First Year Linear Algebra

Tom Szwarcer

thomas.szwarcer@oriel.ox.ac.uk[1]

[1] Please email me any mistakes you see, or with feedback/comments

Acknowledgements

This text is heavily influenced by Andre Lukas' *Vectors and Matrices*, which can be found at `http://www-thphys.physics.ox.ac.uk/people/AndreLukas/V&M/V&Mlecturenotes.pdf` at the time of writing. Some proofs are lifted from this set of notes, and others come from elsewhere, including the depths of my own mind. The structure is also based on his structure with some modifications for clarity.

About This Text

This contains almost everything you need to know for first year linear algebra. The things not included are either because I didn't realise they should have been included (please email me and I'll put them in), things that are heavily covered in A-Level (e.g. intersection of two lines in $\mathbb{R}^3$) or things that I had some other good reason not to include - in most cases I will have pointed you to an alternative resource so you can learn about them there. Do note that this text uses inconsistent notation. Perhaps I will fix that at some point. For the time being think of it as intended to keep you on your toes.

Contents

Chapter 1

Vectors and Vector Spaces

1.1 Vector Spaces

1.1.1 Prerequisites: Fields and Binary Operations

A few ideas it would be helpful to introduce first:

Binary Operations

Say we have a set, S. A binary operation $f : S \times S \to S$ is a mapping such that we take two elements of S (the $S \times S$), operate on them (represented by the $\times$ symbol) and get as a result another element within S. This is a property called **closedness**: when an operation on objects within a set produces another still within the same set.

Fields

A field, say F, is a set of objects equipped with some binary operations that act on its elements. These operations satisfy the field axioms - more on that later. The operations are:

1. **Addition.** Denoted $a + b$ Adding two elements from F results in a third - also part of F.
2. **Multiplication.** Denoted $a \cdot b$. Like addition, the set should be closed under this operation.

Field Axioms

The field axioms are a set of rules that the two operations must obey. They are as follows:

1. **Associativity:** $(a + b) + c = a + (b + c)$ and the same for multiplication.
2. **Commutativity:** $a + b = b + a$
3. **Distributivity:** $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
4. **Presence of identities:** $a + 0 = a$ and $a \cdot 1 = a$. So 0 and 1 are the additive and multiplicative identities respectively.
5. **Presence of inverses:** $a + (-a) = 0$ and $a \cdot a^{-1} = 1$. Note that the latter doesn't apply to 0. Note further it is possible to use the idea of multiplicative inverses to define the division operation, but this is not actually needed to define a field.

1.1.2 Definitions

A vector space (Let's call it V) is a mathematical environment that is equipped with the following:

1. **Vectors:** specifically, a set of them. These are mathematical objects that 'live' within the space and are manipulated within it. Note this is about as general as a description can be; there is no mention of components, nor the magnitude/direction that we might have come to expect when we think of vectors. They are simply abstract objects. We tend to say vectors

within the set are within the space itself, i.e. we don't give the set and the space different names.[1] Simply put, vectors are the elements of the sets that form vector spaces. It is helpful to think of such mathematical 'spaces' as sets with extra structure, in this case operations and rules.

2. **Addition Operation:** this means there is some idea of adding two vectors within the space, say $\mathbf{a} \in V$ and $\mathbf{b} \in V$ to produce $\mathbf{a} + \mathbf{b} = \mathbf{c} \in V$. Crucially $\mathbf{c}$ is also a part of the space: closedness.
3. **Field of Scalars**: see above for more information.
4. **Scalar Multiplication Operation:** there is an idea of combining an element of F with an element of V to produce an element still within F (closedness, once again).
5. **Rules:** similarly to the field axioms, the two operations must satisfy a set of rules:
   (a) **Associativity:** $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$
   (b) **Commutativity:** $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
   (c) **Presence of additive identity:** $\mathbf{a} + \mathbf{0} = \mathbf{a}$
   (d) **Presence of additive inverse:** $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$
   (e) **Presence of scalar multiplication identity:** $1\mathbf{a} = \mathbf{a}$
   (f) **Compatibility of scalar multiplication with field multiplication:** $p(q\mathbf{a}) = (pq)\mathbf{a}$
   (g) **Distributivity:** $p(\mathbf{a} + \mathbf{b}) = p\mathbf{a} + p\mathbf{b}$ and $(p + q)\mathbf{a} = p\mathbf{a} + q\mathbf{a}$

Again, note we can define the subtraction operation from the above, but all we need to define vector spaces is addition and scalar multiplication.

1.2 Remarks

Claim 1. $-(-\boldsymbol{v}) = \boldsymbol{v}$

Proof.

$$\begin{aligned} -(-\mathbf{v}) &= -(-\mathbf{v}) + \mathbf{0} \\ &= -(-\mathbf{v}) + [\mathbf{v} + (-\mathbf{v})] \\ &= \mathbf{v} + [(-\mathbf{v}) + (-(-\mathbf{v}))] \\ &= \mathbf{v} + \mathbf{0} \\ &= \mathbf{v} \end{aligned}$$

□

Claim 2. $0\boldsymbol{v} = \boldsymbol{0}$

Proof.

$$\begin{aligned} 0\mathbf{v} &= (a - a)\mathbf{v} \\ &= (a\mathbf{v}) + (-(a\mathbf{v})) \\ &= \mathbf{0} \end{aligned}$$

□

Claim 3. $(-1)\boldsymbol{v} = -\boldsymbol{v}$

Proof.

$$\begin{aligned} \mathbf{0} &= (1 + (-1))\mathbf{v} \\ \mathbf{0} &= \mathbf{v} + (-1)\mathbf{v} \\ -\mathbf{v} &= (-1)\mathbf{v} \end{aligned}$$

□

[1] This will become clearer when the concept of span is introduced. In brief, the space is 'constructed' by adding and scaling vectors in the set so the set forms the 'meat' of the vector space.

Claim 4. $a\boldsymbol{0} = \boldsymbol{0}$

Proof.

$$\begin{aligned} a\mathbf{0} &= a(0\mathbf{v}) \\ &= (a \cdot 0)\mathbf{v} \\ &= 0\mathbf{v} \\ &= \mathbf{0} \end{aligned}$$

□

Claim 5. $a\boldsymbol{v} = \boldsymbol{0} \implies \boldsymbol{v} = 0 \mid a \neq 0$

Proof.

Consider vector $\mathbf{w}$ s.t. $\mathbf{w} \neq \mathbf{0}$. Then we add $a\mathbf{w}$ to both sides:

$$\begin{aligned} a\mathbf{v} + a\mathbf{w} &= a\mathbf{w} \\ a(\mathbf{v} + \mathbf{w}) &= a\mathbf{w} \\ \mathbf{v} + \mathbf{w} &= \mathbf{w} \\ \mathbf{v} &= \mathbf{0} \end{aligned}$$

□

Claim 6. *All vector spaces contain* $\boldsymbol{0}$

Proof. All vector spaces V are equipped with a field, say F. All fields contain 0 as they all have an additive identity. Hence all vector spaces contain the vector $0\mathbf{v}$ for some $\mathbf{v} \in V$ which we showed equalled $\mathbf{0}$. □

1.3 Sub Vector Spaces

1.3.1 Definitions

Consider a vector space V over a field F. Now consider a subset of that space[2], $W \subseteq V$. If W behaves as a vector space on its own when taken in combination with the aforementioned operations and the same field as V, we say it's a subspace of V. In other words, if W is a subset of V, is closed under addition/scalar multiplication, and is over the same field as V then W is a subspace of V.

1.3.2 Checking a Subspace

The key idea here is that if a set behaves as its own vector space and is a subset of another vector space then it's a subspace. Assuming the set in question (say W) is a subset of V, we need to think about how to check if W behaves like a vector space. This means we need to check closedness under the operations of V, namely addition and scalar multiplication. In theory, one should also check all rules that these operations must follow, but in practice checking for the presence of a zero vector suffices. In summary, to check if W is a subspace of V, check for the following:

1. Presence of $\mathbf{0}$
2. Closedness under vector addition
3. Closedness under scalar multiplication.

[2]To highlight the ambiguity here once more: 'space' is used as shorthand for the set of vectors that forms the space. Again, see the section on span for further clarification

1.4 Linear Combinations

1.4.1 Definitions

A linear combination of vectors is one which involves only addition and scalar multiplication. In other words a general linear combination of vectors $\mathbf{v}_i \in V$ would be $\sum_i \alpha_i \mathbf{v}_i$ for some $\alpha_i \in F$.

1.4.2 Linear Independence

Consider the following statement:

$$\sum_i \alpha_i \mathbf{v}_i = \mathbf{0} \implies \alpha_i = 0 \text{ for all i.} \tag{1.1}$$

This says no matter how you try, there doesn't exist a way to combine vectors $\mathbf{v}_i$ linearly such that the end result $= \mathbf{0}$; unless, of course, you scale them all down to zero and then add them up. If the above statement holds, we describe the set of vectors as being linearly independent.

Claim 7. *A set S is linearly dependent $\iff$ $\boldsymbol{v}_j \in S$ can be written as a linear combination of the other elements.*

Proof.

($\Rightarrow$)

$\sum_i \alpha_i \mathbf{v}_i = \mathbf{0}$ has a solution that has at least one $\alpha_i \neq 0$, say for example α_j.

$$\begin{aligned} \sum_i \alpha_i \mathbf{v}_i &= \mathbf{0} \\ \alpha_j \mathbf{v}_j + \sum_{i \neq j} \alpha_i \mathbf{v}_i &= \mathbf{0} \\ \mathbf{v}_j &= \sum_{i \neq j} -\frac{\alpha_i}{\alpha_j} \mathbf{v}_i \end{aligned}$$

□

Proof.

($\Leftarrow$)

$$\mathbf{v}_j = \sum_{i \neq j} \beta_i \mathbf{v}_i$$

Now check linear independence:

$$\begin{aligned} \mathbf{0} &= \sum_i \alpha_i \mathbf{v}_i \\ \mathbf{0} &= \alpha_j \mathbf{v}_j + \sum_{i \neq j} \alpha_i \mathbf{v}_i \\ \mathbf{0} &= \alpha_j \sum_{i \neq j} \beta_i \mathbf{v}_i + \sum_{i \neq j} \alpha_i \mathbf{v}_i \end{aligned}$$

This has a solution when $\beta_i = \alpha_i$ and $\alpha_j = -1$. Hence the statement doesn't imply $\alpha_i = 0$ for all α_i so the set is not linearly independent.

□

1.4.3 Span

This is where the word 'space' in 'vector space' comes in. Span is defined as follows:

$$Span(S) := \left\{ \sum_i \alpha_i \mathbf{v}_i \mid \forall \alpha_i \in F \right\} \tag{1.2}$$

if S is the set containing all $\mathbf{v}_i$. In other words, the set of all possible linear combinations of the set's elements. It is useful to think of this as some sort of 'space'. Lets consider vectors in $\mathbb{R}^3$ to form a geometrical interpretation of this. Start by considering the set of a single vector $\mathbf{v}$. In this case, no matter how you add and scale this vector, you'll always end up somewhere along the same line, a line that's in the vector's direction. Similarly, take the case you start with two vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$. If they're linearly independent, then the only 'space' you can build with them is a plane. The first one can be used to build an infinite line; if the second is linearly independent, then it will have some component that points off the line. This, through scaling, can be used to build any vector within a plane. With three linearly independent vectors you can build a 3 dimensional space.

Claim 8. *Span(V) forms a sub vector space of V*

Proof.

