Lecture 1 (8/31)

Our main object of interest this semester will be *vector spaces*, so we dive straight in with a definition. This notion does include vectors as you may have learned about in high school or calculus classes, but it describes a more general set of objects.

The definitions here (and in the book) always assume that we are taking a vector space "over $\mathbb{R}$", meaning that $\mathbb{R}$ forms the set of so-called *scalars*. However, everything we do will hold for the scalars being any *field*. You do not need to know the definition of a field, but another field that you may be familiar with are the complex numbers $\mathbb{C}$, for example.

Two.I.1. (p.84)

Definition. A *vector space* (over $\mathbb{R}$) is a set of *vectors* V equipped with two operations $+$ and $\cdot$, such that the following holds: p.84

(1) the set V is closed under vector addition: $\vec{v}+\vec{w}\in V$ for all vectors $\vec{v},\vec{w}\in V$,

(2) vector addition is commutative: $\vec{v}+\vec{w}=\vec{w}\vec{v}$ for all $\vec{v},\vec{w}\in V$,

(3) vector addition is assiative: $(\vec{v}+\vec{w})+\vec{u}=\vec{v}+(\vec{w}+\vec{u})$ for all $\vec{v},\vec{w},\vec{u}\in V$,

(4) there is a zero vector $\vec{0}\in V$ such that $\vec{v}+\vec{0}=\vec{v}$ for all $\vec{v}\in V$,

(5) each $\vec{v}\in V$ has an additive inverse $\vec{w}\in V$ such that $\vec{v}+\vec{w}=\vec{0}$,

(6) the set V is closed under scalar multiplication, that is $r\cdot\vec{v}\in V$, for all $\vec{v}\in V$ and $r\in\mathbb{R}$,

(7) scalar multiplication distributes over scalar addition: $(r+s)\cdot\vec{v}=r\cdot\vec{v}+s\cdot\vec{v}$, for all $\vec{v}\in V$ and $r,s\in\mathbb{R}$,

(8) scalar multiplication distributes over vector addition: $r\cdot(\vec{v}+\vec{w})=r\cdot\vec{v}+r\cdot\vec{w}$, for all $\vec{v},\vec{w}\in V$ and $r\in\mathbb{R}$,

(9) ordinary multiplication of scalars associates with scalar multiplication: $(rs)\cdot\vec{v}=r\cdot(s\cdot\vec{v})$, for all $\vec{v}\in V$ and $r,s\in\mathbb{R}$,

(10) multiplication by the scalar 1 is the identity operation: $1\cdot\vec{v}=\vec{v}$, for all $\vec{v}\in V$.

Example. The vector space we will be working with the most is the one you are likely used to: $\mathbb{R}^n$. For example,

$$\mathbb{R}^3=\left\{\begin{pmatrix}x\\y\\z\end{pmatrix}\mid x,y,z\in\mathbb{R}\right\}.$$

Vector addition is defined as

$$\begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix}+\begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix}=\begin{pmatrix}x_1+x_2\\y_1+y_2\\z_1+z_2\end{pmatrix},\ \text{for}\ \begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix},\begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix},\in\mathbb{R}^3$$

and scalar multiplication is defined as

$$r\cdot\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}rx\\ry\\rz\end{pmatrix},\ \text{for}\ \begin{pmatrix}x\\y\\z\end{pmatrix},\in\mathbb{R}^3,\ r\in\mathbb{R}.$$

Example. In $\mathbb{R}^2$, lines through the origin are vector spaces. The example in the book shows how we can verify this using the step by step definition. For example, the line Ex. 1.3 p.85

$$L = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid y = 3x \right\} = \left\{ \begin{pmatrix} x \\ 3x \end{pmatrix} \mid x \in \mathbb{R} \right\} = \left\{ x \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} \mid x \in \mathbb{R} \right\}$$

is a vector space. However, lines in $\mathbb{R}^2$ that do not pass through the origin are not vector spaces. Can you show why?

Example. Polynomials of bounded degree with real coefficients can be treated as a vector space over $\mathbb{R}$. We define Ex. 1.11 p.90

$$\mathcal{P}_n = \{a_0 + a_1 x + a_2 x^2 \cdots + a_n x^n \mid a_0, \ldots, a_n \in \mathbb{R}\}.$$

Then, vector addition is given by

$$\begin{aligned} a(x) + b(x) &= (a_0 + a_1 x + a_2 x^2 \cdots + a_n x^n) + (b_0 + b_1 x + b_2 x^2 \cdots + b_n x^n) \\ &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 \cdots + (a_n + b_n)x^n, \text{ for all } a(x), b(x) \in \mathcal{P}_n, \end{aligned}$$

and scalar multiplication by

$$r \cdot a(x) = r \cdot (a_0 + a_1 x + a_2 x^2 \cdots + a_n x^n) = ra_0 + ra_1 x + ra_2 x^2 \cdots + ra_n x^n, \text{ for all } a(x) \in \mathcal{P}_n,\ r \in \mathbb{R}.$$

Example. Sets of $n \times m$ matrices with real number entries form a vector space, with entry-wise addition and scalar multiplication (just as with vectors). For example: Ex. 1.11 p.90

$$\mathcal{M}_{2\times 2} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R} \right\},$$

with vector addition given by

$$M_1 + M_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} a+w & b+x \\ c+y & d+z \end{pmatrix}, \text{ for } M_1, M_2 \in \mathcal{M}_{2\times 2},$$

and scalar multiplication given by

$$r \cdot M = r \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ra & rb \\ rc & rd \end{pmatrix}, \text{ for } M \in \mathcal{M}_{2\times 2}, r \in \mathbb{R}.$$

Example. The set $\{f \mid f : \mathbb{R} \to \mathbb{R}\}$ of real functions in one variable can be treated as a vector space, with addition given by Ex. 1.12 p.90

$$(f_1 + f_2)(x) = f_1(x) + f_2(x),$$

and scalar multiplication by

$$(r \cdot f)(x) = rf(x).$$

For the following lemma, try to write out proofs yourself, based on the definition of a vector space. The proofs are also in the book.

Lemma. *In any vector space V, for any $\vec{v} \in V$ and $r \in \mathbb{R}$, we have* Lemma 1.16 p.92

(1) $0 \cdot \vec{v} = \vec{0}$

(2) $(-1 \cdot \vec{v}) + \vec{v} = \vec{0}$

(3) $r \cdot \vec{0} = \vec{0}$.

Two.I.2. (p.96)

Definition. For any vector space, a subspace is a subset that is itself a vector space (under the inherited operations).

Example. The *trivial subspace* that consists of just the zero vector is always a subspace of any vector space. Similarly, the vector space itself is a subspace.

Example. The x-axis is a subspace of $\mathbb{R}$, given by $\left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^2 \mid x \in \mathbb{R} \right\}$.

Example. We have that $\mathcal{P}_2$ is a subspace of $\mathcal{P}_3$. In fact, any $\mathcal{P}_n$ is a subspace of $\mathcal{P}_m$ as long as $n \leq m$.

Definition. A *linear combination* of a set of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$ is an expression of the form

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n.$$

Note that this results in another vector in V.

The following Lemma is *very* helpful when it comes to determining whether something is a vector space. Most of the time, the vector spaces we are dealing with are subsets of known vector spaces. In that case, most of the vector space axioms are satisfied by the inherited operations, and we only need to check two of them (the closure related ones).

Lemma. *For any nonempty subset S of a vector space, under the inherited operations the following statements are equivalent.* Lemma 2.9 p.98

(1) S is a subspace of that vector space.

(2) S is closed under vector addition and under scalar multiplication. Formally:

$$\vec{v} + \vec{w} \ inS \ \text{and} \ r \cdot \vec{v} \in S \ \text{for all} \ \vec{v}, \vec{w} \in S, r \in \mathbb{R}.$$

(3) S is closed under taking linear combinations of two elements. Formally:

$$r_1 \cdot \vec{v} + r_2 \cdot \vec{w} \ \ \text{for all} \ \vec{v}, \vec{w} \in S, \ r_1, r_2 \in \mathbb{R}.$$

Lecture 2 (9/2)

Example. Let S be the following subset of $\mathbb{R}^3$:

$$S = \left\{ \begin{pmatrix} a \\ b \\ a+b \end{pmatrix} \mid a, b \in \mathbb{R} \right\}.$$

To show that S is closed under vector addition we checl that

$$\begin{pmatrix} a \\ b \\ a+b \end{pmatrix} + \begin{pmatrix} a \\ d \\ c+d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d \\ a+b+c+d \end{pmatrix} \in S,$$

and to check scalar multiplication, we see that

$$r \cdot \begin{pmatrix} a \\ b \\ a+b \end{pmatrix} = \begin{pmatrix} ra \\ rb \\ ra+rb \end{pmatrix} \in S.$$

Therefore, S is a subset of $\mathbb{R}^3$.

Example. Suppose that, instead we define

$$S = \left\{ \begin{pmatrix} a \\ b \\ a+b+1 \end{pmatrix} \mid a, b \in \mathbb{R} \right\}.$$

Then, S is not a subspace. For example, we see immediately that it does not contain the zero vector. (And it will also fail closure under both vector addition and scalar multiplication.)

Definition. For a set of vectors $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ we define the *span* of S as the set Def. 2.13 p.100

$$[S] = \{a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n \mid a_1, a_2, \dots, a_n \in \mathbb{R}\}.$$

In other words, the span of a set is the set of all linear combinations of that set.

Example. Is the vector $\binom{2}{3}$ in the span of the set $\{\binom{2}{2}, \binom{0}{3}\}$?

Yes, since $\binom{2}{3} = 1 \cdot \binom{2}{2} + \frac{1}{3} \cdot \binom{0}{3}$.

The following lemma follows quite directly from Lemma 2.9 p.98. The key idea is that if two vectors $\vec{v}$ and $\vec{w}$ are both linear combinations of a set S, i.e. they are in $[S]$, then any linear combination of $\vec{v}$ and $\vec{w}$ is also in $[S]$. Try to fill in the details of this proof yourself. (The proof is also in the book.)

Lemma. *In a vector space, the span of any subset is a subspace.* Lemma 2.15 p.100

Two.II (p.108)

Definition. A set of vectors is *linearly independent* if none of the vectors in the set is a linear combination of the others. Another way of saying this is that a set $S = \{\vec{v}_1, \dots, \vec{v}_n\}$ is linearly independent if the equation Def. 1.4 p.109

$$a_1\vec{v}_1 + \cdots + a_n\vec{v}_n = \vec{0}$$

only has one (trivial, meaning $a_1 = \cdots = a_n = 0$) solution. Otherwise, we say that the set is *linearly dependent.*

The following Corollary gives yet an alternative way of phrasing this. Try to prove this Corollary yourself from the previous definition (note that it is a bidirectional statement). The key idea is that if S is linearly independent, then $\vec{v}$ itself is in the span of S, but not in the span of $S \setminus \{\vec{v}\}$.

Corollary. *A set S is linearly independent if and only if the removal of any vector $\vec{v}$ from S shrinks the span, i.e. $[S \setminus \{\vec{v}\}] \subsetneq [S]$, $\forall \vec{v} \in S$.* Cor. 1.14 p.113

The next Corollary follows from the previous: if $\vec{v} \in S$ can be written as a linear combination of other vectors of S, then removing $\vec{v}$ does not shrink the span. Therefore, we can repeatedly remove vectors from a linearly dependent set without shrinking the span, until there is no more dependence left. Try to write this up formally yourself.

