 \end{smallmatrix} \right] \right\}$. We wish to have an orthonormal basis for this eigenspace, so instead we choose the basis

$$\left\{ \frac{1}{3} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

To find the eigenspace corresponding to $\lambda = 9$ we have

$$A - 9I = \begin{bmatrix} -4 & -4 & -2 \\ -4 & -4 & -2 \\ -2 & -2 & -1 \end{bmatrix} \sim \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

giving a basis $\{\vec{v}_1, \vec{v}_2\} = \left\{ \left[\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix} \right], \left[\begin{smallmatrix} -1 \\ 0 \\ 2 \end{smallmatrix} \right] \right\}$. Note that $\vec{v}_1$ and $\vec{v}_2$ are not orthogonal, so we must perform Gram-Schmidt to obtain an orthogonal basis for this eigenspace.

Performing Gram-Schmidt to obtain an orthogonal basis $\{\vec{w}_1, \vec{w}_2\}$ gives $\vec{w}_1 = \vec{v}_1$ and

$$\vec{w}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2} \vec{w}_1 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}.$$

Now $\{\vec{w}_1, \vec{w}_2\}$ is an orthogonal basis for the eigenspace corresponding to $\lambda = 9$. To obtain an orthonormal basis $\{\vec{u}_1, \vec{u}_2\}$ we set

$$\vec{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{u}_2 = \frac{1}{\sqrt{18}} \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}.$$

Alas, $U^* A U = D$ where

$$U = \begin{bmatrix} \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ \frac{1}{3} & 0 & \frac{4}{\sqrt{18}} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

9.2 Inner Products and Hermitian Matrices

Our diversion into unitary diagonalisation was motivated by finding matrix representatives for inner products. While unitary diagonalisation is important without this application, we now are in a position to answer our motivating questions about which matrices give rise to inner products.

Recall that if we have a vector space $\mathbb{V}$ with a basis $\mathcal{B}$, a Hermitian matrix A, and define

$$\langle \vec{v}, \vec{w} \rangle = [\vec{w}]_{\mathcal{B}}^* A [\vec{v}]_{\mathcal{B}}$$

then this function satisfies axioms 1, 2, and 3 of being an inner product. In order to work out what conditions we need to satisfy axiom 4, let's utilise what we now know about Hermitian matrices: they are unitarily diagonalisable!

Suppose $U^*AU = D$ where U is unitary and D is diagonal. Then

$$\langle \vec{v}, \vec{w} \rangle = [\vec{w}]_{\mathcal{B}}^* U D U^* [\vec{v}]_{\mathcal{B}} = (U^*[\vec{w}]_{\mathcal{B}})^* D (U^*[\vec{v}]_{\mathcal{B}}).$$

Since U^* is invertible, we can think of it as a change of basis matrix. In fact, we have $U^*[\vec{v}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{C}}$ for all $\vec{v} \in \mathbb{V}$, where $\mathcal{C}$ is some other basis (it's not important what it is). So we now have

$$\langle \vec{v}, \vec{w} \rangle = [\vec{w}]_{\mathcal{C}}^* D [\vec{v}]_{\mathcal{C}}.$$

Remember that D is a diagonal matrix with entries along the diagonal equal to the eigenvalues of A, and these eigenvalues are real. There is hope now to work out what restrictions, if any, we need to impose on D (and thus the eigenvalues of A) to ensure we obtain an inner product.

Let

$$[\vec{w}]_{\mathcal{C}} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, \quad [\vec{v}]_{\mathcal{C}} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Then

$$\langle \vec{v}, \vec{w} \rangle = [\vec{w}]_{\mathcal{C}}^* D [\vec{v}]_{\mathcal{C}} = \lambda_1 v_1 \overline{w}_1 + \cdots + \lambda_n v_n \overline{w}_n.$$

Therefore

$$\langle \vec{v}, \vec{v} \rangle = \lambda_1 |v_1|^2 + \cdots + \lambda_n |v_n|^2 .$$

Lecture 32 - August 20

Therefore $\langle \vec{v}, \vec{v} \rangle \geq 0$ if and only if $\lambda_1, \ldots, \lambda_n \geq 0$. Furthermore, if we want $\langle \vec{v}, \vec{v} \rangle = 0$ to imply $\vec{v} = \vec{0}$ we must have that none of the λ_i are equal to 0.