The span contains $\mathbf{0}$ as the zero vector can be obtained from a linear combination of any vector, namely by multiplying it by the scalar 0. This scalar is guaranteed by the field axioms to be within F.

A vector in the span is a linear combination of the original vectors in V. Doing a linear combination of vectors in the span, therefore, will just be another linear combination of the original vectors, hence there is closedness under the relevant operations. Hence Span(V) is a subspace of V.

□

1.4.4 Minimal Sets

Suppose we have a set $S = \{\mathbf{u}, \mathbf{v}, \mathbf{u}+\mathbf{v}\}$. Clearly this set isn't linearly independent. Its span is the 'space[3]' that can be built by combining its elements. Clearly, nowhere new can be reached by using $\mathbf{u}+\mathbf{v}$ alongside $\mathbf{u}$ and $\mathbf{v}$ alone - hence $\mathbf{u}+\mathbf{v}$ can be removed from the set without changing the span. The same applies to a set $S' = \{\mathbf{v}, \alpha\mathbf{v}\}$; $\alpha\mathbf{v}$ can be removed without changing the space[4] that can be built from the set. We can therefore define a minimal set as the set in which all elements that don't change the span have been removed.

1.5 Basis and Dimension

1.5.1 Basis

Definition

A basis set is a minimal set that spans the space entirely. We have seen that Span(V) is a subspace of V; if it is V itself, then we say the set spans the entire space. Hence, you can obtain a basis for any vector space by removing vectors that don't change the span from the set that forms the space. A basis set is always linearly independent, as if it wasn't there would be vectors that could be removed without changing the span.
A general vector within a space can be written as a linear combination of the basis vectors - this is because by definition they span the space. Say we have a basis set $\mathbf{v}_i$; we may write an arbitrary vector $\mathbf{v}$ like so:

$$\mathbf{v} = \sum_i \alpha_i \mathbf{v}_i$$

Which we say is relative to a set of coordinates, α_i.

[3] I'm using inverted commas because vectors aren't necessarily within actual space. For example, it's possible to construct vector spaces of matrices or of functions - clearly these don't point anywhere in 3D space. This is an abstract 'space' - keep this in mind as I'll drop the inverted commas from now on.

[4] More precisely, the dimension of the space. See the section on basis and dimension.

Claim 9. *A vector within a space is associated with a* ***unique*** *set of coordinates with respect to a particular basis set.*

Proof.

Suppose coordinates are not unique, i.e. we may write $\mathbf{v}$ with respect to two sets of coordinates:

$$\mathbf{v} = \sum_i \alpha_i \mathbf{v}_i = \sum_i \beta_i \mathbf{v}_i$$
$$\sum_i (\alpha_i - \beta_i)\mathbf{v}_i = \mathbf{0}$$

As $\mathbf{v}_i$ are linearly independent, the linear independence condition demands that $\alpha_i - \beta_i = 0$ for all i.

$$\alpha_i = \beta_i$$

□

1.5.2 Dimension

Definition

The dimension of the space is the number of basis vectors required to span it. You may be familiar with the $\mathbf{i}, \mathbf{j}, \mathbf{k}$ basis vectors in 3D space. There are three basis vectors and the space is of dimension 3. There doesn't have to be one basis for a particular space - as long as your set is linearly independent and spans the space, it is one. This means there could be many different basis sets for a particular space. There is no reason to think that different basis sets should have the same dimension - it is reasonable to think that with respect to one basis one might need 3 vectors to span the space but with respect to another it might be possible to find a more 'efficient' way of spanning it with only 2. Lets check this.

Claim 10. *Dimension doesn't depend on the basis chosen.*

Proof.

Consider a basis set $\mathbf{v}_1 \dots \mathbf{v}_n$ of V and a set of vectors $\mathbf{w}_1 \dots \mathbf{w}_m$ s.t. $\mathbf{w}_i \in V$. We may write

$$\mathbf{w}_1 = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$
$$\mathbf{v}_1 = \frac{1}{\alpha_1}\left(\mathbf{w}_1 - \sum_{i=2}^{n} \alpha_i \mathbf{v}_i\right)$$

Notice what we have done; we have written $\mathbf{v}_1$ as a linear combination of $\mathbf{w}_1$ and the other $\mathbf{v}_i$. Consider Span($\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_n\}$): if we were to replace $\mathbf{v}_1$ with our new expression in the span, the only thing that would change would be $\mathbf{v}_1$ being exchanged[5] for $\mathbf{w}_1$. This is because the span already includes the other $\mathbf{v}_i$; they can be removed from the expression without changing the span, as the linear combination $\sum_{i=2}^{n} -\frac{\alpha_i}{\alpha_1}\mathbf{v}_i$ is already accounted for within the span by the other $\mathbf{v}_i$. We now repeat the process, exchanging all $\mathbf{v}_2 \dots \mathbf{v}_n$ with $\mathbf{w}_2 \dots \mathbf{w}_n$ to give $V = \text{Span}(\{\mathbf{w}_1 \dots \mathbf{w}_n\})$.

For $m > n$, $\mathbf{w}_{n+1} \dots \mathbf{w}_m$ are still vectors within V. Hence they must be spanned by $\{\mathbf{w}_1 \dots \mathbf{w}_n\}$ implying they are a linear combination of $\{\mathbf{w}_1 \dots \mathbf{w}_n\}$ and thus the set $\{\mathbf{w}_1 \dots \mathbf{w}_m\}$ is linearly dependent.

In the case $m < n$, not all of the $\mathbf{v}_i$ will be replaced by $\mathbf{w}_i$ so it isn't possible for $\mathbf{w}_i$ to span the space.

Now suppose $\mathbf{w}_i \dots \mathbf{w}_m$ forms a basis for V. This immediately implies the $m < n$ case is impossible, as basis sets must span the whole space. Further, it implies the $m > n$ case is impossible as we showed that the set $\{\mathbf{w}_1 \dots \mathbf{w}_m\}$ is linearly dependent, which basis sets cannot be. Hence the only remaining possibility is $m = n$. We already know n is the dimension of another basis set and as there has been no loss of generality, we have shown the dimension of all other basis sets must also be n.

□

Remarks

Claim 11. *If dim*$(V) = n$ *then any linearly independent* $\{\boldsymbol{v}_1 \dots \boldsymbol{v}_n\}$ *form a basis for* V

Proof.

We take $\{\mathbf{v}_1 \dots \mathbf{v}_n\}$ to be linearly independent and the dimension of V to be n. Assume the set doesn't span V and therefore doesn't form a basis. Now suppose we complete the set to a basis with $\{\mathbf{v}_{n+1} \dots \mathbf{v}_m\}$ to form a set that does in fact span V. This yields a contradiction as we know the number of elements in a basis set must be n. Hence, the assumption is false and $\{\mathbf{v}_1 \dots \mathbf{v}_n\}$ does span V and is therefore a basis.

□

Claim 12. *If* W *is a subspace of* V *and dim*(W) = *dim*(V) *then* $W = V$*.*

Proof.

Let $\dim(W) = \dim(V) = n$. Hence it is possible to find a set of n linearly independent vectors $\{\mathbf{v}_1 \dots \mathbf{v}_n\}$ that form a basis for both sets. Spaces are built from the span of a set of vectors and basis sets span the whole space. We may therefore write the statement $V = \text{Span}(\{\mathbf{v}_1 \dots \mathbf{v}_n\})$ and $W = \text{Span}(\{\mathbf{v}_1 \dots \mathbf{v}_n\})$ so $V = W$.

□

Chapter 2

Linear Maps

2.1 Types and Properties of Linear Map

2.1.1 Definition

A linear map $f : X \to Y$ is a map that maps vectors in a vector space X (the domain) to vectors in Y (the codomain) with the following linear properties:

$$f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}) \tag{2.1}$$

$$f(\alpha\mathbf{v}) = \alpha f(\mathbf{v}) \tag{2.2}$$

Where $\alpha \in F$, the field which X and Y are over.

2.1.2 Image and Kernel

Image

The image of a map, $\text{Im}(f)$, is defined as follows:

$$\text{Im}(f) := \{f(\mathbf{x}) \mid \mathbf{x} \in X\} \tag{2.3}$$

i.e. the set of all vectors $f(\mathbf{x})$ that is produced when applying f to all vectors $\mathbf{x} \in X$. It's the 'section' of Y that can be reached by the action of f on any $\mathbf{x}$. The dimension of the image space is also known as the rank, or $\text{rk}(f)$

Kernel

The kernel of a map, $\ker(f)$, is defined as follows:

$$\ker(f) := \{\mathbf{x} \in X \mid f(\mathbf{x}) = \mathbf{0}\} \tag{2.4}$$

i.e. the set of vectors that map to the zero vector under the action of f.

Claim 13. *Zero vectors map to zero vectors under linear maps*

Proof.

$$\begin{aligned} f(\mathbf{0}) &= f(0\mathbf{v}) \\ &= 0f(\mathbf{v}) \\ &= \mathbf{0} \end{aligned}$$

Hence $\mathbf{0} \in \ker(f)$ for all linear f.

□

It is clear that $\text{Im}(f)$ and $\ker(f)$ are subsets of Y and X. But are they subspaces?

Claim 14. *Im(f) is a subspace of Y*

Proof.

We saw in Claim 13 that the zero vector maps to itself so clearly it's part of the image. Now we show closedness:

Let $\mathbf{u}, \mathbf{v} \in X$. Then $\alpha\mathbf{u} + \beta\mathbf{v} \in X$ as X is closed under addition and scalar multiplication by definition. Let $f(\mathbf{u}), f(\mathbf{v}) \in \text{Im}(f)$. Then:

$$\alpha f(\mathbf{u}) + \beta f(\mathbf{v}) = f(\alpha\mathbf{u} + \beta\mathbf{v})$$

which is clearly still within the image as it's a vector obtained from the action of f on a vector within X. Hence it's a subspace of Y.

□

Claim 15. *ker(f) is a subspace of X*

Proof.

We saw in Claim 13 that the zero vector maps to itself so clearly it's part of the kernel. Now we show closedness:

Let $\mathbf{u}, \mathbf{v} \in \ker(X)$. Then $f(\mathbf{u}) = f(\mathbf{v}) = \mathbf{0}$ by definition. Then:

$$\begin{aligned} f(\alpha\mathbf{u} + \beta\mathbf{v}) &= \alpha f(\mathbf{u}) + \beta f(\mathbf{v}) \\ &= \alpha\mathbf{0} + \beta\mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

which means $\alpha\mathbf{u} + \beta\mathbf{v} \in \ker f$. Hence the kernel is closed under the relevant operations and thus it's a subspace of X.

□

2.1.3 Types of Map

Injective

Each element in Y is mapped to by at most one element in X.

Surjective

Each element in Y is mapped to by at least one element in X.

Bijective

Both injective and surjective. Each element in Y is mapped to by exactly one element in X.
It is also worth remembering that a linear and bijective map between two vector spaces is known as a vector space isomorphism.

Claim 16. *f surjective $\iff$ Im(f) $= Y$*

Proof.

(⇒)

By definition there are no $\mathbf{y} \in Y$ that are not mapped to by at least one $\mathbf{x} \in X$. The image therefore clearly 'covers' the entire space.

□

Proof.

(⇐)

The entire space is mapped to by f. This means every element in Y is mapped to by at least one element in X which is the definition of surjectivity.

□

Claim 17. *Im(f) = Y* $\iff$ *dim(Im(f)) = dim(Y)*

Proof.