Corollary. *In a vector space, any finite set of vectors has a linearly independent subset with the same span.* Cor. 1.17 p.115

Two.III.1 (p.121)

Definition. A linearly independent set of vectors that span a vector space V is called a basis for V. Def. 1.1 p.121

Example. For $\mathbb{R}^n$ we call the following set the standard basis:

$$\mathcal{E}_n = \left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right) = (\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n).$$

In this case, it is rather easy to see that every vector in $\mathbb{R}^n$ is a linear combination of this basis, since

$$\begin{pmatrix} x_1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + \cdots + x_n \cdot \vec{e}_n.$$

The following Theorem describes a very important and useful property of bases of vector spaces. They give us a unique expression for every vector in the space. This means that if we have a basis, we can identify vectors with their expression in terms of that basis.

Theorem. *A subset S of a vector space V is a basis if and only if every vector in the space can be expressed as a linear combination of the subset in exactly one way.* Thm. 1.12 p.123

The key to proving this Theorem is the following Lemma:

Lemma. *A set S is linearly independent if and only if every vector $\vec{w} \in [S]$ can be expressed in only one way as a linear combination of S.*

Math!. Bidirectional statements are "if and only if" statements. Mathematicians often write "iff" for short, or we use double-arrow: $A \Leftrightarrow B$. To prove such a statement, we need to show $A \Rightarrow B$ as well as $B \Rightarrow A$. To make it easier to follow for the reader, it is best to separate the two directions clearly.

Math!. We will prove both directions of this lemma by proving the contrapositive statements. Contrapositive means that $A \Rightarrow B$ is the same as saying not-$B \Rightarrow$ not-A. For example, if I say that if it rains then I will bring an umbrella, then that is the same as saying that if I don't bring my umbrella it must not be raining.

Proof. "$\Rightarrow$" Suppose that $\vec{w} \in [S]$. We know that $\vec{w}$ can be expressed as a linear combination of $S = \{\vec{v}_1, \ldots, \vec{v}_n\}$, so assuming that it cannot be expressed in exactly one way is the same as assuming that it can be expressed in at least two ways (i.e. we've ruled out 0). So, suppose $\vec{w}$ can be expressed in two different ways as a linear combination of S. Then we have

$$\begin{aligned}\vec{w} &= a_1\vec{v}_1 + \cdots + a_n\vec{v}_n \\ &= b_1\vec{v}_1 + \cdots + b_n\vec{v}_n\end{aligned}$$

with $a_i \neq b_i$ for at least one value of $1 \leq i \leq n$. This implies that we have a non-trivial solution to the relation

$$\vec{w} - \vec{w} = (a_1 - b_1)\vec{v}_1 + \cdots + (a_n - b_n)\vec{v}_n = \vec{0}.$$

Therefore, the set S is linearly dependent.

"$\Leftarrow$" The reverse direction of the proof looks very similar. Suppose that the set S is linearly dependent. Then we have a non-trivial relation

$$c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0},$$

i.e. not all values of c_i, $1 \leq i \leq n$, are 0. Then if $\vec{w} \in [S]$, we have some expression $\vec{w} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n$ and we can find a second, different expression of $\vec{w}$ by adding $\vec{0}$ to both sides of the equation and obtaining $\vec{w} = \vec{w} + \vec{0} = (a_1 + c_1)\vec{v}_1 + \cdots + (a_n + c_n)\vec{v}_n$. Therefore we have at least two ways of writing $\vec{w}$ as a linear combination of S.

□

The following definition hinges on the fact that all bases of a vector space have the same cardinality (number of elements). We have not proven that this is true yet, so that will be the focus of the next lecture. The reason that this is important is that it shows that the dimension of a vector space is an inherent property of the space, and does not depend on the choice of basis used to describe the vector space.

Definition. The dimension of a vector space V, denoted $\dim(V)$ is the number of elements in a basis for V. Def. 2.5 p.131

Lecture 3 (9/7)

Two.III.2 (p.129)

The Exchange Lemma is foundational to the concept of bases. My favorite way of phrasing the exchange lemma is (I think) easier to understand/remember, but one step harder to prove than the way it is phrased in the following lemma. So, we will start with proving Lemma 2.3, then rephrasing and then using it to argue that all bases have the same cardinality.

Lemma. *Suppose that $B = (\vec{\beta}_1, \vec{\beta}_2, \dots, \vec{\beta}_n)$ is a basis for a vector space V, and* Lemma 2.3 p.130

$$\vec{v} = c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n, \text{ such that } c_i \neq 0.$$

Then, replacing $\vec{\beta}_i$ by $\vec{v}$ gives a new basis for V.

Math!. We will prove some of this lemma by contradiction. This means that we will first assume that what we are trying to prove is not true, and then show that this implies an impossibility. Often proofs will just start by saying "Suppose that" followed by the negation of the statement we're trying to prove, and then the reader understands that this will be a proof by contradiction. I like to be extra clear and add something along the lines of "Suppose that... (for the sake of contradiction)".

Math!. A common form of a statement in math is the statement that two sets are the same: $A = B$ for two sets. This means that every element in A is also an element in B, and every element in B is also an element in A. In other words, it means that $A \subseteq B$ (A is a subset of B) and $B \subseteq A$ (B is a subset of A).

Math!. Proving that $A \subseteq B$. As we said in the previous item, this is the same as saying that every element of A is also an element of B. We can prove this by using the following proof format: First, suppose that $x \in A$ (where x must be an arbitrary element of A, since we want the following to be true for all of them). Then, if we can prove that this leads to the conclusion that $x \in B$, we are done.

Proof. We let this new basis be $\hat{B} = B \setminus \{\vec{\beta}_i\} \cup \{\vec{v}\}$. To show that $\hat{B}$ is a basis for V we need to show that it spans V and that it is linearly independent.

- **Linear independence:**
  Suppose that $\hat{B}$ is not linearly independent (for the sake of contradiction). Then we have some nontrivial relation
  $$d_1\vec{\beta}_1 + \cdots + d_i\vec{v} + \cdots + d_n\vec{\beta}_n = \vec{0}.$$
  This relation is nontrivial, so we know that at least some d_j is nonzero. We cannot have $d_i = 0$, since then this would turn out to be a nontrivial relation on B, which is not possible. So, $d_i \neq 0$. Now we can replace $\vec{v}$ with its expression in terms of B, and obtain
  $$d_1\vec{\beta}_1 + \cdots + d_i(c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n) + \cdots + d_n\vec{\beta}_n = \vec{0}$$
  but this can be rewritten as
  $$(d_1 + d_ic_1)\vec{\beta}_1 + \cdots + d_ic_i\vec{\beta}_i + \cdots + (d_n + d_ic_n)\vec{\beta}_n = \vec{0}$$
  and since we have $d_ic_i \neq 0$ this is a nontrivial relation on B. This is not possible, since B is a basis and therefore linearly independent. This completes the proof, since we showed that if we assume that $\hat{B}$ is linearly independent we must accept a contradiction. Therefore it must be false and we see that $\hat{B}$ is linearly independent as we had hoped.
- Next, we want to show that $\hat{B}$ spans V. In fact, we want to show that $\hat{B}$ spans exactly V (and nothing more), since we are trying to say that $\hat{B}$ is an alternative basis for V. So, the statement we are trying to prove is $[B] = [\hat{B}]$. We split this up into the two parts:
  - $[\hat{B}] \subseteq [B]$. Suppose that $\vec{w} \in [\hat{B}]$. Then, (Note that I am reusing the symbols $d_1, \dots, d_n$ here. They are not the same constants as in the previous part. However, $c_1, \dots, c_n$ are still the same since that is how we defined $\vec{v}$. It's always tricky to decide when to reuse symbols so that we don't run out of letters/confuse the reader.)
    $$\begin{aligned}\vec{w} &= d_1\vec{\beta}_1 + \cdots + d_i\vec{v} + \cdots + d_n\vec{\beta}_n, \text{ for some } d_1, \dots, d_n \in \mathbb{R},\\ &= d_1\vec{\beta}_1 + \cdots + d_i(c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n) + \cdots + d_n\vec{\beta}_n\\ &= (d_1 + d_ic_1)\vec{\beta}_1 + \cdots + d_ic_i\vec{\beta}_i + \cdots + (d_n + d_ic_n)\vec{\beta}_n.\end{aligned}$$
    Since this is a linear combination of elements of B, we see that $\vec{w} \in [B]$.

– $[B] \subseteq [\hat{B}]$. First, we rewrite the expression of $\vec{v}$ in terms of B so that we separate $\vec{\beta}_i$. This proof is somewhat the reverse of the previous.

$$\begin{aligned}\vec{v} &= c_1\vec{\beta}_1 + \cdots + c_i\vec{\beta}_i + \cdots + c_n\vec{\beta}_n \\ \vec{\beta}_i &= -\tfrac{c_1}{c_i}\vec{\beta}_1 + \cdots + \tfrac{1}{c_i}\vec{v} + \cdots - \tfrac{c_n}{c_i}\vec{\beta}_n\end{aligned}$$

Note that it is important here to remember that $c_i \neq 0$. Otherwise there would be a problem with the fractions.
Supose that $w \in [B]$. Then we have (once again, new $d_1, \ldots, d_n$)

$$\begin{aligned}\vec{w} &= d_1\vec{\beta}_1 + \cdots + d_i\vec{\beta}_i + \cdots + d_n\vec{\beta}_n, \text{ for some } d_1, \ldots, d_n \in \mathbb{R}, \\ &= d_1\vec{\beta}_1 + \cdots + d_i\left(-\tfrac{c_1}{c_i}\vec{\beta}_1 + \cdots + \tfrac{1}{c_i}\vec{v} + \cdots - \tfrac{c_n}{c_i}\vec{\beta}_n\right) + \cdots + d_n\vec{\beta}_n \\ &= \left(d_1 - d_i\tfrac{c_1}{c_i}\right)\vec{\beta}_1 + \cdots + \tfrac{d_i}{c_i}\vec{v} + \cdots + \left(d_n - d_i\tfrac{c_n}{c_i}\right)\vec{\beta}_n.\end{aligned}$$

The constant coefficients look a bit complicated here, but the only thing that matters is that they are real numbers. What we have is an expression of $\vec{w}$ as a linear combination of the elements of $\hat{B}$. Therefore, $\vec{w} \in [\hat{B}]$.

□

Lecture 4 (9/9)

Now, we can rewrite the exchange lemma in my favorite way.

Lemma. *If $B_1 = (\vec{\alpha}_1, \ldots, \vec{\alpha}_n)$ and $B_2 = (\vec{\beta}_1, \ldots, \vec{\beta}_m)$ are two bases for a vector space V, then for any $\alpha_i \in B_1$, there exists a $\vec{\beta}_j \in B_2$ such that $B_1 \setminus \{\vec{\alpha}_i\} \cup \{\vec{\beta}_j\}$ (replace $\vec{\alpha}_i$ by $\vec{\beta}_j$ in B_1) is also a basis for V.*

Proof. By Lemma 2.3 p. 130, all we need is to show that B_2 contains a $\vec{\beta}_j$ such that in the expression $\vec{\beta}_j = c_1\vec{\alpha}_1 + \cdots + c_n\vec{\alpha}_n$ we have $a_i \neq 0$. Suppose (for the sake of contradiction) that there is no such $\vec{\beta}_j$. Then every vector in B_2 can be expressed as a linear combination of vectors in $B_1 \setminus \{\vec{\alpha}_i\}$. Sinve $\vec{\alpha}_i \in V$ and B_2 is a basis for V, we can write

$$\begin{aligned}\vec{\alpha}_i &= a_1\vec{\beta}_1 + \cdots + a_m\vec{\beta}_m \\ &= a_1(c_{11}\vec{\alpha}_1 + \cdots + c_{1i-1}\vec{\alpha}_{i-1} + c_{1i+1}\vec{\alpha}_{i+1} + \cdots + c_{1n}\vec{\alpha}_n) + \ldots \\ &\quad + a_m(c_{m1}\vec{\alpha}_1 + \cdots + c_{mi-1}\vec{\alpha}_{i-1} + c_{mi+1}\vec{\alpha}_{i+1} + \cdots + c_{mn}\vec{\alpha}_n).\end{aligned}$$

Again, we have somewhat complicated constants, but the point is that we can write this in the form

$$\vec{\alpha}_i = d_1\vec{\alpha}_1 + \cdots + d_{i-1}\vec{\alpha}_{i-1} + d_{i+1}\vec{\alpha}_{i+1} + \cdots + d_n\vec{\alpha}_n.$$

This implies that $\vec{\alpha}_i$ is dependent on the other elements of B_1, which implies that B_1 is not linearly independent. This is a contradiction, and therefore, there must exist a $\vec{\beta}_j$ as described.