This discussion leads to the following proposition. The proof is left as an exercise, but you can use the discussion we just had to formalise a proof.

Proposition 62. *Let $\mathbb{V}$ be a vector space with basis $\mathcal{B} = \{\vec{\beta}_1, \ldots, \vec{\beta}_n\}$. Then $\langle\ ,\ \rangle$ is an inner product if and only if there is an $n \times n$ Hermitian matrix A with positive eigenvalues such that*

$$\langle \vec{v}, \vec{w} \rangle = [\vec{w}]_{\mathcal{B}}^* A [\vec{v}]_{\mathcal{B}}.$$

If there is such a matrix, it must be given by

$$A = \begin{bmatrix} \left\langle \vec{\beta}_1, \vec{\beta}_1 \right\rangle & \cdots & \left\langle \vec{\beta}_n, \vec{\beta}_1 \right\rangle \\ \vdots & \ddots & \vdots \\ \left\langle \vec{\beta}_1, \vec{\beta}_n \right\rangle & \cdots & \left\langle \vec{\beta}_n, \vec{\beta}_n \right\rangle \end{bmatrix}.$$

Proof. Exercise. ■

As you may have experienced, sometimes it is difficult to prove or disprove that a potential inner product statisfies axiom 4. Proposition 62 gives us an algorithmic way to check. We simply choose a basis, create the matrix in the proposition, check that the matrix actually performs the inner product for us, and compute its eigenvalues!

Example. Is
$$\langle a+bx, c+dx\rangle = 2a\overline{c} + (1+i)b\overline{c} + (1-i)a\overline{d} + 3b\overline{d}$$
an inner product on $\mathcal{P}_1(\mathbb{C})$? Let's find out.

Let $\mathcal{B}$ be the standard basis for $\mathcal{P}_1(\mathbb{C})$. Our candidate matrix is given by
$$A = \begin{bmatrix} \langle 1,1\rangle & \langle x,1\rangle \\ \langle 1,x\rangle & \langle x,x\rangle \end{bmatrix} = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}.$$
First thing we need to check is that this matrix is Hermitian, which it is. Now we need to check that it actually performs the inner product for us. That is, we need to check $\langle p,q\rangle = [q]_{\mathcal{B}}^* A[p]_{\mathcal{B}}$. We have
$$[c+dx]_{\mathcal{B}}^* A[a+bx]_{\mathcal{B}} = \begin{bmatrix} \overline{c} & \overline{d} \end{bmatrix} \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a\overline{c} + (1+i)b\overline{c} + (1-i)a\overline{d} + 3b\overline{d} \end{bmatrix}.$$
Therefore this matrix does the trick! So, to check whether or not it's an inner product, we need to compute the eigenvalues and make sure they're all positive. We have
$$\begin{aligned} |A-\lambda I| &= \begin{vmatrix} 2-\lambda & 1+i \\ 1-i & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)(3-\lambda) - (1+i)(1-i) \\ &= \lambda^2 - 5\lambda + 4 \\ &= (\lambda-4)(\lambda-1). \end{aligned}$$
Since the eigenvalues are 1 and 4, which are both positive, Proposition 62 allows us to conclude that this is indeed an inner product.

Example. On assignment 5 you were asked to decide whether or not
$$\left\langle \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \right\rangle = 2v_1w_1 + 2v_2w_2 + 2v_3w_3 - v_1w_2 - v_2w_1 - v_2w_3 - v_3w_2$$
is an inner product on $\mathbb{R}^3$. Let's see if we can use all this Hermitian matrix nonsense to answer the question.

Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the standard basis for $\mathbb{R}^3$. Our candidate matrix for this potential inner product is
$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$
This matrix is certainly Hermitian, which is a good start. Furthermore we have
$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 2v_1w_1 + 2v_2w_2 + 2v_3w_3 - v_1w_2 - v_2w_1 - v_2w_3 - v_3w_2 \end{bmatrix}$$
so the matrix does actually perform the inner product for us. Finally, you can compute its eigenvalues and get them to be $\{2, 2+\sqrt{2}, 2-\sqrt{2}\}$, all of which are positive! Therefore this is an inner product.