(⇒)

If they are the same space then we can choose the same basis set for them - with the same number of basis vectors and therefore the same dimension. I hate using the word obviously but in this case I fear I have no choice.

□

Proof.

(⇐)

Im(f) is a subspace of Y. As proved in Claim 12, if they have the same dimension then they must be the same space.

□

Claim 18. *f injective* $\iff$ *ker(f) =* $\{\mathbf{0}\}$

Proof.

(⇒)

We showed in Claim 13 that the kernel always contains the zero vector. Hence the zero vector is already mapped to by one vector - itself. This means no others can map to it (so the kernel has no other elements) as by definition injective maps have each element in the codomain being mapped to by at most one in the domain. There's no room for anything else to map to $\mathbf{0}$.

□

Proof.

(⇐)

Suppose we have $f(\mathbf{x}_1) = f(\mathbf{x}_2) \mid \mathbf{x}_1, \mathbf{x}_2 \in X$. Then:

$$\begin{aligned} f(\mathbf{x}_1) - f(\mathbf{x}_2) &= \mathbf{0} \\ f(\mathbf{x}_1 - \mathbf{x}_2) &= \mathbf{0} \end{aligned}$$

Which implies $\mathbf{x}_1 - \mathbf{x}_2$ is within $\ker(f)$. But it's given that the kernel only contains the zero vector. Therefore:

$$\begin{aligned} \mathbf{x}_1 - \mathbf{x}_2 &= \mathbf{0} \\ \mathbf{x}_1 &= \mathbf{x}_2 \end{aligned}$$

So we have $f(\mathbf{x}_1) = f(\mathbf{x}_2) \implies \mathbf{x}_1 = \mathbf{x}_2$, i.e. it isn't possible for two different $\mathbf{x} \in X$ to map to the same $\mathbf{y} \in \text{Im}(f)$; the $\mathbf{x}_i$ must be the same. This is simply saying each element in Y is mapped to by at most one element in X - the definition of injectivity.

□

And as a brief aside, the dimension of the space spanned by the zero vector alone is zero. This is because the zero vector only spans itself, creating a space that can be reached by not going anywhere (not using, scaling or adding any vectors), i.e. with zero basis vectors.

2.1.4 Inverse and Identity Maps

Identity Maps

An identity map is defined as $\text{Id}_X : X \to X$, i.e. a map from a space to itself such that $\text{Id}_X(\mathbf{x}) = \mathbf{x}$.

Inverse Maps

The inverse map of f defined to be the map such that when composed onto f, the identity map is produced. In other words:

$$f^{-1} : Y \to X$$

such that

$$\begin{aligned} f^{-1} \circ f &= \text{Id}_X \\ f \circ f^{-1} &= \text{Id}_Y \end{aligned}$$

The distinction between identity maps is important. Say you start at X, apply f to get to Y and then f^{-1} to get to X again. You have mapped from X to X which is then an identity map on X, and vice versa if you start at Y and apply f^{-1} to get back to X, etc..

Claim 19. *f^{-1} exists $\iff$ f is bijective*

Proof.

(⇒)

Suppose we have $f(\mathbf{x}_1) = f(\mathbf{x}_2) \mid \mathbf{x}_1, \mathbf{x}_2 \in X$. Then:

$$\begin{aligned} f(\mathbf{x}_1) &= f(\mathbf{x}_2) \\ f^{-1} \circ f(\mathbf{x}_1) &= f^{-1} \circ f(\mathbf{x}_2) \\ \mathbf{x}_1 &= \mathbf{x}_2 \end{aligned}$$

Which implies injectivity as seen in Claim 18.

We now need to show that there exists an $\mathbf{x}$ that maps to each $\mathbf{y}$. But if f^{-1} exists, we are guaranteed to have $\mathbf{y} = f \circ f^{-1}(\mathbf{y})$, with $f^{-1}(\mathbf{y})$ being the relevant $\mathbf{x} \in X$. In summary, for each $\mathbf{y}$, we know for sure that it'll be mapped to by $\mathbf{x} = f^{-1}(\mathbf{y})$. So f must be surjective. This means f is bijective, as we already showed it was injective.

□

Proof.

($\Leftarrow$)

Each $\mathbf{y}$ is mapped to by one $\mathbf{x}$. All we have to do is define the map f^{-1} that takes us the other way:

$$f(\mathbf{x}) = \mathbf{y} \in Y$$
$$f^{-1}(\mathbf{y}) = \mathbf{x} \in X$$

Which clearly satisfies the requirements for an inverse map as set out in this section - check it yourself.

□

Claim 20. $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$

Proof.

$$(f \circ g)^{-1} \circ f \circ g = \mathrm{Id}$$
$$(f \circ g)^{-1} \circ f = g^{-1}$$
$$(f \circ g)^{-1} = g^{-1} \circ f^{-1}$$

□

2.2 Dimension Theorem

2.2.1 Proof

Claim 21. $f : X \to Y$. *f is linear. Let* $n = dim(X)$, $k = dim(ker(f))$, $r = rk(f)$. *Then* $r + k = n$.

Proof.

Let $\mathbf{v}_1 \dots \mathbf{v}_k$ be a basis for $\ker(f)$. Now let's complete the basis by adding $\mathbf{v}_{k+1} \dots \mathbf{v}_n$ linearly independent vectors to form a basis for X. Remember, this can be done as we proved earlier the kernel is a subspace of the domain.

The image of f is the set of vectors that can be reached by the action of f on vectors in X. Therefore, for some vector $\mathbf{w} \in \mathrm{Im}(f)$, we have

$$\mathbf{w} = f\left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_i\right)$$
$$= \sum_{i=1}^{n} \alpha_i f(\mathbf{v}_i)$$

As $f(\mathbf{v}_1) \ldots f(\mathbf{v}_k) = \mathbf{0}$ as $\mathbf{v}_1 \ldots \mathbf{v}_k \in \ker(f)$, we can drop these from the sum:

$$\mathbf{w} = \sum_{i=k+1}^{n} \alpha_i f(\mathbf{v}_i)$$

So an arbitrary vector in the image can be reached by$f(\mathbf{v}_{k+1}) \ldots f(\mathbf{v}_n)$, which implies that that set spans the image space.

Let's now try and prove $f(\mathbf{v}_{k+1}) \ldots f(\mathbf{v}_n)$ is a linearly independent set, by contradiction. Assume a solution to

$$\sum_{i=k+1}^{n} \alpha_i f(\mathbf{v}_i) = f\left(\sum_{i=k+1}^{n} \alpha_i \mathbf{v}_i\right) = \mathbf{0}$$

exists with $\alpha_j \neq 0$ for some number (≥ 0) of j's such that $n \geq j \geq k+1$. The above expression implies

$$\sum_{i=k+1}^{n} \alpha_i \mathbf{v}_i \in \ker(f).$$

This is impossible as the basis for $\ker(f)$ is $\mathbf{v}_1 \ldots \mathbf{v}_k$. As a result of the contradiction, $\alpha_i = 0$ for all i - which is sufficient criteria to show linear independence.

So $f(\mathbf{v}_{k+1}) \ldots f(\mathbf{v}_n)$ is linearly independent and spans the image. Therefore, it forms a basis for the image. We can find its dimension by counting the basis vectors - there's $n - k$ of them.

$$r = n - k$$
$$r + k = n$$

□

2.2.2 Consequences

An interesting observation regarding dimension and type of map can be made:

Claim 22. *f bijective $\Rightarrow$ dim(X) = dim(Y).*

Proof.

f bijective implies $\dim(\ker(f)) = 0$ which implies $\mathrm{rk}(f) = \dim(X)$. Also, f bijective implies Im(f) = Y. Clearly $\dim(X) = \dim(Y)$ follows.

□

2.3 Matrices

2.3.1 Coordinate Maps

Definition

A coordinate map $\psi : F^n \to X$ is a map from a set of n objects in the field F to a vector space X, such that[1] $\dim(X) = n$. ψ acts on scalars to 'build' a vector within X, with respect to a particular basis of X. This idea is a bit difficult to pin down with words. Essentially, we might write a vector as $\mathbf{v} = \sum_i \alpha_i \mathbf{v}_i$. It is convenient, however, to represent the same vector as the more familiar column vector: $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \end{pmatrix}$ which is written with implied respect to the basis set $\mathbf{v}_i$.

[1] The fact the column vector and the vector space are the same dimension will be important in ensuring the map is bijective, as you'll see later.

This representation makes it convenient to manipulate vectors symbolically, as we will see, with matrices. In particular, F^n means the vector space formed by column vectors of size n with entries all within the field F. The role of the coordinate map is to take a column vector, isolate the coordinates from it and use them to build a vector with respect to a basis of X. For example, we might have the vector $\left(\begin{smallmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{smallmatrix}\right)$ with respect to a basis set $\mathbf{v}_i$. We then use the map ψ like so:

$$\psi \begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \sum_1^3 \alpha_i \mathbf{v}_i$$

Again, the idea here is the coordinate map builds 'actual' vectors with respect to a basis from column vectors.

Properties

Lets think about what type of map these maps are. First, we know from the definition ($\psi : F^n \to X$) that the basis vectors used in the 'building' process are $\in X$. Secondly, we know by definition that the dimension of the column vector and the vector space ψ maps them to are the same. This implies that in the general case, the image of ψ is a linear combination of n out of n basis vectors of X; this means that the image covers the codomain X completely, as a general linear combination of all basis vectors must span the space by the definition of a basis. Now, remembering claim 16, this is exactly the condition for ψ to be surjective. Further, remembering claim 17, we know the dimension of the image is equal to the dimension of the codomain, namely n. But n is the dimension of the domain as well. Hence, by the dimension theorem, the dimension of the kernel is 0 and thus by claim 18 ψ is injective. Therefore ψ is bijective, and thus invertible; there exists a map ψ^{-1} which takes you from a vector to a column vector. This can also be seen intuitively from the fact that coordinates of a vector are unique - there is only one actual vector corresponding to each column vector. Finally, it's left as an exercise to the reader to show that coordinate maps are linear[2].

2.3.2 Why Matrices are Linear Maps

Say we have a map $f : X \to Y$. We know the vector spaces X and Y have corresponding column vector spaces, say of dimension n and m respectively. We like to work symbolically with column vectors, so we'd like to find an object that represents the action of f on these column vectors. The situation is as follows:

$$\begin{array}{ccccc} & & f & & \\ & X & \longrightarrow & Y & \\ \varphi & \uparrow & & \uparrow & \psi \\ & F^n & \longrightarrow & F^m & \\ & & A & & \end{array}$$

So f takes us from X to Y, our coordinate maps φ and ψ take us from column vector spaces to actual vector spaces and our mystery object, A, takes us between column vectors. By simply following the diagram from the tail of A's arrow to its tip, we can find an equivalent expression for A's action:

$$A(\boldsymbol{\alpha}) = \psi^{-1} \circ f \circ \varphi(\boldsymbol{\alpha})$$

Where $\boldsymbol{\alpha}$ is the n-dimensional column vector we want f to act on. Now, let $\mathbf{v}_i$ be a basis for X and $\mathbf{w}_i$ be a basis for Y. We can make the following observations:

$$A(\boldsymbol{\alpha}) = \psi^{-1} \circ f \left(\sum_{i=1}^n \alpha_i \mathbf{v}_i\right)$$

$$A(\boldsymbol{\alpha}) = \psi^{-1} \left(\sum_{i=1}^n \alpha_i f(\mathbf{v}_i)\right)$$

We know $f(\mathbf{v}_i) \in Y$ so we may expand it in terms of Y's basis:

$$f(\mathbf{v}_i) = \sum_{j=1}^m \beta_{ji} \mathbf{w}_j$$

[2] Sorry, I really couldn't be bothered to add and format this in. It should be straightforward enough.