□

In the previous result we wrote that B_1 has n elements and B_2 has m elements, because we did not know for sure yet that these must be equal. Now, we can finally prove this.

Theorem. *For finite-dimensional vector spaces, all bases have the same number of elements. (The same cardinality.)* Thm 2.4 p.131

Proof. Suppose that V is a finite-dimensional vector space and that B_1, B_2 are two bases, with $|B_1| < |B_2|$ (for the sake of contradiction). If B_1 has elements that are not in B_2, then we can replace them with elements from B_2 one by one, by the previous lemma, to obtain a new basis B_1' for V, with $B_1 \subsetneq B_2$. This implies that any remaining elements in $B_2 \setminus B_1'$ can be written as linear combinations of B_1', but then they can be written in terms of other elements of B_2. This is a contradiction, since B_2 must be linearly independent. □

The following two corollaries from the book use very similar arguments. They are very useful and you should try to remember them. Try to prove them yourself, and check the proofs in the book if necessary.

Corollary. *Any linearly independent set in V can be extended to a basis.* Cor. 2.12 p.132

Corollary. *Any spanning set of V can be shrunk to a basis.* Cor. 2.13 p.132

Definition. For a given vectors space V and basis B, we can express each element in V in terms of (as a linear combination) of B. All we need to define an element $\vec{v} \in V$ as a linear combination of $B = (\vec{\beta}_1, \ldots, \vec{\beta}_n)$ is the unique list of coefficients $a_1, \ldots, a_n$ in the expression $\vec{v} = a_1\vec{\beta}_1 + \ldots a_n\vec{\beta}_n$. We write this as Def. 1.13 p.124

$$\text{Rep}_B(\vec{v}) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}_B .$$

Example. If we let $B = (\binom{2}{0}, \binom{1}{3})$ be a basis for $\mathbb{R}^2$. Then, for example, we have

$$\text{Rep}_B\left(\binom{1}{1}\right) = \binom{1/3}{1/3}_B,$$

since we have

$$\binom{1}{1} = \frac{1}{3}\binom{2}{0} + \frac{1}{3}\binom{1}{3}.$$

Example. Consider the subspace

$$S = \{a_0 + a_1x + a_2x^2 \in \mathcal{P}_2 \mid a_0 = a_2\}.$$

This space has a basis (for example) given by $B = (1 + x^2, x)$. We have $2 - x + 2x^2 \in S$, and we can write

$$2 - x + 2x^2 = 2 \cdot (1 + x^2) + (-1) \cdot x,$$

which gives

$$\text{Rep}_B(2 - x + 2x^2) = \binom{2}{-1}_B.$$

Matrix basics

We will introduce matrices here, mainly as functions that map vectors to vectors. More precisely, they are so-called linear transformations of vectors spaces. A matrix is an m-by-n array of numbers. We refer to the elements of a matrix A by a_{ij}, meaning the element in the ith row and jth column:

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}.$$

Definition. A *linear transformation* is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that it can be written as $T(\vec{x}) = A\vec{x}$, where A is an $m \times n$ matrix.

In order to understand how to multiply matrices with vectors, we first define the dot product of vectors. First of all, we denote vectors as follows:

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{pmatrix}.$$

The dot product of two vectors $\vec{v} \cdot \vec{w}$ is defined as follows:

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n.$$

Note that the dot product of two vectors is a scalar, and that it is only defined if the two vectors have the same dimension. When we multiply a matrix with a vector, we can think of it in terms of dot products, either by thinking of A as a set of horizontal row vectors:

$$A\vec{x} = \begin{pmatrix} \leftarrow & \vec{w_1} & \rightarrow \\ \leftarrow & \vec{w_2} & \rightarrow \\ & \vdots & \\ \leftarrow & \vec{w_n} & \rightarrow \end{pmatrix} \vec{x} = \begin{pmatrix} \vec{w_1} \cdot \vec{x} \\ . \\ . \\ . \\ \vec{w_n} \cdot \vec{x} \end{pmatrix},$$

or as a set of vertical column vectors:

$$A\vec{x} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_m \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \vec{v_1} + \ldots + x_n \vec{v_n}.$$

In either case, the result is a vector in $\mathbb{R}^m$. Suppose that we are given an equation $A\vec{x} = \vec{w}$ where $\vec{x}$ is unknown. Then

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$$

can be rewritten as (if we think of the dot product version of multiplication)

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= w_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= w_2 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= w_m \end{aligned}$$

A system of linear equations! From now on, we will think of systems of linear equations as equivalent to these matrix equations. Also, not that if we rewrite $A\vec{x} = \vec{w}$ with the column-vector version of multiplication, we see that this is equivalent to solving

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_m \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \vec{v_1} + \ldots + x_n \vec{v_n} = \vec{w}.$$

This is equivalent to asking whether $\vec{w}$ is in the span of the column vectors of A.

Lecture 5 (9/14)

The *column rank* of a matrix is the dimension of the (sub)space spanned by its column vectors. Similarly, the *row rank* of a matrix is the dimension of the (sub)space spanned by its rows. We will see that these two ranks are always equal, which will help us to define a single notion of rank of a matrix.

Example. Consider the matrix

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

The span of its columns is

$$\left[\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right) \right].$$

Since $\left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)$ is not a scalar multiple of $\left(\begin{smallmatrix} 1 \\ 1 \\ 2 \end{smallmatrix}\right)$, we see that this is an independent set, and therefore this is a 2-dimensional subspace. Similarly, we see that

$$\left[\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right) \right]$$

also is a 2-dimensional space (in fact this is $\mathbb{R}^2$).

We will return to column and row ranks after we discuss using matrices to solve systems of equations, but try to keep them in the back of your mind in the mean-time, and think about how they might relate to the solution sets.

Reduced row echelon form

We will start representing systems of linear equations as matrices, and then solving them using matrix operations. This method is known as **Gauss-Jordan elimination**. For example, the system

$$\begin{aligned} 3x + 6y - z &= 3 \\ 2x - 4y + 3z &= 2 \end{aligned}$$

can be represented by its **coefficient matrix**, which contains the coefficients of each variable in the equations:

$$\begin{pmatrix} 3 & 6 & -1 \\ 2 & -4 & 3 \end{pmatrix}$$

or its **augmented matrix**, which contains all of the information in the system of equations:

$$\left(\begin{array}{ccc|c} 3 & 6 & -1 & 3 \\ 2 & -4 & 3 & 2 \end{array}\right).$$

In order to solve the system of equations, we would like to write the augmented matrix into a standardized form, the so-called **reduced row echelon form**. A matrix in reduced row echelon form satisfies the following properties: Ch. One p.50

- The leading (first non-zero) entry in each row is 1.
- A column that contains such a leading entry has 0 everywhere else.

- The rows are in order of the position of their leading coefficients.

Example. Here are some examples of matrices in RREF, together with solutions to the system. Write down the corresponding equations yourself. In the last example, the variable z does not have its own "leading 1", and is therefore free. We give it an arbitrary value t. The variable y is then dependent on $z = t$. (We could have let z depend on $y = t$ instead, but the rref gives us a canonical method in which the earlier variable receives the leading 1 and is named as the dependent variable.)

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

- Inconsistent
- 0 solutions
(one of the equations is 0=1)

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right).$$

- Consistent
- Independent
- 1 solution

$x = 1$
$y = 2$
$z = 3$

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

- Consistent
- Dependent
- ∞ solution

$x = 1$
$y + 2z = 2 \to y = 2 - 2t$
$z = t$

The reduced row echelon form of a matrix is unique, and can be achieved by using a series of the following **elementary row operations**.

- Swapping two rows.
- Multiplying a row by a non-zero scalar.
- Adding a multiple of one row to another row.

We use the following general algorithm for rewriting a matrix into its RREF.

Algorithm 1 RREF

for row $i = 1 \dots n$ **do**
  let a_{ij} be the leading coefficient of row i;
  divide row i by a_{ij} so that a_{ij} becomes 1;
  subtract row i a_{kj} times from all other rows so that a_{kj} becomes 0 for all $k \neq i$;
end for
arrange the rows in order of the position of their leading coefficients;

Once you have written your system of linear equations as an augmented matrix, and then written the augmented matrix into reduced row echelon form, you can find the solutions.

- If the RREF matrix contains a row where the final entry is non-zero and all other entries are 0, then the system is **inconsistent** and there are no solutions. This is the equivalent of having one of your equations be $0 = 1$.
- If the RREF is a **diagonal matrix** if you ignore the last column and any 0-rows, then the system is **consistent** and **independent**, and it has exactly one solution. This is equivalent to your system of equations looking like $x = a,\ y = b,\ z = c, \dots$
- Otherwise, the system is **consistent** and **dependent**, and it has infinitely many solutions. This is equivalent to having a set of equations that looks something like $x = a + b \cdot z, y = c + d \cdot z, \dots$. You can let z take any value and this will determine x and y.

Lecture 6 (9/16)

We now return to the row and column ranks of A. First we will argue that the RREF of A has the same row rank as A as well as the same column rank as A. Then, we argue that for a matrix in RREF, the row and column ranks are equal.

Claim. *Row operations do not affect the span of the rows of a matrix A.*

Proof. This proof is a good exercise for understanding vector spans. We need to prove the following statements:

- Reordering the vectors in an ordered set S does not change its span.
- Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, and let $S' = \{c \cdot \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, where $c \neq 0$. Then, $[S] = [S']$.
- Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, and let $S' = \{\vec{v}_1, \vec{v}_2 + c \cdot \vec{v}_1, \ldots, \vec{v}_n\}$, where $c \in \mathbb{R}$. Then, $[S] = [S']$. Note that for this statement, you can start by proving this for $S = \{\vec{v}_1, \vec{v}_2\}$.

□

Claim. *Row operations do not affect the span of the rows of a matrix A.*

Proof. Let the right-hand-side of each equation in a system of linear equations arising from A be 0. Then row operations do not affect the last column in the augmented matrix. Therefore, a solution to the system is a relation on the columns of A, as well as a relation on the columns of the RREF. Since A and its RREF have the same set of relations, they have the same linear dependencies between their column vectors, and therefore the same rank. □

Claim. *The row and column rank of a matrix are equal.*

Proof. In light of the previous two claims, we only need to show that the claim is true for a matrix in RREF form. We will do this by showing that the ranks are equal to the number of leading 1s (sometimes called pivots).

For the row rank, note that the rows with leading 1s are the only non-zero rows. Since each has a 1 in a position where no other row has a non-zero entry, none of them can be written as a linear combination of the others.