9.3 Normal Matrices

We will finish the course by attempting to remove the condition "with real eigenvalues" from the spectral theorem for Hermitian matrices. We say earlier that even though $\left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ is not Hermitian, it is still unitarily diagonalisable. However, we know not all matrices are unitarily diagonalisable (or even diagonalisable for that matter). So we need a class of matrices that contains all Hermitian matrices, but also all matrices which are unitarily diagonalisable but not all of their eigenvalues are real. Normal matrices are where the money is at!

Definition. A square matrix A is **normal** if $AA^* = A^*A$.

For example, the following matrices are normal:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & i & 0 \\ i & 0 & 0 \\ 0 & 0 & i \end{bmatrix} \quad \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4+i & 0 \\ 0 & 0 & \pi - \sqrt{2}i \end{bmatrix} \quad \begin{bmatrix} 2 & 3+i \\ 3-i & 4 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

In fact, we've already seen a whole bunch of normal matrices!

Proposition 63. *Hermitian, unitary, and diagonal matrices are all normal.*

Proof. Exercise. ■

As we can see from the examples above, there are normal matrices that are not unitary, diagonal, or Hermitian.

Now, suppose we have a matrix A which is unitarily diagonalisable, that is $U^*AU = D$ for some unitary U and diagonal D. Since diagonal matrices are normal we have

$$A^*A = (UDU^*)^*(UDU^*) = UD^*DU^* = UDD^*U^* = (UDU^*)(UDU^*)^* = AA^*.$$

So we have just proved that if a matrix is unitarily diagonalisable, it is normal. One of the more wonderful results in linear algebra is that the converse also holds! You are walked through a proof of the next theorem in assignment 6, and we will sketch out a different proof here.

Theorem 64 (Spectral Theorem for Normal Matrices)**.** *A square matrix is normal if and only if it is unitarily diagonalisable.*

Proof sketch. Here is an outline of the steps you would need to take to prove the forward direction. It is a good exercise to fill in the details and turn this into a robust proof

1. Show every square matrix A can be written as $A = B + iC$ where $B = \frac{1}{2}(A + A^*)$ and $C = -\frac{1}{2}(A - A^*)$.

2. Prove that such a B and C are Hermitian.

3. Prove that $BC = CB$ if and only if A is normal.

4. Let X and Y be diagonalisable $n \times n$ matrices. Prove that if $XY = YX$ then there exists an invertible P such that $P^{-1}XP$ and $P^{-1}YP$ are both diagonal matrices (although not necessarily the same diagonal matrix).

5. Returning to our matrices B and C, prove that there is a unitary U such that U^*BU and U^*CU are diagonal.

6. Prove that A is unitarily diagonalisable. ■

There are a few very important gems hidden in this proof. The way we decompose a matrix A into the two matrices B and C in step one should be compared to the real and imaginary part of a complex number. Recall that for any $z \in \mathbb{C}$, $\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z})$ and $\mathrm{Im}(z) = -\frac{1}{2}(z - \overline{z})$. Step 4 is an important lemma in its own right. Two matrices X and Y are said to **commute** if $XY = YX$. Step 4 is more commonly rephrased as "commuting diagonalisable matrices are simultaneously diagonalisable." In fact, the converse is true!

Exercise. Let X and Y be diagonalisable matrices. Prove that $XY = YX$ if and only if there is an invertible P such that $P^{-1}XP$ and $P^{-1}YP$ are both diagonal matrices (not necessarily the same diagonal matrix).

With the spectral theorem for normal matrices at our disposal, we may ask how we actually diagonalise a given normal matrix. The answer is the same as for Hermitian matrices. Diagonalise as usual and perform Gram-Schmidt on each eigenspace to obtain an orthonormal basis of eigenvectors!

10 To infinity and beyond!

We are now at the end of this course, but we've barely downed a couple of drops of the vast ocean of linear algebra.

In most of this course we've focused on finite-dimensional vector spaces. However, if you don't insist on finite dimensions, you get into the wonderful world of topological vector spaces, Banach spaces, Hilbert spaces, and functional analysis to name a few topics.

Although permitting infinite dimensional vector spaces yeilds a wild and wonderful world, there is a comparable world of matrix analysis laying in the finite-dimensional setting. Beautiful and surprising results can be found (look up the Gershgorin circle theorem and the eigenvalue interlacing theorem for examples of such theorems) if you are to delve into this world.

There is more to learn than you ever could imagine, good luck!