Note that the iterated variable here is j, but β has an i subscript as it will depend on the choice of $\mathbf{v}_i$. In other words the expansion of $f(\mathbf{v}_1)$ will have different β coefficients than the expansion of $f(\mathbf{v}_2)$. Now, subbing in this in to our previous expression:

$$A(\boldsymbol{\alpha}) = \psi^{-1}\left(\sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \beta_{ji}\mathbf{w}_j\right)$$

Rearranging:

$$A(\boldsymbol{\alpha}) = \psi^{-1}\left(\sum_{j=1}^{m} \mathbf{w}_j \sum_{i=1}^{n} \alpha_i\beta_{ji}\right)$$

Now notice how $\sum_{i=1}^{n} \alpha_i\beta_{ji}$ is just a scalar, lets call it γ_j.

$$A(\boldsymbol{\alpha}) = \psi^{-1}\left(\sum_{j=1}^{m} \gamma_j\mathbf{w}_j\right)$$

Now notice what we have - a linear combination of Y's basis vectors with a set of coordinates γ_j. This is exactly the setup that ψ^{-1} can deconstruct into column vectors! Lets do that then:

$$A(\boldsymbol{\alpha}) = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \alpha_i\beta_{1i} \\ \sum_{i=1}^{n} \alpha_i\beta_{2i} \\ \vdots \\ \sum_{i=1}^{n} \alpha_i\beta_{mi} \end{pmatrix}$$

$$A(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha_1\beta_{11} \\ \alpha_1\beta_{21} \\ \vdots \\ \alpha_1\beta_{m1} \end{pmatrix} + \begin{pmatrix} \alpha_2\beta_{12} \\ \alpha_2\beta_{22} \\ \vdots \\ \alpha_2\beta_{m2} \end{pmatrix} + \cdots + \begin{pmatrix} \alpha_n\beta_{1n} \\ \alpha_n\beta_{2n} \\ \vdots \\ \alpha_n\beta_{mn} \end{pmatrix}$$

$$A(\boldsymbol{\alpha}) = \alpha_1\begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{m1} \end{pmatrix} + \alpha_2\begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \vdots \\ \beta_{m2} \end{pmatrix} + \cdots + \alpha_n\begin{pmatrix} \beta_{1n} \\ \beta_{2n} \\ \vdots \\ \beta_{mn} \end{pmatrix}$$

Notice that the column vectors are simply those representing the image of the corresponding $f(\mathbf{v}_i)$, but in column vector form. Another way of writing this would be:

$$A(\boldsymbol{\alpha}) = \alpha_1\boldsymbol{\beta}^1 + \alpha_2\boldsymbol{\beta}^2 + \cdots + \alpha_n\boldsymbol{\beta}^n$$

This is an incredibly useful result. It describes the procedure for symbolically carrying out the action of a linear map on a column vector to produce another column vector. It says that the procedure is to take the first coordinate of the input column vector and multiply by the image of the first domain basis under the action of f.[3] Then, repeat for the second and so on and add them all up. If you haven't seen this result before, try it out with a couple of 2x2 matrices and convince yourself this is in fact equivalent to the traditional method of matrix-vector multiplication.
We would like to separate the α parts and the $\boldsymbol{\beta}^i$ parts of our expression for $A(\boldsymbol{\alpha})$ as this will allow us to extract the object that represents A. Lets group the $\boldsymbol{\beta}^i$ column vectors together into a matrix, and put $\boldsymbol{\alpha}$ separately to the side:

$$A(\boldsymbol{\alpha}) = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \ddots & & \vdots \\ \vdots & & & \\ \beta_{m1} & \dots & & \beta_{mn} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

[3] This image must be in coordinate vector form and **with respect to the basis of the codomain.**

From this we can notice three things:

1. The object A is fully defined by the images of the codomain's basis vectors. These are put in column vector form and arranged into a matrix to give A.
2. An $m \times n$ matrix is a map from $F^n \to F^m$. Hence an $n \times n$ matrix maps between two column vector spaces of the same dimension.
3. For a matrix and column vector to be multiplicable, the number of columns of the matrix must equal the number of rows of the vector. This ensures the vector is of the same dimension as the domain of the matrix's corresponding map.

Finally, in order to ensure we do the aforementioned coordinate multiplication with image vectors and sum procedure, we will have to figure out some weird way of doing the symbolic application of A to $\boldsymbol{\alpha}$. The method that follows this procedure is exactly the strange way we normally multiply matrices and vectors.

2.3.3 Derivation of Matrix Multiplication Rules

We have seen that matrices represent linear maps between column vector spaces. Linear maps can be composed, so we should now derive the relationship that allows two matrices to be composed, a process we call matrix multiplication.
Consider a map $g : Y \to Z$ such that the following is satisfied:

$$\begin{array}{ccccc} & & g & & \\ & Y & \longrightarrow & Z & \\ \psi & \uparrow & & \uparrow & \omega \\ & F^m & \longrightarrow & F^l & \\ & & B & & \end{array}$$

This implies we can make the map $h : X \to Z$ which satisfies $h = g \circ f$. Also, if B is the matrix representing the action of g on an m dimensional coordinate vector, then we can compose the matrices to produce $C(\boldsymbol{\alpha}) = BA(\boldsymbol{\alpha})$. Here, C represents the action of h on an n dimensional column vector to produce an l dimensional one. Notice how the matrices are composed in the same order the maps are - right to left.
Lets go through the mildly tedious process of finding C. From the diagram we have

$$B(\boldsymbol{\alpha}) = \omega^{-1} \circ g \circ \psi(\boldsymbol{\alpha})$$

But we want $B(A(\boldsymbol{\alpha}))$ instead[4], so we just replace:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1} \circ g \circ \psi(A(\boldsymbol{\alpha}))$$

From our previous result we already know what $A(\boldsymbol{\alpha})$ is - we can just sub it in:

$$A(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i \boldsymbol{\beta}^i$$

$$B(A(\boldsymbol{\alpha})) = \omega^{-1} \circ g \circ \sum_{i=1}^{n} \alpha_i \psi(\boldsymbol{\beta}^i)$$

Using the linearity of ψ. As $\boldsymbol{\beta}^i$ is a column vector, and ψ is a coordinate map with respect to the $\mathbf{w}_i$ basis, $\psi(\boldsymbol{\beta}^i)$ will build a vector with coordinates that are the components of $\boldsymbol{\beta}^i$, namely β_{ji}:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1} \circ g \circ \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{m} \beta_{ji} \mathbf{w}_j \right)$$

[4] If this isn't clear, think about how a composition is just applying one map after another; this is equivalent to applying the second to the result, or output, of the first.

And using the linearity of g:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1}\left(\sum_{i=1}^{n}\alpha_i\sum_{j=1}^{m}\beta_{ji}g(\mathbf{w}_j)\right)$$

We can expand the image of the basis vectors $g(\mathbf{w}_j)$ in the same manner as before:

$$g(\mathbf{w}_j) = \sum_{k=1}^{l}\delta_{kj}\mathbf{z}_k$$

Again noting the two subscripts required for δ as it depends on the choice of $\mathbf{w}_j$ that the image is taken of. Here, $\mathbf{z}_k$ are the basis vectors of the space Z. Substituting in:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1}\left(\sum_{i=1}^{n}\alpha_i\sum_{j=1}^{m}\beta_{ji}\sum_{k=1}^{l}\delta_{kj}\mathbf{z}_k\right)$$

Rearranging:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1}\left(\sum_{k=1}^{l}\mathbf{z}_k\underbrace{\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\beta_{ji}\delta_{kj}\right]}_{\varepsilon_k}\right)$$

The object in square brackets is just a scalar varying with k, so we can call it ε_k. We have:

$$B(A(\boldsymbol{\alpha})) = \omega^{-1}\left(\sum_{k=1}^{l}\varepsilon_k\mathbf{z}_k\right)$$

And as ω^{-1} 'deconstructs' vectors made of $\mathbf{z}_k$, we can easily see that this will produce

$$B(A(\boldsymbol{\alpha})) = \begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_l\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{n}\alpha_i\sum_{j=1}^{m}\beta_{ji}\delta_{1j}\\ \sum_{i=1}^{n}\alpha_i\sum_{j=1}^{m}\beta_{ji}\delta_{2j}\\ \vdots\\ \sum_{i=1}^{n}\alpha_i\sum_{j=1}^{m}\beta_{ji}\delta_{lj}\end{pmatrix}$$

$$B(A(\boldsymbol{\alpha})) = \alpha_1\begin{pmatrix}\sum_{j=1}^{m}\beta_{j1}\delta_{1j}\\ \sum_{j=1}^{m}\beta_{j1}\delta_{2j}\\ \vdots\\ \sum_{j=1}^{m}\beta_{j1}\delta_{lj}\end{pmatrix} + \alpha_2\begin{pmatrix}\sum_{j=1}^{m}\beta_{j2}\delta_{1j}\\ \sum_{j=1}^{m}\beta_{j2}\delta_{2j}\\ \vdots\\ \sum_{j=1}^{m}\beta_{j2}\delta_{lj}\end{pmatrix} + \cdots + \alpha_n\begin{pmatrix}\sum_{j=1}^{m}\beta_{jn}\delta_{1j}\\ \sum_{j=1}^{m}\beta_{jn}\delta_{2j}\\ \vdots\\ \sum_{j=1}^{m}\beta_{jn}\delta_{lj}\end{pmatrix}$$

We can see $B(A(\boldsymbol{\alpha}))$ is in the same form as $A(\boldsymbol{\alpha})$, so clearly the equivalent matrix C exists. As before, we just have to separate off the α_i and group the columns into a matrix:

$$B(A(\boldsymbol{\alpha})) = \begin{pmatrix}\sum_{j=1}^{m}\beta_{j1}\delta_{1j} & \sum_{j=1}^{m}\beta_{j2}\delta_{1j} & \dots & \sum_{j=1}^{m}\beta_{jn}\delta_{1j}\\ \sum_{j=1}^{m}\beta_{j1}\delta_{2j} & \ddots & & \vdots\\ \vdots & & & \\ \sum_{j=1}^{m}\beta_{j1}\delta_{lj} & \dots & & \sum_{j=1}^{m}\beta_{jn}\delta_{lj}\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix} = C(\boldsymbol{\alpha})$$

From which we can easily see what the equivalent matrix C is.