For the column rank, first consider the columns that have leading 1s. Since each has a 1 in a position where no other leading-1 column has a non-zero entry, none of them can be written as a linear combination of the others. So, this set is linearly independent. Any other non-zero column has entries only in positions for which there is a leading-1 column, and therefore can be written as a linear combination of the leading-1 columns. This shows that the leading-1 columns are a basis for the column span, with dimension equal to the number of leading 1s. □

Lectures 7-9

Review + Midterm 1.

Lecture 10 (9/30)

We are now starting in Chapter Three, which is on maps between spaces. Ch. Three p.173

Math!. The mathematical notation for a map (or function) from a set A to a set B is $f : A \to B$. This means that for every $a \in A$ there is a unique $b \in B$ such that $f(a) = b$. We say that such a map is *injective* if there are no two values in A mapping to the same value in B, i.e. $f(a_1) = f(a_2)$ implies that $a_1 = a_2$. We say that a map is *surjective* if every $b \in B$ has some $a \in A$ mapping to it, i.e. for every $b \in B$ there exists an $a \in A$ such that $f(a) = b$. A map is *bijective* if it is both injective and surjective.

Definition. Let V and W be two vector spaces. A *homomorphism* is a map $f : V \to W$ that "preserves the structure of the vector space V". Formally, we have that

$$f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2), \ \ \forall \vec{v}_1, \vec{v}_2 \in V,$$

i.e. the map preserves vector addition, and

$$f(r \cdot \vec{v}) = r \cdot f(\vec{v}), \ \ \forall r \in \mathbb{R}, \vec{v} \in V,$$

i.e. it preserves scalar multiplication. This is equivalent to saying that the map preserves linear combinations:

$$f(r_1 \cdot \vec{v}_1 + r_2 \cdot \vec{v}_2) = r_1 \cdot f(\vec{v}_1) + r_2 \cdot f(\vec{v}_2).$$

Definition. A map $f : V \to W$ is an *isomorphism* if it is a bijective homomorphism. An isomorphism $f : V \to V$ is called an *automorphism*. Def. 1.3 p.172

In the book (and in the lecture), we prove the following Lemma.

Lemma. *The inverse of an isomoprhism is also an isomorphism.* Lem. 2.1 p.183

This is an important Lemma, because it helps us describe pairs of vector spaces as isomorphic to one another, meaning equivalent in some way. More precisely, the isomorphism relation on vector spaces turns out to be an *equivalence relation*.

Math!. A relation R on a set A is a set of ordered pairs of elements from A. Formally, $R \subseteq A^2$. We call R an equivalence relation, denoted $a \sim b$ (meaning $(a, b) \in R$) if R is

- reflexive: $a \sim a$ for all $a \in A$,
- symmetric: $a \sim b \ \Leftrightarrow \ b \sim a$, for all $a, b \in A$,
- transitive: $a \sim b, \ b \sim c \ \Rightarrow \ a \sim c$ for all $a, b, c \in A$.

Equivalence relations give rise to partitions of a set A: a classification of the elements of A into classes, such that a and b are in the same class if and only if $a \sim b$, and every element is in exactly one class.

By the way, you have seen another type of equivalence relation already, but we didn't call it that. We can say that two matrices are equivalent if they have the same rref. Our next challenge is to prove the following Theorem, as it will help us show that all vector spaces of the same dimension are isomorphic, and therefore all n-dimensional vector spaces are isomorphic to $\mathbb{R}^n$.

Theorem. *Isomorphism defines an equivalence relation on vector spaces.* Thm. 2.2 p.184

Lecture 11 (10/5)

We will now start to prove the Theorem above (Thm. 2.2 p.184). It turns out that the classification given by isomorphism as an equivalence relation on vector spaces is quite easy to characterize: vector spaces are isomorphic if and only if they have the same dimension. Therefore, we prove the following instead:

Theorem. *For two vector spaces U and W, we have that $U \simeq W$ if and only if* $\dim U = \dim W$. Thm. 2.3 p.185

Proof. See proof in the book. □

We now have, in particular, that any n-dimensional vector space is simply isomorphic to $\mathbb{R}^n$. Therefore, we can use our knowledge and tools (linear dependence, rref, etc) for $\mathbb{R}^n$ on other vector spaces. In the following sections, we will start to see that linear maps $f : V \to W$ can always be thought of as matrix functions, such that $f(\vec{v}) = A\vec{v}$. In order to do that, we do need that $\vec{v}$ is a vector in the sense that you are used to: a vertical array of n real numbers. Because of the isomorphism between V and $\mathbb{R}^n$, we can indeed do that.

Lecture 12 (10/7)

Today we look at two important subspaces that are related to a linear map $f : V \to W$: the null space and range space. These contain the information "lost" and "retained" from V, respectively. Please do not use this as definitions; it's just to help you gain intuition.

Definition. The *null space* or *kernel* of a linear map $f : V \to W$ is the inverse image of $\vec{0}_W$: Def. 2.11 p.204

$$\mathcal{N}(f) = f^{-1}(\vec{0}_W) = \{\vec{v} \in V \mid f(\vec{v}) = \vec{0}_W\}.$$

We also define the *nullity* of f as the dimension of the kernel:

$$\text{nullity}(f) = \dim \mathcal{N}(f).$$

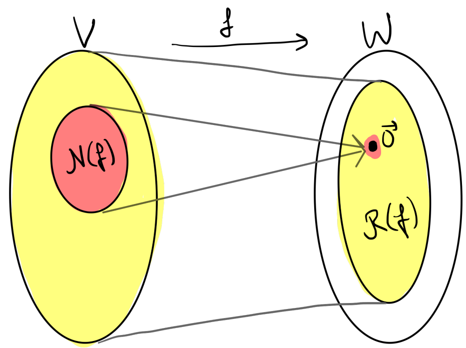
Definition. The *range space* of a linear map $f : V \to W$ is defined as: Def. 2.2 p.200

$$\mathcal{R}(f) = \{f(\vec{v}) \mid \vec{v} \in V \mid f(\vec{v}) = \{\vec{w} \in W \mid f(\vec{v}) = \vec{w} \text{ for some } \vec{v} \in V\}.$$

We also define the *rank* of f as the dimension of the kernel:

$$\text{rank}(f) = \dim \mathcal{R}(f).$$

Note that we already defined the rank of matrices. You will see that we are talking about the same rank in both cases, as we connect linear maps to matrices.
It is important here to note that $\mathcal{N}(f)$ is a subspace (as we shall see) of V, while $\mathcal{R}(f)$ is a subsspace of W. Sketch:



Indeed, these are both vector spaces themselves.

Lemma. *For any linear map $f : V \to W$, null space $\mathcal{N}(f)$ is a subspace of V.* Lemma 2.10 p.204

Lemma 2.10 on p.204 is in fact a bit stronger, as it shows that the inverse image of any subspace is in fact a subspace. The trivial subspave $\{\vec{0}_W\}$ is then a special case of that.

Lemma. *For any linear map $f : V \to W$, range space $\mathcal{R}(f)$ is a subspace of W.* Lemma 2.1 p.200

Now, we are ready to prove a very important theorem. It turns out that the rank and nullity of a linear map are two sides of the same coin. Of the dimensions of V, some are lost and the rest is retained. Or, more formally:

Theorem. *For any linear map $f : V \to W$, we have* Thm. 2.14 p.205

$$\text{rank}(f) + \text{nullity}(f) = \dim V.$$

Note that $\dim W$ is not relevant to this statement. Since $\mathcal{R} \subseteq W$, we do know that

$$\text{rank}(f) \leq \dim W.$$

Example. Consider the following map $f : \mathcal{P}_3 \to \mathcal{P}_2$, defined by

$$f(p(x)) = \frac{d}{dx}p(x),$$

or

$$f(a + bx + cx^2 + dx^3) = b + 2cx + 3dx^2.$$

What is the null space of f? This is exactly the set of polynomials with 0 derivative everywhere, i.e. constant functions of the form $p(x) = a$. Therefore, we can let $\langle 1 \rangle$ be a basis for $\mathcal{N}(f)$, and we see that $\text{nullity}(f) = 1$. What is the range space of f? We see that every $p(x) \in \mathcal{P}_2$ is the derivative of some $q(x) \in \mathcal{P}_3$. Namely, a polynomial of the form $a + bx + cx^2$ is the derivative of $ax + (b/2)x^2 + (c/3)x^3$, for example. Therefore, this map is surjective and $\mathcal{R}(f) = \mathcal{P}_2$, and $\text{rank}(f) = 3$.

Lecture 13 (10/12)

The following Theorem summarizes what we know about invertible maps (maps that don't "lose any information"). We have proven some parts carefully in class, other parts you can try as an exercise or refer to in the book.

Theorem. *For a linear map $f : V \to W$, $\dim V = n$, the following are equivalent:* Thm. 2.20 p.207

(1) The map f is injective (or 1-1)

(2) f has an inverse function f^{-1} which is a linear map $\mathcal{R}(f) \to V$

(3) $\text{nullity}(f) = 0$

(4) $\text{rank}(f) = n$

(5) if $\langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$ is a basis for V then $\langle f(\vec{\beta}_1), \ldots, f(\vec{\beta}_n) \rangle$ is a basis for $\mathcal{R}(f)$.

Inverse Images

As we have seen, some linear maps are fully invertible, such that $f^{-1} : W \to V$. We call such maps isomorphisms (or automorphisms if the domain and codomain are equal). Even if a map is not invertible, we can still define an *inverse image*, for $S \subseteq W$: p.201

$$f^{-1}(S) = \{\vec{v} \in V \mid f(\vec{v}) \in S\}.$$

For example $f^{-1}(\mathcal{R}(f)) = V$ and $f^{-1}(\vec{0}) = \mathcal{N}(f)$.

If the map f is not injective and we have two distinct elements $\vec{v}_1, \vec{v}_2$ such that $f(\vec{v}_1) = f(\vec{v}_2)$, then their difference is an element in $\mathcal{N}(f)$, because

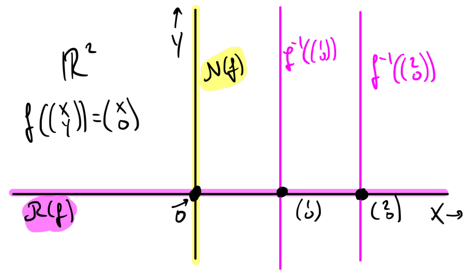
$$f(\vec{v}_1 - \vec{v}_2) = f(\vec{v}_1) - f(\vec{v}_2) = \vec{0},$$

which implies that $\vec{v}_1 - \vec{v}_2 \in \mathcal{N}(f)$.

Example. Consider the projection in $\mathbb{R}^2$ onto the x-axis, such that $f\left(\binom{x}{y}\right) = \binom{x}{0}$. Then $\mathcal{R}(f) = [(\,1//0\,)]$ and $\mathcal{N}(f) = [(\,0//1\,)]$. For each $\vec{w} = (\,w//0\,) \in \mathcal{R}(f)$ we have that

$$f^{-1}(\vec{w}) = \{\binom{w}{a} \mid a \in \mathbb{R}\}.$$

We can sketch this as follows:



Example. Let $f : \mathcal{P}_2 \to \mathcal{P}_2$ with $f(p(x)) = \frac{d}{dx}p(x)$. Then what is $f^{-1}(1 + 2x)$? This is the set of all $p(x) = a + bx + cx^2$ such that $\frac{d}{dx}p(x) = 1 + 2x$. These are polynomials of the form $q(x) = c + x + x^2$ for any $x \in \mathbb{R}$. Indeed any such two functions differ only in a constant, and nullity$(f) = [1]$.