We have arrived at the result:

$$(BA)_{pq} = C_{pq} = \sum_{j=1}^{m} \delta_{pj}\beta_{jq}$$

But δ_{pq} are just the elements of the matrix representing g (namely B) and β_{pq} are those of the matrix representing f (namely A). Hence:

$$(BA)_{pq} = C_{pq} = \sum_{j=1}^{m} B_{pj}A_{jq}$$

2.4 Change of Basis

2.4.1 Basis Translation Process

We have found an expression for a matrix that represents the action of a linear map on a coordinate vector. Recall the relationships

$$\begin{array}{ccccc} & & f & & \\ & X & \longrightarrow & Y & \\ \varphi & \uparrow & & \uparrow & \psi \\ & F^n & \longrightarrow & F^m & \\ & & A & & \end{array}$$

$$A = \psi^{-1} \circ f \circ \varphi$$

This matrix is linked to two basis sets: the first being the basis of X (lets call it $\mathbf{x}_i$) which φ is with respect to and the second being the basis of Y (lets call it $\mathbf{y}_i$ which ψ is with respect to. Suppose we wanted to replace these basis sets with $\mathbf{x}'_i$ and $\mathbf{y}'_i$ respectively, and find the equivalent matrix with respect to these new basis sets. Our new relationship would be

$$A' = (\psi')^{-1} \circ f \circ \varphi'$$

We can add in some strategically placed identity maps that won't change the overall expression:

$$A' = (\psi')^{-1} \circ (\psi \circ \underbrace{\psi^{-1}) \circ f \circ (\varphi}_{A} \circ \varphi^{-1}) \circ \varphi'$$
$$A' = (\psi')^{-1} \circ \psi \circ A \circ \varphi^{-1} \circ \varphi'$$

Now let's think about what happens when we give A' an input coordinate vector with respect to $\mathbf{x}'_i$, say $\boldsymbol{\alpha}'$:

1. φ' builds a vector with respect to $\mathbf{x}'_i$
2. φ^{-1} deconstructs the vector into a coordinate vector with respect to $\mathbf{x}_i$, which is a form that A can understand
3. A acts on this coordinate vector to produce a coordinate vector with respect to $\mathbf{y}_i$
4. ψ constructs a vector with respect to $\mathbf{y}_i$
5. ψ^{-1} deconstructs this into a coordinate vector with respect to $\mathbf{y}'_i$

So whereas before we had $\mathbf{x}_i$ in, $\mathbf{y}_i$ out, we now have $\mathbf{x}'_i$ in and $\mathbf{y}'_i$ out through a rather complicated translation process.

2.4.2 Priming and Unpriming Matrices

Lets consider the composition $(\psi')^{-1} \circ \psi$. This is a map $p_Y : F^m \to F^m$ and so has an equivalent matrix, which we will call P_Y. Lets look at what P_Y actually does; first it takes in a coordinate vector with respect to $\mathbf{y}_i$ and then does some constructing and deconstructing with the end result of outputting a coordinate vector with respect to $\mathbf{y}'_i$. Therefore, we call P_Y a priming matrix as it converts a coordinate vector with respect to an unprimed basis into one with respect to a primed basis. In particular, the Y subscript indicates it's a priming matrix for the vector space Y, given a set of primed and unprimed basis sets.

We would like to show that this map is bijective and thus has an inverse. It is impossible for p_Y to map different unprimed coordinate vectors to the same primed one as that would imply different vectors in one basis were the same vector in the other - this is clearly impossible as vectors exist independently of basis, they can simply be described relative to one. Further, it isn't possible for an unprimed coordinate vector to not have an equivalent primed coordinate vector as this would imply the primed basis didn't span the space. This is impossible by the definition of basis. Hence, priming maps must have a 1:1 correspondence between primed and unprimed coordinate vectors and are therefore bijective. As we showed earlier, all bijective maps are invertible.
Notice that $\varphi^{-1} \circ \varphi' = ((\varphi')^{-1} \circ \varphi)^{-1}$. This is in the form of the inverse of a priming map, so it must be a map that unprimes coordinate vectors. We can call its inverse map $p_X : F^n \to F^n$ with equivalent matrix P_X. Again, notice the X subscript: given a primed set and an unprimed set of basis vectors of X, P_X converts between them. Overall then, we can write A' in matrix form:

$$A' = P_Y A P_X^{-1} \tag{2.5}$$

Clearly then, the process of changing the basis a matrix is with respect to is a case of finding P_X and P_Y, and then simply substituting in.
It would also be convenient to know how to explicitly find priming matrices. It happens that in most cases, one tends to find the unpriming matrices and then invert them. Suppose we have a primed basis $\mathbf{x}'_i$ and an unprimed one $\mathbf{x}_i$. Suppose further we are given that

$$\mathbf{x}'_i = \sum_j \alpha_{ji} \mathbf{x}_j$$

i.e. we are give the coordinates of the primed set with respect to the unprimed set. Now lets see what we would get if we arranged these coordinates into a matrix:

$$\begin{pmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{pmatrix}$$

We already know the columns of the matrix are the images of the basis vectors. But the ith column is precisely the column vector of $\mathbf{x}'_i$ with respect to $\mathbf{x}_i$. If we give this matrix $\mathbf{e}_i$ as an input - which can be interpreted as the representation of the ith primed basis vector with respect to its *own* basis set - it gives back the representation of the ith primed basis vector with respect to the unprimed set. Hence, this matrix is an unpriming matrix. This is a bit tricky to wrap your head around, so the following example should hopefully clear it up.

2.4.3 Example

Suppose we have two vector spaces X and Y, of dimension 3 and 2 respectively. X has an unprimed basis set $\mathbf{x}_i$ and primed set

$$\mathbf{x}'_1 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{x}'_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \quad \mathbf{x}'_3 = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}$$

with respect to $\mathbf{x}_i$. Just to be crystal clear, this means $\mathbf{x}'_1 = 2\mathbf{x}_1 + 0\mathbf{x}_2 + 1\mathbf{x}_3$. Also, Y has an unprimed basis set of $\mathbf{y}_i$ and primed set

$$\mathbf{y}'_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad \mathbf{y}'_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

with respect to $\mathbf{y}_i$.
Suppose we have a map $a : X \to Y$ with an equivalent matrix mapping between F^3 and F^2:

$$A = \begin{pmatrix} 1 & 3 & -2 \\ 4 & 0 & 1 \end{pmatrix}.$$

If you read the section on matrices as linear maps, you should know that the numbers in the matrix are with respect to the basis $\mathbf{y}_i$. We would like to find, however, this matrix with respect to the primed basis sets. In other words, the matrix that takes in coordinate vectors that are with respect to $\mathbf{x}'_i$ and yields coordinate vectors with respect to $\mathbf{y}'_i$.

Using Priming Matrices

Let's look at the statement $\mathbf{x}'_1 = \left(\begin{smallmatrix} 2 \\ 0 \\ 1 \end{smallmatrix}\right)$. Notice $\mathbf{x}'_1 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$ with respect to the primed basis set, so some map exists that takes in $\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$ with respect to the primed set of X and yields $\left(\begin{smallmatrix} 2 \\ 0 \\ 1 \end{smallmatrix}\right)$ with respect to the unprimed set. This map is therefore the unpriming map of X, namely p_X^{-1}.
Recall that the columns of a matrix are simply the coordinate vectors of the images of the domain's basis set, with respect to the codomain's basis set. Of course, this still applies to priming matrices and thus, as the basis of the domain of P_X^{-1} is the primed set, and $\left(\begin{smallmatrix} 2 \\ 0 \\ 1 \end{smallmatrix}\right)$ is the image of the first basis vector in said domain[5] then this must form the first column of P_X^{-1}. Repeating this for the other columns yields

$$P_X^{-1} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & -1 & 3 \end{pmatrix}$$

Similarly for the $\mathbf{y}_i$ set, we have

$$P_Y^{-1} = \begin{pmatrix} 3 & 0 \\ 2 & -1 \end{pmatrix}$$

We'll cover how to find matrix inverses later, but for now just trust me that

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

Now all that's left is to substitute into equation 2.5, yielding

$$A' = \begin{pmatrix} 0 & 2 & 0 \\ -9 & 1 & -3 \end{pmatrix}$$

Without Priming Matrices

It is possible to do the 'translation' work of the priming and unpriming matrices ourselves, and therefore find A' without explicitly calculating them. Recall again that the columns of A' will be the images of the primed X basis vectors. We can find these images by translating the primed basis vectors to be with respect to the unprimed ones, using A to find their images and then translating back.
No actual work needs to be done translating the primed X vectors in to the primed ones: they're already written with respect to the unprimed ones. So now we just operate on them with A:

$$A\mathbf{x}'_1 = \begin{pmatrix} 0 \\ 9 \end{pmatrix}$$

$$A\mathbf{x}'_2 = \begin{pmatrix} 6 \\ 3 \end{pmatrix}$$

$$A\mathbf{x}'_3 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

[5] I.e. the first primed basis vector in X

Which are with respect to $\mathbf{y}_i$. Hence, all we have to do is convert these into $\mathbf{y}'_i$ and we're done.

$$\begin{pmatrix} 0 \\ 9 \end{pmatrix} = a\mathbf{y}_1 + b\mathbf{y}_2$$
$$\begin{pmatrix} 0 \\ 9 \end{pmatrix} = \begin{pmatrix} 3a \\ 2a - b \end{pmatrix}$$

Which implies $a = 0$ and $b = -9$. This means $A\mathbf{x}_1 = \left(\begin{smallmatrix} 0 \\ -9 \end{smallmatrix}\right)$ with respect to $\mathbf{y}'_i$. Repeating for the other basis vectors and then arranging the results into columns:

$$A' = \begin{pmatrix} 0 & 2 & 0 \\ -9 & 1 & -3 \end{pmatrix}$$

Which is the result we arrived at earlier.

Chapter 3

Scalar Products

3.1 Properties

3.1.1 Definition

The hermitian scalar product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ is a map from two vectors within a vector space V to the complex numbers. The map must satisfy the following properties:

1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle^*$ (take conjugate when swapping arguments)
2. $\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{v} \rangle + \beta\langle \mathbf{u}, \mathbf{w} \rangle$ (linear in second argument)
3. $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ if $\mathbf{v} \neq \mathbf{0}$ (positive definite)

Regarding the third point, if one of the vectors within a scalar product is zero we may take out a scalar of 0 due to linearity making the whole thing zero.

3.1.2 Magnitude of a Vector

We can define the magnitude of a vector $|\mathbf{v}| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Claim 23. *$|\boldsymbol{a}|$ is a purely real quantity.*

Proof.

$$|\mathbf{a}|^2 = \langle \mathbf{a}, \mathbf{a} \rangle$$

Clearly $\langle \mathbf{a}, \mathbf{a} \rangle = \langle \mathbf{a}, \mathbf{a} \rangle^*$ from the conjugate flip property, so $|\mathbf{a}|^2$ is a purely real quantity. Suppose, in the general case, $|\mathbf{a}| = \alpha + \beta i \mid \alpha, \beta \in \mathbb{R}$:

$$|\mathbf{a}|^2 = \alpha^2 + 2\alpha\beta i - \beta^2 \in \mathbb{R}$$

Which implies either $\alpha = 0$ or $\beta = 0$. Consider the former:

$$|\mathbf{a}|^2 = -\beta^2 < 0$$

which is impossible by the positive definite property. Hence, $\beta = 0$ so $|\mathbf{a}| \in \mathbb{R}$.

□

3.1.3 Function Application Uniqueness

Consider linear maps $f : X \to X$ and $g : X \to X$ with $\mathbf{v}, \mathbf{w} \in X$.

Claim 24. $\langle \boldsymbol{v}, f(\boldsymbol{w}) \rangle = \langle \boldsymbol{v}, g(\boldsymbol{w}) \rangle \implies f = g$

Proof.

$$\begin{aligned} \langle \mathbf{v}, f(\mathbf{w})\rangle - \langle \mathbf{v}, g(\mathbf{w})\rangle &= 0 \\ \langle \mathbf{v}, f(\mathbf{w}) - g(\mathbf{w})\rangle &= 0 \end{aligned}$$

As this must hold for all cases, including the ones where $\mathbf{v} \neq \mathbf{0}$, then $f(\mathbf{w}) = g(\mathbf{w})$. Since this must hold for all $\mathbf{w}$, then $f = g$.

□

Claim 25. $\langle f(\boldsymbol{v}), \boldsymbol{w}\rangle = \langle g(\boldsymbol{v}), \boldsymbol{w}\rangle \implies f = g$

Proof.

$$\begin{aligned} \langle \mathbf{w}, f(\mathbf{v})^* &= \langle \mathbf{w}, g(\mathbf{v})^* \\ (\langle \mathbf{w}, f(\mathbf{v})\rangle - \langle \mathbf{w}, g(\mathbf{v})\rangle)^* &= 0 \\ \langle f(\mathbf{v}) - g(\mathbf{v}), \mathbf{w}\rangle^* &= 0 \\ \langle \mathbf{w}, f(\mathbf{v}) - g(\mathbf{v})\rangle &= 0 \end{aligned}$$

As this must hold for all cases, including the ones where $\mathbf{w} \neq \mathbf{0}$, then $f(\mathbf{v}) = g(\mathbf{v})$. Since this must hold for all $\mathbf{v}$, then $f = g$.

□

3.1.4 Cauchy-Schwarz Inequality

Claim 26. $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq |\boldsymbol{x}||\boldsymbol{y}|$

Proof.

Consider $\langle \mathbf{x} + t\mathbf{y}, \mathbf{x} + t\mathbf{y}\rangle$ with $t \in \mathbb{C}$.

$$\begin{aligned} 0 &\leq \langle \mathbf{x} + t\mathbf{y}, \mathbf{x} + t\mathbf{y}\rangle \\ &= \langle \mathbf{x} + t\mathbf{y}, \mathbf{x}\rangle + \langle \mathbf{x} + t\mathbf{y}, t\mathbf{y}\rangle \\ &= \langle \mathbf{x}, \mathbf{x} + t\mathbf{y}\rangle^* + \langle t\mathbf{y}, \mathbf{x} + t\mathbf{y}\rangle^* \\ &= \langle \mathbf{x}, \mathbf{x}\rangle + (t\langle \mathbf{x}, \mathbf{y}\rangle)^* + t\langle \mathbf{x}, \mathbf{y}\rangle + |t|^2\langle \mathbf{y}, \mathbf{y}\rangle \\ &= \langle \mathbf{x}, \mathbf{x}\rangle + |t|^2\langle \mathbf{y}, \mathbf{y}\rangle + 2\mathrm{Re}(t\langle \mathbf{x}, \mathbf{y}\rangle) \end{aligned}$$

Now choose $t = -\frac{\langle \mathbf{x}, \mathbf{y}\rangle^*}{\langle \mathbf{y}, \mathbf{y}\rangle}$

$$\begin{aligned} &= |\mathbf{x}|^2 + \frac{|\langle \mathbf{x}, \mathbf{y}\rangle|^2}{\langle \mathbf{y}, \mathbf{y}\rangle} - 2\frac{|\langle \mathbf{x}, \mathbf{y}\rangle|^2}{\langle \mathbf{y}, \mathbf{y}\rangle} \\ &= |\mathbf{x}|^2 - \frac{|\langle \mathbf{x}, \mathbf{y}\rangle|^2}{|\mathbf{y}|^2} \geq 0 \\ |\mathbf{x}|^2|\mathbf{y}|^2 &\geq |\langle \mathbf{x}, \mathbf{y}\rangle|^2 \\ |\langle \mathbf{x}, \mathbf{y}\rangle| &\leq |\mathbf{x}||\mathbf{y}| \end{aligned}$$

□

3.1.5 Triangle Inequality

Claim 27. $|\boldsymbol{a} + \boldsymbol{b}| \leq |\boldsymbol{a}| + |\boldsymbol{b}|$

Proof.

$$\begin{aligned}
|\mathbf{a}+\mathbf{b}|^2 &= \langle \mathbf{a}+\mathbf{b}, \mathbf{a}+\mathbf{b}\rangle \\
&= |\mathbf{a}|^2 + |\mathbf{b}|^2 + (\langle \mathbf{a}, \mathbf{b}\rangle + \langle \mathbf{a}, \mathbf{b}\rangle^*) \\
&= |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2\mathrm{Re}(\langle \mathbf{a}, \mathbf{b}\rangle) \\
|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}|^2 - |\mathbf{b}|^2 &\leq 2|\langle \mathbf{a}, \mathbf{b}\rangle| \leq 2|\mathbf{a}||\mathbf{b}| && \text{(using Cauchy-Schwarz)} \\
|\mathbf{a}+\mathbf{b}|^2 &\leq (|\mathbf{a}| + |\mathbf{b}|)^2 \\
|\mathbf{a}+\mathbf{b}| &\leq |\mathbf{a}| + |\mathbf{b}|
\end{aligned}$$

□

3.1.6 Angles Between Vectors

The Cauchy-Schwarz inequality allows us to define the extent to which one side is less than the other by defining the angle θ between two vectors such that

$$\cos\theta := \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{|\mathbf{x}||\mathbf{y}|} \tag{3.1}$$

which is valid when $\langle \mathbf{x}, \mathbf{y}\rangle$ is a real number. This yields the interpretation of the real scalar product being the magnitude of the projection of one vector in the direction of the other.

3.1.7 Orthogonality

We say two vectors are orthogonal if their scalar product is zero. This has the consequence of mandating orthogonal and non-zero vectors to be linearly independent.

Claim 28. *Consider a set of vectors S such that $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = 0$ ($\boldsymbol{v}_i, \boldsymbol{v}_j \in S, \boldsymbol{v}_i \neq \boldsymbol{0}$) for all i, i.e. all vectors in S are pairwise orthogonal and non-zero. Then all vectors in the set are linearly independent.*

Proof.

Linear independence condition:

$$\sum_i \alpha_i \mathbf{v}_i = 0$$

Take scalar product of both sides:

$$\begin{aligned}
\left\langle \mathbf{v}_j, \sum_i \alpha_i \mathbf{v}_i \right\rangle &= 0 \\
\sum_i \alpha_i \langle \mathbf{v}_j, \mathbf{v}_i\rangle &= 0
\end{aligned}$$

The left side is zero for all combinations except j with itself due to pairwise orthogonality. Hence we have

$$\alpha_j \langle \mathbf{v}_j, \mathbf{v}_j\rangle = 0$$

And due to the positive definite property $\langle \mathbf{v}_j, \mathbf{v}_j\rangle > 0$ so α_j must =0. As we never specified $\mathbf{v}_j$ apart from it being within S, $\alpha_j = 0$ for all j. Hence they are linearly independent.

□

3.2 Orthonormal Basis Sets

3.2.1 Definition

In this section we will make use of the Kronecker Delta symbol, δ. This will be defined in section 6.1.1; skip ahead to there and then come back if you aren't already familiar with it.
A set of vectors $\boldsymbol{\varepsilon}_i$ is said to be orthonormal if $\langle \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j \rangle = \delta_{ij}$. Therefore, they are all mutually orthogonal and normalised.

3.2.2 Gram-Schmidt Procedure

This is the procedure for creating an orthonormal basis set from an arbitrary basis set of vectors $\mathbf{v}_i$.

Begin with $\mathbf{v}_1$. Normalise to produce $\boldsymbol{\varepsilon}_1$:

$$\boldsymbol{\varepsilon}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}$$

Now take $\mathbf{v}_2$. We subtract off all components in the direction of $\boldsymbol{\varepsilon}_1$ to ensure orthogonality:

$$\mathbf{v}_2' = \mathbf{v}_2 - \langle \boldsymbol{\varepsilon}_1, \mathbf{v}_2 \rangle \boldsymbol{\varepsilon}_1$$

The scalar product extracts the magnitude of these components and the normalised factor of $\boldsymbol{\varepsilon}_1$ gives the components in the relevant direction. Now normalise:

$$\boldsymbol{\varepsilon}_2 = \frac{\mathbf{v}_2'}{|\mathbf{v}_2'|}$$

and repeat for the rest of the vectors in the set. For the ith vector, subtract off all the components in the directions of $\boldsymbol{\varepsilon}_1 \dots \boldsymbol{\varepsilon}_{i-1}$, and then normalise.

Notice that $\text{Span}(\{\mathbf{v}_i\}) = \text{Span}(\{\boldsymbol{\varepsilon}_i\})$. Clearly $\text{Span}(\mathbf{v}_1) = \text{Span}(\boldsymbol{\varepsilon}_1)$ as they differ by a rescaling. Also, notice $\text{Span}(\{\mathbf{v}_1, \mathbf{v}_2\}) = \text{Span}(\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2\})$ as $, \boldsymbol{\varepsilon}_2$ differs from $\mathbf{v}_2$ by a scale factor and a linear combination of $\boldsymbol{\varepsilon}_1$, which we agreed was already in the span and thus doesn't affect it. The same observation can be repeated for $\boldsymbol{\varepsilon}_3$ and so on, such that $\text{Span}(\{\mathbf{v}_i\}) = \text{Span}(\{\boldsymbol{\varepsilon}_i\})$. As $\{\mathbf{v}_i\}$ was defined to be a basis set and has the same span as $\{\boldsymbol{\varepsilon}_i\}$, then the latter is also a basis set.
Thanks to the Gram-Schmidt Procedure we are able to write any vector with respect to an orthonormal basis set, which has many useful consequences.

3.2.3 Finding Coordinates

Lets write a vector $\mathbf{v}$ in terms of an orthonormal basis set:

$$\mathbf{v} = \sum_i \alpha_i \boldsymbol{\varepsilon}_i$$

and take the scalar product with another basis vector within the basis set:

$$\begin{aligned}
\langle \boldsymbol{\varepsilon}_j, \mathbf{v} \rangle &= \left\langle \boldsymbol{\varepsilon}_j, \sum_i \alpha_i \boldsymbol{\varepsilon}_i \right\rangle \\
&= \sum_i \alpha_i \langle \boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_i \rangle \\
&= \sum_i \alpha_i \delta_{ij} \\
&= \alpha_j
\end{aligned}$$

So we have extracted the jth coordinate by taking a scalar product with the jth basis vector in the orthonormal set.

3.2.4 Computation of the Scalar Product

Once again, writing vectors with respect to an orthonormal basis has useful consequences - we can find how to actually compute the scalar product of two vectors.

Let $\mathbf{v} = \sum_i \alpha_i \boldsymbol{\varepsilon}_i, \mathbf{w} = \sum_i \beta_i \boldsymbol{\varepsilon}_i$. Then:

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{w} \rangle &= \left\langle \sum_i \alpha_i \boldsymbol{\varepsilon}_i, \sum_j \beta_j \boldsymbol{\varepsilon}_j \right\rangle \\
&= \sum_j \beta_j \left\langle \sum_i \alpha_i \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j \right\rangle \\
&= \sum_i \sum_j \alpha_i^* \beta_j \langle \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j \rangle \\
&= \sum_i \sum_j \alpha_i^* \beta_j \delta_{ij} \\
&= \sum_i \alpha_i^* \beta_i
\end{aligned}$$

3.2.5 Finding Matrix Elements

Consider a linear map $f : V \to V$ with an equivalent matrix A with entris with respect to an orthonormal basis $\boldsymbol{\varepsilon}_i$ of V. We know the columns of A are the images of $\boldsymbol{\varepsilon}_i$ under the action of f, which allows us to write

$$f(\boldsymbol{\varepsilon}_j) = \sum_i A_{ij} \boldsymbol{\varepsilon}_i$$

Taking scalar products of both sides:

$$\begin{aligned}
\langle \boldsymbol{\varepsilon}_k, f(\boldsymbol{\varepsilon}_j) \rangle &= \left\langle \boldsymbol{\varepsilon}_k, \sum_i A_{ij} \boldsymbol{\varepsilon}_i \right\rangle \\
&= \sum_i A_{ij} \delta_{ki} \\
&= A_{kj}
\end{aligned}$$

3.3 Orthogonal Complement

3.3.1 Definition

Consider a space V with subspace W. We define the orthogonal complement to W as follows:

$$W^\perp := \{\mathbf{v} \in V \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \quad \forall \mathbf{w} \in W\} \tag{3.2}$$

Which in words means the set of vectors in V that are orthogonal to all vectors in the subspace W.