Linear maps as matrices

Now, we will finally show formally that linear maps are exactly the maps that can be expressed as left-multiplication of vectors by matrices. Consider a linear map $f : V \to W$, and let $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ be a basis for V. Then the list $f(\vec{\beta}_1), \dots, f(\vec{\beta}_n)$ fully defines f, since for any $\vec{v} \in V$ we have $\vec{v} = c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n$ and Def. 1.2 p.214

$$f(\vec{v}) = f(c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n) = c_1 f(\vec{\beta}_1) + \cdots + c_n f(\vec{\beta}_n).$$

Suppose $\dim W = m$ and D is a basis for W. Then every $f(\vec{v}) = \vec{w}$ has a representation in terms of D. We let

$$\text{Rep}_D(f(\vec{\beta}_1)) = \begin{pmatrix} h_{1,1} \\ h_{2,1} \\ \vdots \\ h_{m,1} \end{pmatrix}, \text{Rep}_D(f(\vec{\beta}_2)) = \begin{pmatrix} h_{1,2} \\ h_{2,2} \\ \vdots \\ h_{m,2} \end{pmatrix}, \dots, \text{Rep}_D(f(\vec{\beta}_n)) = \begin{pmatrix} h_{1,n} \\ h_{2,n} \\ \vdots \\ h_{m,n} \end{pmatrix}.$$

Then

$$f(\vec{v}) = \begin{pmatrix} h_{1,1} & \cdots & h_{1,n} \\ \vdots & & \vdots \\ h_{m,1} & \cdots & h_{m,n} \end{pmatrix} \text{Rep}_B(\vec{v}) = \text{Rep}_D(\vec{w}),$$

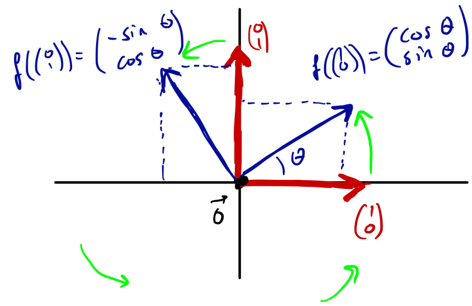
since

$$\begin{pmatrix} h_{1,1} & \cdots & h_{1,n} \\ \vdots & & \vdots \\ h_{m,1} & \cdots & h_{m,n} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = c1 \begin{pmatrix} h_{1,1} \\ \vdots \\ h_{m,1} \end{pmatrix} + \cdots + c_n \begin{pmatrix} h_{1,n} \\ \vdots \\ h_{m,n} \end{pmatrix} = c_1 f(\vec{\beta}_1)_D + \cdots + c_n f(\vec{\beta}_n)_D.$$

Example. Consider the linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by a rotation of the plane by an angle θ (counterclockwise). Let's use the standard basis $E = \langle \vec{e}_1, \vec{e}_2 \rangle = \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle$ as a basis for both the domain and codomain (they are both $\mathbb{R}^2$). Then we see that $f(\vec{e}_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and $f(\vec{e}_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$. Therefore

$$\text{Rep}_{E,E}(f) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}_{E,E}.$$

Note that we will often omit the basis subscript when working in standard bases.

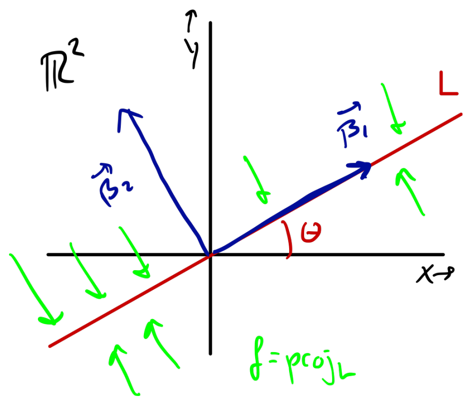


Example. In order to see why a non-standard basis might be useful, consider a projection $\text{proj}_L : \mathbb{R}^2 \to \mathbb{R}^2$ to a line L which is defined by an angle θ with the positive x-axis. We use the basis $B = \langle \vec{\beta}_1, \vec{\beta}_2 \rangle = \langle \left(\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right), \left(\begin{smallmatrix} -\sin\theta \\ \cos\theta \end{smallmatrix}\right) \rangle$ (notice any similarity with the previous example? what is the relation?). Then

$$\text{Rep}_B\, f(\vec{\beta}_1) = \left(\begin{smallmatrix} 1 \\ 0\theta \end{smallmatrix}\right)_B, \text{Rep}_B\, f(\vec{\beta}_2) = \left(\begin{smallmatrix} 0 \\ 0\theta \end{smallmatrix}\right)_B.$$

Therefore,

$$\text{Rep}_{B,B}(f) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0\theta \end{smallmatrix}\right)_{B,B}.$$



Lecture 14 (10/14)

Combining linear maps

Adding/scaling maps

If $f : \mathbb{R}^n \to \mathbb{R}^m$ that maps $\vec{v} \mapsto \vec{w}$ we let cf, for any scalar c, be the map that maps $\vec{v} \mapsto c\vec{w}$. It is not so difficult to deduce that if the matrix A represents f, then cA represents cf, where scalar multiplication of matrices is defined similarly to vectors (entry-wise): Sec. IV.1 p.232

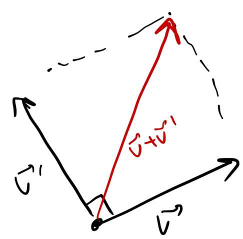
$$cA = c \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} = \begin{pmatrix} ca_{1,1} & \cdots & ca_{1,n} \\ \vdots & & \vdots \\ ca_{m,1} & \cdots & ca_{m,n} \end{pmatrix}.$$

If $f, h : \mathbb{R}^n \to \mathbb{R}^m$ then let $f + h : \mathbb{R}^n \to \mathbb{R}^m$ be defined as $(f+h)(\vec{v}) = f(\vec{v}) + h(\vec{v})$. If $f(\vec{v}) = A\vec{v}$ and $h(\vec{v}) = B\vec{v}$, then $(f+h)\vec{v} = (A+B)\vec{v}$, where matrix addition is defined entry-wise:

$$A + B = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1}+b_{1,1} & \cdots & a_{1,n}+b_{1,n} \\ \vdots & & \vdots \\ a_{m,1}+b_{m,1} & \cdots & a_{m,n}+b_{m,n} \end{pmatrix}.$$

Example. Consider the linear transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates every vector over a $\pi/4$ angle and scales it by $\sqrt{2}$. If $\vec{v}'$ is the vector $\vec{v}$ rotated over $\pi/2$ then we can think of this as $\vec{v} \mapsto \vec{v} + \vec{v}'$. Let f be the identity transformation and h a rotation over $\pi/2$. Then we should obtain $g = h + f$. We have $f\vec{v} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\vec{v}$ and $h\vec{v} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\vec{v}$ (see the rotation example on page 16). Then g has matrix $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

We may also let g' be the linear map given by a rotation over $\pi/4$, which has matrix $\begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$. Then the map g is a scaling of this map by the scalar $\sqrt{2}$, which indeed gives us the representing matrix $\sqrt{2}\begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.



Composing maps

Suppose that we have two linear maps $f : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^p \to \mathbb{R}^m$. Let $h \circ f$ be the composition of the two functions: Sec. IV.2 p.236

$$h \circ f(\vec{v}) = h(f(\vec{v})).$$

Let A be the matrix that represents f and B the matrix that represents h (under the standard bases). Then

$$h \circ f(\vec{v}) = B(A\vec{v}).$$

Here we introduce a new operation on matrices: matrix multiplication. (This is in fact a generalization of matrix with vector multiplication.) If

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{p1} & \dots & a_{pn} \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} b_{11} & \dots & b_{1p} \\ \vdots & & \vdots \\ b_{m1} & \dots & n_{mp} \end{pmatrix},$$

then the ijth entry of the product BA is

$$(BA)_{ij} = \begin{pmatrix} b_{i1} \\ \vdots \\ b_{ip} \end{pmatrix} \cdot \begin{pmatrix} a_{1j} \\ \vdots \\ a_{pj} \end{pmatrix} = b_{i1}a_{1j} + \cdots + b_{ip}a_{pj}.$$

Example. For example,

$$\begin{pmatrix} 1 & -1 \\ 0 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & -3 \\ -1 & -3 \\ 5 & 3 \end{pmatrix}.$$

Example. Matrix multiplication is associative. For example, let $f : \mathbb{R}^3 \to \mathbb{R}^2$ and $h : \mathbb{R}^2 \to \mathbb{R}^3$ be represented by matrices $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{pmatrix}$ respectively. Then we have that

$$h \circ f(\vec{v}) = B(A\vec{v}) = (BA)\vec{v}.$$

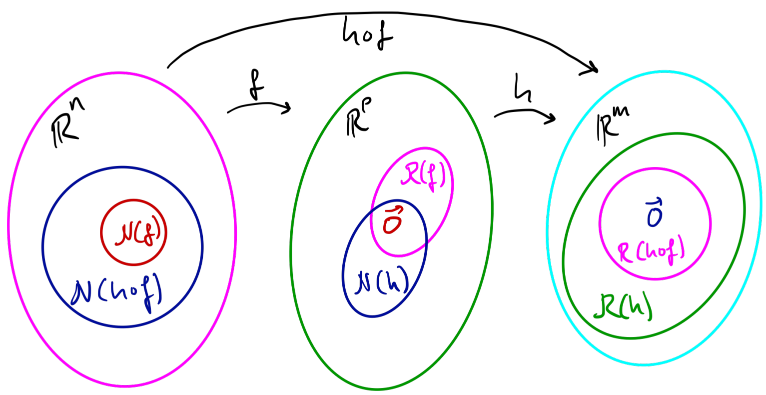
For example, let $\vec{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. Then

$$h \circ f(\vec{v}) = B(A\vec{v}) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{pmatrix} \left(\begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 7 \\ 0 \\ 23 \end{pmatrix},$$

and

$$h \circ f(\vec{v}) = (BA)\vec{v} = \left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 3 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 0 \\ 23 \end{pmatrix}.$$

The following picture respresents the relationship between the different sets, ranges and null spaces of the two maps.



Claim. *We have* $\mathcal{R}(h \circ f) \subseteq \mathcal{R}(h)$.

Proof. Suppose that $\vec{z} \in \mathcal{R}(h \circ f)$. Then there is some $\vec{v} \in \mathbb{R}^n$ such that $h \circ f(\vec{v}) = h(f(\vec{v})) = \vec{z}$. Let $\vec{w} = f(\vec{v}) \in \mathbb{R}^p$. Then $h(\vec{w}) = \vec{z}$, and therefore $\vec{z} \in \mathcal{R}(h)$. □

Claim. *We have* $\mathcal{N}(f) \subseteq \mathcal{N}(h \circ f)$.

Proof. Suppose that $\vec{v} \in \mathcal{N}(f)$. Then $f(\vec{v}) = \vec{0}$, and therefore $h \circ f(\vec{v}) = h(f(\vec{v})) = h(\vec{0}) = \vec{0}$. Therefore $\vec{v} \in \mathcal{N}(h \circ f)$. □

Lecture 15 (10/19)

Inverses

Definition. The identity map $\mathbb{R}^n \to \mathbb{R}^n$ is the map such that $\vec{v} \mapsto \vec{v}$ for all $\vec{v} \in \mathbb{R}^n$. It is represented by the **identity matrix**

$$I_n = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}.$$

You are probably familiar with inverses of functions in one real variable. If $f : \mathbb{R} \to \mathbb{R}$ and the inverse f^{-1} exists, then it is such that every time $f(x) = y$ then $f^{-1}(y) = x$. Then $f \circ f^{-1} = f^{-1} \circ f$ is the identity map. Here we introduce a more nuanced version of inverses: left and right inverses. Sec. IV.4 p.254

Definition. For two functions $f : A \to B$ and $h : B \to A$, we say that f is a *left inverse* of h if $f \circ h$ is the identity map and f is a *right inverse* of h if $h \circ f$ is the identity map. If both hold, we say thaat f is a *two-sided* inverse of h.