3.3.2 Properties

Claim 29. *$W^\perp$ is a subspace of V*

Proof.

By definition $W^\perp \subset V$. Now we check presence of the zero vector and closedness:

$$\langle \mathbf{w}, \mathbf{0} \rangle = 0$$

So $\mathbf{0} \in W^\perp$. Now let $\mathbf{x}, \mathbf{y} \in W^\perp$

$$\langle \mathbf{w}, \alpha\mathbf{x} + \beta\mathbf{y} \rangle = \alpha \langle \mathbf{w}, \mathbf{x} \rangle + \beta \langle \mathbf{w}, \mathbf{y} \rangle = 0$$

□

Claim 30. $W \cap W^{\perp} = \{\mathbf{0}\}$

Proof. Let $\mathbf{x} \in W \cap W^{\perp}$. $\langle \mathbf{w}, \mathbf{x} \rangle = 0$ holds for all $\mathbf{w} \in W$ including $\mathbf{x}$, so $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ which has the only solution $\mathbf{x} = \mathbf{0}$. □

Claim 31. $W \cup W^{\perp} = V$ *and hence dim(W)+dim($W^{\perp}$) = dim(V).*

Proof.

Let $\mathbf{w}_1 \ldots \mathbf{w}_r$ form an orthonormal basis for W so $\dim(W) = r$. Now complete this with $\mathbf{w}_{r+1} \ldots \mathbf{w}_n$ so $\dim(V) = n$ with $\mathbf{w}_i$ being an orthonormal basis for V. Consider $\mathbf{u} \in W^{\perp}$. As $W^{\perp} \subset V$ we may write $\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{w}_i$. Now we'll use the scalar product to find the coordinates. For some $j = 1 \ldots r$:

$$\begin{aligned}
\langle \mathbf{w}_j, \mathbf{u} \rangle &= \left\langle \mathbf{w}_j, \sum_{i=1}^{n} \alpha_i \mathbf{w}_i \right\rangle \\
&= \sum_{i=1}^{n} \alpha_i \langle \mathbf{w}_j, \mathbf{w}_i \rangle \\
&= \alpha_j = 0
\end{aligned}$$

As all vectors in $W^{\perp}$ are orthogonal to all in W and $\mathbf{w}_j \in W$. Hence we can drop the basis vectors $\mathbf{w}_1 \ldots \mathbf{w}_r$:

$$\mathbf{u} = \sum_{i=r+1}^{n} \alpha_i \mathbf{w}_i$$

And as $\mathbf{u}$ is an arbitrary vector in $W^{\perp}$, $\mathbf{w}_{r+1} \ldots \mathbf{w}_n$ span $W^{\perp}$. As they are all mutually orthogonal they are linearly independent and thus $\mathbf{w}_{r+1} \ldots \mathbf{w}_n$ forms a basis for $W^{\perp}$. Taking both the basis sets for W and $W^{\perp}$ together yields the basis set for V, so $W \cup W^{\perp} = V$. Also, by counting vectors in each basis set we arrive at $\dim(W)+\dim(W^{\perp}) = \dim(V)$.

□

3.4 Adjoint Maps

3.4.1 Definition

For a map $f : V \to V$, the adjoint map $f^{\dagger} : V \to V$ satisfies the following property for $\mathbf{v}, \mathbf{w} \in V$:

$$\langle \mathbf{v}, f(\mathbf{w}) \rangle = \langle f^{\dagger}(\mathbf{v}), \mathbf{w} \rangle. \tag{3.3}$$

Properties

Claim 32. $f^{\dagger}$ *is unique.*

Proof.

Suppose for $\mathbf{w} \neq 0$

$$\begin{aligned}
\langle f^{\dagger}(\mathbf{v}), \mathbf{w} \rangle &= \langle g^{\dagger}(\mathbf{v}), \mathbf{w} \rangle \\
(\langle \mathbf{w}, f^{\dagger}(\mathbf{v}) \rangle - \langle \mathbf{w}, g^{\dagger}(\mathbf{v}) \rangle)^* &= 0 \\
\langle \mathbf{w}, f^{\dagger}(\mathbf{v}) - g^{\dagger}(\mathbf{v}) \rangle^* &= 0
\end{aligned}$$

As $\mathbf{w} \neq 0$ we have $f^{\dagger}(\mathbf{v}) - g^{\dagger}(\mathbf{v}) = 0$ from which the claim follows.

□

Claim 33. $(f^\dagger)^\dagger = f$

Proof.

$$\langle \mathbf{v}, f(\mathbf{w})\rangle = \langle f^\dagger(\mathbf{v}), \mathbf{w}\rangle = \langle \mathbf{v}, (f^\dagger)^\dagger(\mathbf{w})\rangle$$

And due to the previously proved uniqueness of functions applied within the scalar product, the claim holds.

□

Claim 34. $(f+g)^\dagger = f^\dagger + g^\dagger$

Proof.

$$\begin{aligned}\langle \mathbf{v}, (f+g)(\mathbf{w})\rangle &= \langle \mathbf{v}, f(\mathbf{w})\rangle + \langle \mathbf{v}, g(\mathbf{w})\rangle \\ &= \langle f^\dagger(\mathbf{v}), \mathbf{w}\rangle + \langle g^\dagger(\mathbf{v}), \mathbf{w}\rangle \\ &= \langle \mathbf{w}, f^\dagger(\mathbf{v})\rangle^* + \langle \mathbf{w}, g^\dagger(\mathbf{v})\rangle^* \\ &= \langle f^\dagger(\mathbf{v}) + g^\dagger(\mathbf{v}), \mathbf{w}\rangle = \langle (f^\dagger + g^\dagger)(\mathbf{v}), \mathbf{w}\rangle\end{aligned}$$

But $\langle \mathbf{v}, (f+g)(\mathbf{w})\rangle$ also equals $\langle (f+g)^\dagger \mathbf{v}, \mathbf{w})\rangle$ so again by uniqueness the claim is proved.

□

Claim 35. $(\alpha f)^\dagger = \alpha^* f^\dagger$

Proof.

$$\langle (\alpha f)^\dagger(\mathbf{v}), \mathbf{w}\rangle = \langle \mathbf{v}, \alpha f(\mathbf{w})\rangle \tag{3.4}$$
$$= \alpha\langle \mathbf{v}, f(\mathbf{w})\rangle \tag{3.5}$$
$$= \alpha\langle f^\dagger(\mathbf{v}), \mathbf{w}\rangle \tag{3.6}$$
$$= \langle \alpha^* f^\dagger(\mathbf{v}), \mathbf{w}\rangle \tag{3.7}$$

And again by uniqueness the claim is proved.

(3.8)

□

Claim 36. $(f \circ g)^\dagger = g^\dagger \circ f^\dagger$

Proof.

$$\begin{aligned}\langle (f\circ g)^\dagger(\mathbf{v}), \mathbf{w}\rangle &= \langle \mathbf{v}, (f\circ g)(\mathbf{w})\rangle \\ &= \langle f^\dagger(\mathbf{v}), g(\mathbf{w}) \\ &= \langle g^\dagger \circ f^\dagger(\mathbf{v}, \mathbf{w})\rangle\end{aligned}$$

And again by uniqueness the claim is proved.

□

Claim 37. $(f^{-1})^\dagger = (f^\dagger)^{-1}$

Proof.

Clearly $\mathrm{Id}^\dagger = \mathrm{Id}$. So:

$$\begin{aligned}\mathrm{Id}^\dagger &= \mathrm{Id} \\ (f^{-1}\circ f)^\dagger &= \mathrm{Id} \\ f^\dagger \circ (f^{-1})^\dagger &= \mathrm{Id} \\ (f^{-1})^\dagger &= (f^\dagger)^{-1}\end{aligned}$$

□

3.4.2 Hermitian Matrices

It would be useful to find an expression for the elements of the matrix corresponding to the adjoint of a particular map. Let A represent $f : V \to V$ and B represent $f^\dagger$. We have the elements of A and B:

$$\begin{aligned}
A_{ij} &= \langle \boldsymbol{\varepsilon}_i, f(\boldsymbol{\varepsilon_j}) \rangle \\
B_{ij} &= \langle \boldsymbol{\varepsilon}_i, f^\dagger(\boldsymbol{\varepsilon_j}) \rangle \\
&= \langle f(\boldsymbol{\varepsilon}_i), \boldsymbol{\varepsilon}_j \rangle \\
&= \langle \boldsymbol{\varepsilon}_i, f(\boldsymbol{\varepsilon}_j) \rangle^* \\
&= A^*_{ji}
\end{aligned}$$

So B is the conjugate transpose of A.

3.5 More Types of Map

3.5.1 Self-Adjoint (Hermitian)

Defined by $f = f^\dagger$. We have:

$$\langle \mathbf{v}, f(\mathbf{w}) \rangle = \langle f(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{w}, f(\mathbf{v}) \rangle^*$$

3.5.2 Unitary

Unitary maps preserve scalar products. For a unitary map $f : V \to V$:

$$\langle f(\mathbf{v}), f(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \tag{3.9}$$

Claim 38. *Unitary maps leave magnitudes unchanged.*

Proof.

$$\begin{aligned}
|f(\mathbf{v})| &= \sqrt{\langle f(\mathbf{v}), f(\mathbf{v}) \rangle} \\
&= \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \\
&= |\mathbf{v}|
\end{aligned}$$

□

Claim 39. *Unitary maps leave angles unchanged.*

Proof.

Let $\theta(\mathbf{u}, \mathbf{v})$ be the angle between $\mathbf{u}$ and $\mathbf{v}$:

$$\theta(\mathbf{u}, \mathbf{v}) = \cos^{-1}\left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{|\mathbf{u}||\mathbf{v}|}\right)$$

By the previous claim, magnitudes are unchanged by unitary maps:

$$\begin{aligned}
\theta(\mathbf{u}, \mathbf{v}) &= \cos^{-1}\left(\frac{\langle f(\mathbf{u}), f(\mathbf{v}) \rangle}{|f(\mathbf{u})||f(\mathbf{v})|}\right) \\
&= \theta(f(\mathbf{u}), f(\mathbf{v}))
\end{aligned}$$

□

Claim 40. *Unitary maps are defined by $f^\dagger \circ f = Id$*

Proof.

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \\ &= \langle \mathbf{u}, f^\dagger \circ f(\mathbf{v}) \rangle\end{aligned}$$

And by function application uniqueness, $f^\dagger \circ f(\mathbf{v}) = \mathbf{v}$. Also, this implies $f^{-1} = f^\dagger$.

□

Claim 41. *$f^\dagger$ is unitary if f is unitary.*

Proof.

Let $\mathbf{v}' = f(\mathbf{v})$ and $\mathbf{w}' = f(\mathbf{w})$.

$$\begin{aligned}\langle \mathbf{v}', \mathbf{w}' \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle \\ &= \langle f^{-1}(\mathbf{v}'), f^{-1}(\mathbf{w}') \rangle\end{aligned}$$

□

3.6 Dual Vector Spaces

3.6.1 Definition

It is not difficult to show that the set of all linear maps between two vector spaces forms a vector space itself. The dual vector space V^* of a vector space V over a field F is defined as the set of all linear maps $\varphi : V \to F$. In other words, the set of linear maps that when applied to elements of V yield elements in F, if we consider F to be a one-dimensional vector space. The elements of V^* are called linear functionals.