Example. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}^3$ be two functions such that $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \overset{f}{\mapsto} \begin{pmatrix} x \\ y \end{pmatrix}$ and $\begin{pmatrix} x \\ y \end{pmatrix} \overset{h}{\mapsto} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$.

Then $h \circ f : \mathbb{R}^3 \to \mathbb{R}^3$ is such that $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \overset{h\circ f}{\mapsto} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$, which is not surjective. However, $f \circ h : \mathbb{R}^2 \to \mathbb{R}^2$ is such that $\begin{pmatrix} x \\ y \end{pmatrix} \overset{f\circ h}{\mapsto} \begin{pmatrix} x \\ y \end{pmatrix}$ is the identity map. Therefore, f is a left inverse of h (or h is a right inverse of f).

We can see this in terms of matrix multiplication as well. The function h has matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ and f has matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. The composition $h \circ f$ has matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq I_3.$$

The composition $f \circ h$ has matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

For most of the remainder of this course, we will talk about linear transformations, i.e. maps $\mathbb{R}^n \to \mathbb{R}^n$. For these, there is only one type of possible inverse: a two-sided inverse. For an $n \times n$ matrix A, we say that A is *invertible* if there exists an $n \times n$ matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$. For 2×2 matrices, there is an easy formula for the matrix inverse. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

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

Note that this formula fails when $ad - bc = 0$, this is exactly when A is not invertible. We will see this again when we talk about determinants of matrices.

Lecture 16 (10/28)

Correlation coefficient

In the topics that follow, we will make heavy use of vector dot products. We have already seen dot products in the context of matrix and vector multiplication, but we haven't looked at any geometric or other kind of interpretation yet. To remind ourselves, the dot product of two vectors $\vec{v}$ andd $\vec{w}$ (of the same dimension) is defined as p. 43

$$\vec{v} \cdot \vec{w} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \cdots + v_n w_n.$$

We can now also define the *length* of a vector $\vec{v}$ in terms of dot products. This is the usual Euclidean length that you are already familiar with. We have

$$|\vec{v}| = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{\vec{v} \cdot \vec{v}}.$$

Before we move on, we take a small detour and look at an application (and alternative interpretation) of the dot product: data correlation. We need one more fact about dot products, which we will not prove here. You can read more about the geometry in Section One.II in the book if you wish. We have

$$|\vec{v} \cdot \vec{w}| = |\vec{v}||\vec{w}| \cos\theta,$$

where θ is the angle between $\vec{v}$ and $\vec{w}$.

Example. Consider 4 mathematicians: Andrea, Braxton, Charlie and Dima. Over one week, we record their coffee consumption and their productivity. We would like to find out if there is a correlation between these two variables. Here is the result:

Mathematician	Cups of coffee	Lemmas proved
A	7	8
B	14	10
C	11	8
D	4	6

We can think of the coffee data and productivity data as two 4-dimensional vectors, where the mathematicians are the dimensions. In that case, it makes sense to think of the origin 0 as the mean, and then write the data as a deviation from the mean. Correlation between two variables described the way they vary with each other, and should not be affected by their averages. For example, it should not matter whether we measure the coffee in cups or in ounces or in droplets. So, we consider the normalized data:

Mathematician	Cups of coffee	Lemmas proved
A	-2	0
B	5	2
C	2	0
D	-5	-2

If two variables are correlated, we expect their signs to match for each dimension, and the deviation from the respective means to be proportional. Geometrically, this means that a positive correlation looks like a small angle between the two vectors, whereas a negative correlation looks like a large angle. An ideal positive correlation should have $\theta = 0$, and ideal negative correlation $\theta = \pi$ and an ideal uncorrelation should have $\theta = \pi/2$. Naturally, we let the **correlation coefficient** r be

$$r = \cos\theta = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}\| \, \|\vec{y}\|}.$$

In our example, we get $\vec{x} = \begin{pmatrix} -2 \\ 5 \\ 2 \\ -5 \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ -2 \end{pmatrix}$, and

$$r = \frac{-2\cdot 0 + 5\cdot 2 + 2\cdot 0 + -5\cdot -2}{\sqrt{(-2)^2+5^2+2^2+(-5)^2}\cdot\sqrt{0^2+2^2+0^2+(-2)^2}} \simeq .93$$

Very high correlation!

Orthogonal Complements

We say that two vectors $\vec{v}$ and $\vec{w}$ are *orthogonal* if $\vec{v}\cdot\vec{w} = 0$. Now, if V is a subspace of $\mathbb{R}^n$, then we say that a vector $\vec{w}$ is **orthogonal to the subspace** V if $\vec{w}$ is orthogonal to **all** vectors in V. So, Three.VI.3 p. 285

$$\vec{w} \text{ is orthogonal to } V \text{ if and only if } \vec{w}\cdot\vec{v} = 0 \ \ \forall \vec{v}\in V.$$

The set of all such vectors is call the **orthogonal complement** of V, denoted $V^\perp$. So,

$$V^\perp = \{\vec{w}\in\mathbb{R}^n | \vec{w}\cdot\vec{v} = 0 \ \ \forall\vec{v}\in V\}.$$

The orthogonal complement has the following properties:

Lemma. *1. $V^\perp$ is a subspace of $\mathbb{R}^n$.*

2. V and $V^\perp$ have only $\vec{0}$ in common.

3. $\dim(V) + \dim(V^\perp) = n$.

4. $(V^{\perp})^{\perp} = V$.

Proof. 1. The set $V^{\perp}$ is exactly the kernel of the projection onto V, because the projection maps exactly those vectors to 0 that are orthogonal to V. We already know that kernels are subspaces.

2. If $\vec{x}$ is in both V and $V^{\perp}$, then $\vec{x} \cdot \vec{x} = 0 = \|\vec{x}\|^2$. So, $\|\vec{x}\| = 0$ and $\vec{x} = \vec{0}$.

3. In the projection onto V, we have that V is the image and $V^{\perp}$ is the kernel, so this is in fact the rank-nullity theorem! The way we will prove this, is by showing that we can find a basis for V and one for $V^{\perp}$ the union of which form a basis for $\mathbb{R}^n$.

4. If $\vec{v} \in V$, then $\vec{v} \cdot \vec{w} = 0$ for all $\vec{w} \in V^{\perp}$. The set $(V^{\perp})^{\perp}$ contains all such vectors, so $\vec{v} \in (V^{\perp})^{\perp}$, and because this was true for all $\vec{v} \in V$ we have that $V \subseteq (V^{\perp})^{\perp}$.
Also, $\dim(V)+\dim(V^{\perp}) = n$ and $\dim(V^{\perp})+\dim((V^{\perp})^{\perp}) = n$. Therefore $\dim(V) = \dim((V^{\perp})^{\perp})$. So, we must have that $(V^{\perp})^{\perp} = V$.

□

The transpose of a matrix

If A is an $n \times m$ matrix, then the **transpose** A^T of A is the $m \times n$ matrix whose entries ij are the entries ji of A. You can think about this as switching the roles of the rows and columns, or as grabbing the matrix by the upper-left and lower-right corner, and flipping it over.

The transpose is a very handy concept for orthogonal matrices, as you'll see in the next section. The following lemma gives some properties of the transpose. You don't need to remember them, just know where to look if you need any of them.

Lemma. *For any matrix A and its transpose A^T*

1. $(A + B)^T = A^T + B^T$
2. $(kA)^T = kA^T$
3. $(AB)^T = B^T A^T$
4. $\operatorname{rank}(A) = \operatorname{rank}(A^T)$
5. $(A^T)^{-1} = (A^{-1})^T$.

Transposes also give us a way to think of the dot product as an ordinary matrix multiplication. For two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$,

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w}.$$

Orthonormal bases

A set of vectors is called **orthonormal** if they are all orthogonal to one another and have length 1. The book mostly focuses on the orthogonal part in this Section, but orthonormal is much more useful, so we might as well focus on that straight away. Three.VI.2 p. 280
If $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_p$ are orthonormal then $\hat{u}_i \cdot \hat{u}_j = 0$ if $i \neq j$ and $\hat{u}_i \cdot \hat{u}_j = 1$ if $i = j$.

Clearly, an orthonormal set is a linearly independent set. An orthonormal set of cardinality p in an p-dimensional subspace V is therefore a basis of V. We call that an **orthonormal basis**.

If we have an orthonormal basis $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_p$ for U, then we can write the projection of $\vec{v}$ onto U as a set of dot products.

Since $\text{proj}_U \vec{v} \in U$, we know that $\text{proj}_U \vec{v} = k_1\hat{u}_1 + \ldots + k_p\hat{u}_p$. We also know that $\vec{v} - \text{proj}_U \vec{w} \in U^\perp$. Together, this means that

$$\begin{aligned}(\vec{v} - k_1\hat{u}_1 - \ldots - k_p\hat{u}_p) \cdot \hat{u}_1 &= 0\\ (\vec{v} - k_1\hat{u}_1 - \ldots - k_p\hat{u}_p) \cdot \hat{u}_2 &= 0\\ &\vdots\\ (\vec{v} - k_1\hat{u}_1 - \ldots - k_p\hat{u}_p) \cdot \hat{u}_p &= 0.\end{aligned}$$

Since, $\hat{u}_i \cdot \hat{u}_j = 0$ for all $i \neq j$, this gives us

$$\begin{matrix}(\vec{v} - k_1\hat{u}_1) \cdot \hat{u}_1 = 0\\ (\vec{v} - k_1\hat{u}_2) \cdot \hat{u}_2 = 0\\ \vdots\\ (\vec{v} - k_p\hat{u}_p) \cdot \hat{u}_p = 0.\end{matrix} \qquad \text{which leads to} \qquad \begin{matrix}k_1 = \vec{v} \cdot \hat{u}_1\\ k_2 = \vec{v} \cdot \hat{u}_2\\ \vdots\\ k_p = \vec{v} \cdot \hat{u}_p.\end{matrix}$$

We now conclude that

$$\text{proj}_U \vec{v} = (\hat{u}_1 \cdot \vec{v})\hat{u}_1 + (\hat{u}_2 \cdot \vec{v})\hat{u}_2 + \ldots + (\hat{u}_p \cdot \vec{v})\hat{u}_p.$$

If we set $U = \mathbb{R}^n$, then we have

$$\vec{v} = (\hat{u}_1 \cdot \vec{v})\hat{u}_1 + (\hat{u}_2 \cdot \vec{v})\hat{u}_2 + \ldots + (\hat{u}_n \cdot \vec{v})\hat{u}_n.$$

This is a fast way to find the unique expression of $\vec{v}$ as a linear combination of the basis vectors.

Orthogonal Projection Matrices

We can use transposes to rewrite dot products as matrix products. This helps us find the matrix of a projection, based on the expression in terms of dot products that we found in the previous section.

$$\begin{aligned}\text{proj}_U \vec{v} &= (\hat{u}_1 \cdot \vec{v})\hat{u}_1 + (\hat{u}_2 \cdot \vec{v})\hat{u}_2 + \ldots + (\hat{u}_p \cdot \vec{v})\hat{u}_p\\ &= \hat{u}_1\hat{u}_1^T\vec{v} + \hat{u}_2\hat{u}_2^T\vec{v} + \ldots + \hat{u}_p\hat{u}_p^T\vec{v}\\ &= (\hat{u}_1\hat{u}_1^T + \hat{u}_2\hat{u}_2^T + \ldots + \hat{u}_p\hat{u}_p^T)\vec{v}\\ &= \begin{pmatrix}\uparrow & \uparrow & & \uparrow\\ \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_p\\ \downarrow & \downarrow & & \downarrow\end{pmatrix}\begin{pmatrix}\leftarrow & \hat{u}_1^T & \rightarrow\\ \leftarrow & \hat{u}_2^T & \rightarrow\\ & \vdots & \\ \leftarrow & \hat{u}_p^T & \rightarrow\end{pmatrix}\vec{v}\\ &= QQ^T\vec{v}.\end{aligned}$$

We have found a matrix for the orthogonal projection in terms of the orthonormal basis! In general, if $V \subseteq \mathbb{R}^n$ has orthonormal basis $\hat{u}_1, \ldots, \hat{u}_m$, then the matrix of the orthogonal projection onto V is

$$A = QQ^T,$$

where

$$Q = \begin{pmatrix}\uparrow & \uparrow & & \uparrow\\ \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_p\\ \downarrow & \downarrow & & \downarrow\end{pmatrix}.$$

Lecture 17 (11/2)

Gram-Schmidt Algorithm

Previously, we found a formula for projections that does not involve a matrix multiplication:

$$\operatorname{proj}_U \vec{v} = (\hat{u}_1 \cdot \vec{v})\hat{u}_1 + (\hat{u}_2 \cdot \vec{v})\hat{u}_2 + \ldots + (\hat{u}_p \cdot \vec{v})\hat{u}_p,$$

where $\hat{u}_1, \ldots, \hat{u}_p$ is an orthonormal basis for U. We often have a basis for U that is not orthonormal, so it will be useful to be able to turn such a basis into an orthonormal one. We can indeed do this, using the **Gram-Schmidt algorithm**.

Suppose we have a basis for a subspace $V \subseteq \mathbb{R}^n$ that is not orthonormal: $\langle \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \rangle$. We can start by turning $\vec{v}_1$ into a unit vector, and let that be our first orthonormal basis vector:

$$\hat{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}.$$

Now, let's fix the second vector. To turn $\vec{v}_2$ into the next valid orthonormal basis vector, we have to make sure that both the constraints of orthogonality and unit length are satisfied. We can make sure that the vector is orthogonal to $\hat{u}_1$ by throwing away its component that is parallel to $\hat{u}_1$:

$$\vec{v}_2^{\perp} = \vec{v}_2 - \vec{v}_2^{\parallel} = \vec{v}_2 - \operatorname{proj}_W \vec{v}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \hat{u}_1)\hat{u}_1,$$

where $W = [\hat{u}_1]$. Then we make it unit length:

$$\hat{u}_2 = \frac{\vec{v}_2^{\perp}}{\|\vec{v}_2^{\perp}\|}.$$

We keep going like this:

$$\vec{v}_3^{\perp} = \vec{v}_3 - \vec{v}_3^{\parallel} = \vec{v}_3 - \operatorname{proj}_W \vec{v}_3 = \vec{v}_3 - (\vec{v}_3 \cdot \hat{u}_1)\hat{u}_1 - (\vec{v}_3 \cdot \hat{u}_2)\hat{u}_2,$$

where $W = [\hat{u}_1, \hat{u}_2]$. Then we set

$$\hat{u}_3 = \frac{\vec{v}_3^{\perp}}{\|\vec{v}_3^{\perp}\|}.$$

Etc:

$$\vec{v}_4^{\perp} = \vec{v}_4 - \vec{v}_4^{\parallel} = \vec{v}_4 - \operatorname{proj}_W \vec{v}_4 = \vec{v}_4 - (\vec{v}_4 \cdot \hat{u}_1)\hat{u}_1 - (\vec{v}_4 \cdot \hat{u}_2)\hat{u}_2 - (\vec{v}_4 \cdot \hat{u}_3)\hat{u}_3,$$

where $W = [\hat{u}_1, \hat{u}_2, \hat{u}_3]$. Then we set

$$\hat{u}_4 = \frac{\vec{v}_4^{\perp}}{\|\vec{v}_4^{\perp}\|}.$$

When we are done, we have an orthonormal basis $\hat{u}_1, \ldots, \hat{u}_m$ for V, where

$$\begin{aligned} \hat{u}_1 &= \frac{\vec{v}_1}{\|\vec{v}_1\|}, \\ \hat{u}_2 &= \frac{\vec{v}_2^{\perp}}{\|\vec{v}_2^{\perp}\|}, \\ \hat{u}_3 &= \frac{\vec{v}_3^{\perp}}{\|\vec{v}_3^{\perp}\|}, \\ &\vdots \\ \hat{u}_p &= \frac{\vec{v}_p^{\perp}}{\|\vec{v}_p^{\perp}\|}, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \vec{v}_2^{\perp} &= \vec{v}_2 - \operatorname{proj}_{[\hat{u}_1]} \vec{v}_2, \\ \vec{v}_3^{\perp} &= \vec{v}_3 - \operatorname{proj}_{[\hat{u}_1, \hat{u}_2]} \vec{v}_3, \\ &\vdots \\ \vec{v}_p^{\perp} &= \vec{v}_p - \operatorname{proj}_{[\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_{p-1}]} \vec{v}_p. \end{aligned}$$

You should be able to check for yourself that this is indeed an orthonormal basis of V.
As an application of projections, I'd like to show you approximate solutions to systems of linear equations.

Example. The projection of a vector $\vec{x}$ onto a subspcae V can be thought of as the vector in V that is the closest possible to $\vec{x}$, out of all the vectors in V. You can see this easily in 2D or 3D using Pythagoras, and it looks similar in more dimensions. So, we have

$$\text{proj}_V\,\vec{x} = \{\vec{v} \in V \text{ such that } \|\vec{x} - \vec{v}\| \leq \|\vec{x} - \vec{w}\| \;\forall \vec{w} \in V\}.$$

For many real-world problems, such as finding a solution to a system of linear equations, a perfect solution may not exist. Instead we look for a solution that is the best possible within our restrictions. We have seen examples of systems of linear equations that are inconsistent. For example, take the system

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We can see that

$$\text{im}(A) = [(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix})]$$

and that

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin \text{im}(A)$$

. In this case, any solution $\vec{x}$ to $A\vec{x} = \vec{b}$ would be wrong, but can we find the solution for $\vec{x}$ that is the "least wrong", or the closest possible to a solution?

The vector on the line spanned by $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ that is the closest to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is exactly the projection of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ onto the line $\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right]$. In other words, we are looking for

$$\text{proj}_{\text{im}(A)}\,\vec{b} = A\vec{x}^*.$$

We call this $\vec{x}^*$ the **least-squares solution** of the system $A\vec{x} = \vec{b}$. This is the solution hat minimizes the length of the vector $\vec{b} - A\vec{x}^*$ and therefore the sum of the squares of its entries. The length of $\vec{b} - A\vec{x}^*$ is called the **error**. If the system $A\vec{x} = \vec{b}$ is consistent, then we can achieve a zero error solution.

First we need two results about the image and kernel of A and its transpose.

Lemma. $(\text{im}(A))^\perp = \ker(A^T)$

Proof. Let $A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$. Then the image of A is the span of the set $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$. The space $(\text{im}(A))^\perp$ is exactly the set of all vectors $\vec{x} \in \mathbb{R}^n$ such that $\vec{x} \cdot \vec{v}_i = 0$, for $1 \leq i \leq m$. We can also write this as

$$\begin{pmatrix} \leftarrow & \vec{v}_1^T & \rightarrow \\ \leftarrow & \vec{v}_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & \vec{v}_m^T & \rightarrow \end{pmatrix} \vec{x} = \vec{0}.$$

Therefore $(\text{im}(A))^\perp = \ker(A^T)$. ☐

Lemma. $\ker(A) = \ker(A^T A)$

Proof. Any $A\vec{x} \neq \vec{0}$ in the image of A is not in $\ker(A^T)$ (because it is not orthogonal to all vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$). So, if $\vec{x} \notin \ker(A)$, then $\vec{x} \notin \ker(A^T A)$.
Also, if $\vec{x} \in \ker(A)$, then clearly $\vec{x} \in \ker(A^T A)$. Therefore, $\ker(A) = \ker(A^T A)$. □

If $\vec{x}^*$ is the least-squares solution to $A\vec{x} = \vec{b}$, then

$$\begin{aligned}
&A\vec{x}^* = \operatorname{proj}_{\operatorname{im}(A)} \vec{b}, \\
&\vec{b} - \operatorname{proj}_{\operatorname{im}(A)} \vec{b} \in (\operatorname{im}(A))^\perp, \\
&\Rightarrow\ A^T(\vec{b} - A\vec{x}^*) = \vec{0}, \\
&\Rightarrow\ A^T A\vec{x}^* = A^T\vec{b}.
\end{aligned}$$

The equation $A^T A\vec{x}^* = A^T\vec{b}$ is called the **normal equation** of $A\vec{x} = \vec{b}$.

If $\ker(A) = \vec{0}$, then $\ker(A^T A) = \vec{0}$ and $A^T A$ is invertible, so we can write

$$\vec{x}^* = (A^T A)^{-1} A^T \vec{b}.$$

This also means that

$$\operatorname{proj}_{im(A)} \vec{b} = A\vec{x}^* = A(A^T A)^{-1} A^T \vec{b}.$$

Lecture 18 (11/4)

Inverting a matrix

The **inverse** of a linear transformation $T(\vec{x})$, denoted $T^{-1}(\vec{x})$, is another LT such that $T^{-1}(T(\vec{x})) = \vec{x}$ and $T(T^{-1}(\vec{y})) = \vec{y}, \forall \vec{x} \in X, \vec{y} \in Y$, where X and Y are the domain and target space of T, respectively. Note that Y is the domain and X the target space of T^{-1}.

A LT $T : X \to Y$ is **invertible** if and only if $T(\vec{x}) = \vec{y}$ has a unique solution $\vec{x} \in X$ for each $\vec{y} \in Y$.

A square matrix A is invertible if its linear transformation is invertible.

To find the inverse of A, consider the augmented matrix

$$\left(A \mid I_n\right),$$

where I_n is the **identity matrix**, the square $n \times n$ matrix with 1s on the diagonal and 0s everywhere else:

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

If

$$\operatorname{rref}\left(\left(A \mid I_n\right)\right) = \left(I_n \mid B\right),$$

then $A^{-1} = B$. So,

$$\operatorname{rref}\left(\left(A \mid I_n\right)\right) = \operatorname{rref}\left(\left(I_n \mid A^{-1}\right)\right).$$

Determinants

The *determinant* of an $n \times n$ matrix is an invariant that holds information about the independence, angles and magnitudes of its columns. You should know how to find determinants of 2×2 and 3×3 matrices. The computation gets quite involved for most larger matrices, so I encourage you to find out how to find determinants by computer. p. 326
For a 2×2 matrix A, the determinant, denoted, $|A|$, or $\det A$, is given by

$$\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = ad - bc.$$

For a 3×3 matrix, the determinant is given by

$$\left|\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}\right| = aei + bfg + cdh - hfa - idb - gec.$$

Eigenvectors and eigenvalues

Let A be an $n \times n$ matrix. A nonzero vector $\vec{v}_i \in \mathbb{R}^n$ is called an *eigenvector* of A if $A\vec{v}_i = \lambda_i \vec{v}_i$ for some $\lambda_i \in \mathbb{R}$. Clearly, if $\vec{v}_i$ is an eigenvector of A, then $c\vec{v}_i$ for any nonzero $c \in \mathbb{R}$ is also an eigenvector of A. The value $\lambda_i \in \mathbb{R}$ associated with the vector $\vec{v}_i$ is called the *eigenvalue*. Five.II.3 p. 412

Finding eigenvalues and eigenvectors is easy. You already have the necessary skills! We know that if $\vec{v}$ is an eigenvector of A, then $A\vec{v} = \lambda\vec{v}$. We can rewrite this as

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

This means that λ is an eigenvalue of A whenever the matrix $A - \lambda I_n$ has a nonzero kernel. We know that this is the case when the determinant is 0. This gives us the equation we need to solve to find the eigenvalues:

$$\left|A - \lambda I_n\right| = 0.$$

The expression $\left|A - \lambda I_n\right|$ takes the form of an n-degree polynomial in λ, and we call this the **characteristic polynomial** of A, which is sometimes denoted by $f_A(\lambda)$. We call the equation $f_A(\lambda) = 0$ the **characteristic equation** of A. When A is a 2×2, and in some particular cases a 3×3 matrix, you should be able to solve the characteristic equation by hand. You may need to review some algebra and complex numbers for that. By the fundamental theorem of algebra, we know that roots of the characteristic polynomial exist, but we are not guaranteed that they are real. Some matrices have real eigenvalues/eigenvectors, some have complex ones, and some have both. Let's do an example of a matrix with real eigenvalues/eigenvectors. Let,

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

We have

$$A - \lambda I_2 = \begin{pmatrix} 0 - \lambda & 1 \\ -2 & -3 - \lambda \end{pmatrix},$$

which gives

$$f_A(\lambda) = |A - \lambda I_2| = \begin{vmatrix} 0 - \lambda & 1 \\ -2 & -3 - \lambda \end{vmatrix} = \lambda^2 + 3\lambda + 2 = (1 + \lambda)(2 + \lambda).$$

This polynomial has roots $\lambda_1 = -1$ and $\lambda_2 = -2$. To find the corresponding eigenvectors, we go back to the equation $(A - \lambda_i I_n)\vec{v} = \vec{0}$. All eigenvectors will be in the kernel of $A - \lambda_i I_n$ (although that kernel also includes $\vec{0}$, which is not an eigenvector). We call the kernel of $A - \lambda_i I_n$ the **eigenspace** of λ_i, denoted E_{λ_i}. So,

$$E_{-1} = \ker\left(\begin{pmatrix} 1 & 1 \\ -2 & -2 \end{pmatrix}\right) = \left[\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right], \quad E_{-2} = \ker\left(\begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix}\right) = \left[\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right].$$

Therefore, the two eigenvectors are $\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\vec{v}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.
Let's turn to a matrix with complex eigenvalues: a rotation matrix. Suppose that

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b \neq 0$ (so the rotation angle is not an integer multiple of π). Then following the same mechanism as above we obtain $\lambda_1 = a + bi$ and $\lambda_2 = a - bi$, and

$$E_{a+bi} = \ker\left(\begin{pmatrix} -ib & -b \\ b & -ib \end{pmatrix}\right) = \left[\begin{pmatrix} i \\ 1 \end{pmatrix}\right], \quad E_{a-bi} = \ker\left(\begin{pmatrix} ib & -b \\ b & ib \end{pmatrix}\right) = \left[\begin{pmatrix} -i \\ 1 \end{pmatrix}\right].$$

So, $\vec{v}_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}$ and $\vec{v}_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Algebraic and geometric mulitplicities

If a λ_i is a root of $f_A(\lambda)$, then we say that λ_i has **algebraic multiplicity** k, denoted $\text{almu}(\lambda_i)$ if we can write Def. 3.7 p. 418

$$f_A(\lambda) = (\lambda_i - \lambda)^k g(\lambda),$$

where $g(\lambda)$ is an $n-k$ degree polynomial which does not have λ_i as a root. You may expect that if an eigenvalue has algebraic multiplicity k, that means that there are k linearly independent eigenvectors with that eigenvalue. This is not necessarily true, though. All we know is that there are up to k such eigenvectors. We call the $\dim(E_{\lambda_i})$ the **geometric multiplicity** of λ_i, denoted $\text{gemu}(\lambda_i)$. We have

$$\text{gemu}(\lambda_i) \leq \text{almu}(\lambda_i).$$

(In order to prove that rigorously, we'd need a few determinant tricks that we won't do, but you can Google it if you are curious.
We know that, by the fundamental theorem of algebra, we always have n roots, that need not be real. However, when n is odd, we are guarantueed to have at least 1 real root.

Lemma. *For odd n, every $n \times n$ matrix has at least one real eigenvalue.*

Lecture 19 (11/16)

Change of basis and similar matrices

You are already familiaar with change of basis as a linear transformation. The standard basis is often not the best choice of basis for a given linear transformation, and that is why it's important for you to be able to smoothly transition between different bases for a subspace. An important lesson we will learn from this is that many transformations represent the same transformation with respect to different bases. For example, a rotation of a plane will look different when you look at it from different angles, but it keeps its essence. Of course we need to formalize this idea of "essence". Five.II p. 402
If $\mathfrak{B} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ is a basis for a vector space V, then (as you know), we can write every $\vec{x} \in V$ uniquely as a linear combination

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n.$$

As long as we know what the basis is, we can then express $\vec{x}$ purely in terms of these coordinates $c_1, c_2, \ldots, c_n$. You may not realize this, but that is exactly what we do when we write $\vec{x}$ the way we do "normally", *i.e.* when working with respect to the standard basis, we simply just forget to even mention what basis the coordinates refer to. We let

$$[\vec{x}]_{\mathfrak{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

be the coordinate vector of $\vec{x}$ with respect to $\mathfrak{B}$. We will use a matrix S that has the basis $\mathfrak{B}$ as its column vectors:

$$S = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

so that we can write:

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n = S[\vec{x}]_{\mathfrak{B}}.$$

If S is a square matrix, then S is invertible (why?), and we have

$$[\vec{x}]_{\mathfrak{B}} = S^{-1}\vec{x}.$$

Let f be a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$, and let $\mathfrak{B} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ be a basis for $\mathbb{R}^n$. Then there should (why?) be a matrix B that directly transforms $[\vec{x}]_{\mathfrak{B}}$ into $[f(\vec{x})]_{\mathfrak{B}}$. So,

$$B[\vec{x}]_{\mathfrak{B}} = [f(\vec{x})]_{\mathfrak{B}},\ \forall \vec{x} \in \mathbb{R}^n.$$

This matrix B is called the $\mathfrak{B}$-matrix of f.

How do we construct B from f and $\mathfrak{B}$? We have

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n,$$

So,

$$\begin{aligned}
[f(\vec{x})]_{\mathfrak{B}} &= [f(c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n)]_{\mathfrak{B}} \\
&= [f(c_1\vec{v}_1) + f(c_2\vec{v}_2) + \ldots + f(c_n\vec{v}_n)]_{\mathfrak{B}} \\
&= [c_1 f(\vec{v}_1) + c_2 f(\vec{v}_2) + \ldots + c_n f(\vec{v}_n)]_{\mathfrak{B}} \\
&= [c_1 f(\vec{v}_1)]_{\mathfrak{B}} + [c_2 f(\vec{v}_2)]_{\mathfrak{B}} + \ldots + [c_n f(\vec{v}_n)]_{\mathfrak{B}} \\
&= c_1[f(\vec{v}_1)]_{\mathfrak{B}} + c_2[f(\vec{v}_2)]_{\mathfrak{B}} + \ldots + c_n[f(\vec{v}_n)]_{\mathfrak{B}} \\
&= \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ [f(\vec{v}_1)]_{\mathfrak{B}} & [f(\vec{v}_2)]_{\mathfrak{B}} & \cdots & [f(\vec{v}_n)]_{\mathfrak{B}} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \\
&= \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ [T(\vec{v}_1)]_{\mathfrak{B}} & [T(\vec{v}_2)]_{\mathfrak{B}} & \cdots & [f(\vec{v}_n)]_{\mathfrak{B}} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} [\vec{x}]_{\mathfrak{B}} \\
&= B[\vec{x}]_{\mathfrak{B}}
\end{aligned}$$

If we have $\vec{y} = A\vec{x}$, and $[\vec{y}]_{\mathfrak{B}} = B[\vec{x}]_{\mathfrak{B}}$, then

$$\vec{y} = A\vec{x} = AS[\vec{x}]_{\mathfrak{B}}$$

and

$$\vec{y} = S[\vec{y}]_{\mathfrak{B}} = SB[\vec{x}]_{\mathfrak{B}}.$$

Therefore,

$$AS = SB$$

$$B = S^{-1}AS, \text{ or } A = SBS^{-1}.$$

We can use the following picture to understand this translation:

$$\begin{array}{ccc} \vec{x} & \xrightarrow{A} & f(\vec{x}) \\ S \uparrow \downarrow S^{-1} & & S \uparrow \downarrow S^{-1} \\ [\vec{x}]_{\mathfrak{B}} & \xrightarrow{B} & [f(\vec{x})]_{\mathfrak{B}} \end{array}$$

If there exists an invertible matrix S such that matrices A and B are related by $AS = SB$, or $A = SBS^{-1}$, then we call A and B *similar*. This means that they represent the same linear transformation, just with respect to a different basis. We write $A \sim B$. Similarity is an **equivalence relations**.

Theorem. *Similar matrices, $A \sim B$, have the following features ~~not~~ in common:*

- *characteristic polynomial* $\left\{ \begin{array}{l} \textit{eigenvalues,} \\ \textit{algebraic multiplicities,} \\ \textit{trace,} \\ \textit{determinant,} \\ \textit{invertibility,} \end{array} \right.$
- *geometric multiplicities* $\left\{ \begin{array}{l} \textit{rank*,} \\ \textit{nullity*,} \\ \textit{diagonalizability,} \end{array} \right.$
- ~~*eigenvectors*~~,
- ~~*kernel*~~,
- ~~*image*~~.

**assuming we already have same characteristic polynomial.*

Diagonal matrices

Recall that the ith column of the $\mathfrak{B}$-matrix corresponds to the vector $[f(\vec{v}_i)]_{\mathfrak{B}}$. If the matrix B only has nonzero entries on its main diagonal, this transformation becomes very easy to understand. We call such a matrix a **diagonal** matrix. For example: Five.II p. 408

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

tells us that

$$f(\vec{v}_1) = \vec{v}_1,\ f(\vec{v}_2) = 2\vec{v}_2,\ f(\vec{v}_3) = 3\vec{v}_3.$$

In other words, B is diagonal if for every vector in the basis $\mathfrak{B}$, we have $T(\vec{v}_i) = c_i\vec{v}_i$. If a matrix A is similar to some diagonal matrix B, then we say that A is **diagonalizable**. It is not hard to see that if the the matrix B is diagonal, we have that $[f(\vec{v}_i)]_{\mathfrak{B}} = \lambda_i[\vec{v}_i]_{\mathfrak{B}}$, and therefore $f(\vec{v}_i) = A\vec{v}_i = \lambda_i\vec{v}_i$. In other words, if A is diagonalizable, A has an eigenbasis! Conversely, if we already know that A has an eigenbasis, it is very easy to diagonalize A.

Theorem. *A square matrix A is diagonalizable if and only if it has an eigenbasis. If so, we can diagonalize A, with eigenbasis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, where $A\vec{v}_i = \lambda_i\vec{v}_i$, by writing*

$$A = SBS^{-1} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}^{-1}.$$