Zero Functional

The zero functional is the element of V^* that doesn't change a functional when it is added to said functional. In other words, $\varphi + \varphi^0 = \varphi$. As these are maps, we can apply them to an arbitrary vector $\mathbf{v} \in V$, such that:

$$\begin{aligned}(\varphi + \varphi^0)(\mathbf{v}) &= \varphi(\mathbf{v}) \\ \varphi(\mathbf{v}) + \varphi^0(\mathbf{v}) &= \varphi(\mathbf{v}) \\ \varphi^0(\mathbf{v}) &= 0\end{aligned}$$

As $\varphi(\mathbf{v})$ is a scalar, we can subtract it off both sides and know what we're getting. So the zero functional is the functional which maps to the scalar 0.

3.6.2 Dual Basis

Claim 42. *For a basis $\boldsymbol{v}_1 \dots \boldsymbol{v}_n$ of V there exists a dual basis $d_1 \dots d_n$ of V^* such that $d_i(\boldsymbol{v}_j) = \delta_{ij}$.*

Proof.

As always, we must check this supposed basis set is linearly independent and spans V^*. Beginning with the former:

$$\sum_i \alpha_i d_i = \varphi^0$$

where φ^0 is the zero functional. We can apply both sides to $\mathbf{v}_j$ which should yield the same result:

$$\begin{aligned}
\sum_i \alpha_i d_i(\mathbf{v}_j) &= \varphi^0(\mathbf{v}_j) \\
\sum_i \alpha_i d_i(\mathbf{v}_j) &= 0 \\
\sum_i \alpha_i \delta_{ij} &= 0 \\
\alpha_j &= 0
\end{aligned}$$

So as $j = 1 \dots n$ and is arbitrary, we have satisfied the condition for linear independence. Now we need to show that the dual basis spans V^*. In other words, we need to show for an arbitrary functional $\varphi \in V^*$ we have $\varphi = \sum_i \alpha_i d_i$.

Let us, as an aside, suppose this is true. This will let us find out what α_i should be, and then let us make a guess at the form of φ. We would have

$$\begin{aligned}
\varphi(\mathbf{v}_j) &= \sum_i \alpha_i d_i(\mathbf{v}_j) \\
\varphi(\mathbf{v}_j) &= \alpha_j
\end{aligned}$$

So suppose we have a functional $\phi = \sum_i \varphi(\mathbf{v}_i) d_i$. We now need to show $\phi = \varphi$. This can be done by showing that, for an arbitrary vector $\mathbf{v} \in V$, $\phi(\mathbf{v}) = \varphi(\mathbf{v})$. But as $\mathbf{v}$ can be expanded as a linear combination of basis vectors of V:

$$\begin{aligned}
\phi(\mathbf{v}) &= \varphi(\mathbf{v}) \\
\phi\left(\sum_i \beta_i \mathbf{v}_i\right) &= \varphi\left(\sum_i \beta_i \mathbf{v}_i\right) \\
\sum_i \beta_i \phi(\mathbf{v}_i) &= \sum_i \beta_i \varphi(\mathbf{v}_i)
\end{aligned}$$

So it is sufficient to show that $\phi(\mathbf{v}_i) = \varphi(\mathbf{v}_i)$ for all basis vectors $\mathbf{v}_i$ within V's basis set. Lets find $\phi(\mathbf{v}_i)$:

$$\begin{aligned}
\phi(\mathbf{v}_j) &= \sum_i \varphi(\mathbf{v}_i) d_i(\mathbf{v}_j) \\
&= \sum_i \varphi(\mathbf{v}_i) \delta_{ij} \\
&= \varphi(\mathbf{v}_j)
\end{aligned}$$

Hence we have $\phi = \varphi = \sum_i \varphi(\mathbf{v}_i) d_i$ and thus the proposed dual basis does indeed span V^* as an arbitrary functional can be written as a linear combination of elements of the dual basis set. Therefore, the proposed dual basis is a basis for V^* as it satisfies both criteria for a basis set.

□

In physics literature an alternative notation is used. See the section on dual spaces in Andre Lukas' notes for a full explanation. For the purposes of the first year linear algebra course, it is more helpful to understand this concept with notation we have already become comfortable with and then introduce new notation when it is needed in future years.

3.6.3 Application to Bilinear Forms

Non-Degenerate Symmetric Bilinear Forms

We say a symmetric bilinear form $\langle \cdot, \cdot \rangle$ on a real vector space V is non-degenerate if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in V$ implies $\mathbf{v} = 0$. In other words, there are no vectors that are orthogonal to all other vectors with respect to the orthogonality defined by the bilinear form in question. To form an intuitive picture, consider $V = \mathbb{R}^3$ and take the bilinear form to be the real scalar product. Clearly it is not possible to find a vector $\mathbf{v}$ within this space perpendicular to all others as one could always scale $\mathbf{v}$ to $\mathbf{v}' = \alpha\mathbf{v}$ which would not be perpendicular to $\mathbf{v}$. Hence, the real scalar product is non-degenerate.

Claim 43. *Suppose we have a real vector space V equipped with a symmetric bilinear form D $\langle\cdot,\cdot\rangle$. If we define the maps $D_1, D_2 : V \to V^*$ with $D_1(\boldsymbol{v}) = \langle\boldsymbol{v},\cdot\rangle$, $D_2(\boldsymbol{v}) = \langle\cdot,\boldsymbol{v}\rangle$, we have D non degenerate $\iff$ D_1 and D_2 are isomorphisms.*

Proof.

$(\Rightarrow)$

$\langle\cdot,\cdot\rangle$ is bilinear, i.e. linear in each argument so D_1 and D_2 are linear. Therefore, we just need to show they are bijective. Firstly, by definition $\ker(D_1) = \{\mathbf{v} \in V \mid \langle\mathbf{v},\cdot\rangle = \varphi^0\}$. For a particular choice of $\mathbf{v}$, say $\mathbf{x} \in \ker(D_1)$, we have $D_1(\mathbf{x}) = \varphi^0$ and therefore

$$D_1(\mathbf{x})(\mathbf{y}) = \varphi^0(\mathbf{y})$$

for some arbitrary $\mathbf{y} \in V$. By the definition of the zero functional, we have

$$D_1(\mathbf{x})(\mathbf{y}) = 0.$$

But by the non-degeneracy condition we have that this imples $\mathbf{x} = \mathbf{0}$ as $\mathbf{y}$ was arbitrary. This means that the kernel can only contain $\mathbf{0}$. We have seen that this implies injectivity by claim 18. As a further consequence, $\dim(\ker(D_1)) = 0$, and thus by the dimension theorem $\dim(\mathrm{Im}(D_1)) = \dim(V)$, which implies surjectivity by claim 16. Hence, D_1 is bijective. Almost identical arguments can be applied to D_2 to show that it is also bijective.

$(\Leftarrow)$

We have D_1 is bijective. Hence, it is injective, and $\ker(D_1) = \{\mathbf{0}\}$. Hence, if we have

$$D_1(\mathbf{x}) = \varphi^0 \implies \mathbf{x} = \{\mathbf{0}\}$$

By the definition of the kernel. Suppose the equality holds, and thus $\mathbf{x} = \mathbf{0}$. We can apply both sides to an arbitrary $\mathbf{y} \in V$:

$$D_1(\mathbf{x})(\mathbf{y}) = \varphi^0(\mathbf{y})$$
$$\langle\mathbf{x},\mathbf{y}\rangle = 0$$

Which still implies $\mathbf{x} = \mathbf{0}$ as we've done the same operation to both sides. This is the very definition of non-degeneracy as $\mathbf{y}$ is arbitrary so it must hold for all $\mathbf{y} \in V$.

□

Chapter 4

Matrices

4.1 Unitary Matrices

4.1.1 Properties

Claim 44. *Unitary matrices have orthonormal columns.*

Proof.

Consider a unitary matrix written with respect to an orthonormal basis $\boldsymbol{\varepsilon}_i$. Recall we can define unitary maps by $f^\dagger \circ f = \mathrm{Id}$. If we use the matrix A to represent the map f, we have $f^\dagger = A^\dagger$ where the dagger represents taking the Hermitian conjugate of the matrix, i.e. the conjugate transpose. Hence:

$$\begin{aligned} A^\dagger A &= I \\ \sum_k (A^\dagger)_{ik} A_{kj} &= \delta_{ij} \\ \sum_k (A^*)_{ki} A_{kj} &= \delta_{ij} \\ \langle \mathbf{A}^i, \mathbf{A}^j \rangle &= \delta_{ij} \end{aligned}$$

by the definition of the scalar product of vectors with respect to an orthonormal basis. Here, I is the identity matrix and $\mathbf{A}^i$ is the ith column of the matrix A taken as a column vector.

□

4.2 Rank

4.2.1 Definition

Rank was previously defined as the dimension of the image of a map. Further, we showed that the columns of a matrix are the images of the domain's basis vectors. Also, we can show that these columns span $\mathrm{Im}(f)$. Considering an arbitrary vector in $\mathrm{Im}(f)$ to be $f(\mathbf{w})$, where $\mathbf{w}$ is some vector within the domain with basis vectors $\mathbf{v}_i$.

$$\begin{aligned} f(\mathbf{w}) &= f(\sum_i \alpha_i \mathbf{v}_i) \\ &= \sum_i \alpha_i f(\mathbf{v}_i) \end{aligned}$$

So clearly an arbitrary vector in the image can be reached by linear combinations of the images of the basis vectors - but these are precisely the matrix columns. Hence, linearly independent columns form a basis for the image of a matrix. As rank is the dimension of the image, and linearly independent matrix columns form a basis for the image, we just need to count the number of linearly independent columns to find the rank.

4.2.2 Row Rank = Column Rank

It can be proved that the row rank, i.e. the number of linearly independent row vectors in a matrix, is equal to column rank, which is the 'traditional' rank - number of linearly independent column vectors.

Claim 45. *Row rank = column rank*

Proof.

Take an arbitrary mxn matrix A with a column rank c. Hence, there are c linearly independent columns. Let it have a row rank r. Define the intersection matrix B such that each element of B is both in a linearly independent row and column of B. To be more specific, define the following ordered sets:

$$\mathcal{C} = \{\text{column numbers of the linearly independent columns of } A\text{, ascending order}\}$$
$$\mathcal{R} = \{\text{row numbers of the linearly independent rows of } A\text{, ascending order}\}$$

And let the ith element of $\mathcal{C}$ or $\mathcal{R}$ be $\mathcal{C}_i, \mathcal{R}_i$ respectively. For example, consider the following example where membership of the set of linearly independent rows/columns is denoted by $\checkmark \backslash \times$

$$\begin{array}{c c} & \begin{array}{cccc} \checkmark & \times & \checkmark & \checkmark \end{array} \\ \begin{array}{c} \checkmark \\ \times \\ \checkmark \end{array} & \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{pmatrix} \end{array}$$

Which would yield the following matrix for B:

$$B = \begin{pmatrix} A_{11} & A_{13} & A_{14} \\ A_{31} & A_{33} & A_{34} \end{pmatrix}$$

Now as shown before, matrices are maps between column vector spaces. Suppose, we use B to represent a map $b : X \to Y$. In this case, B is an $r \times c$ matrix which corresponds to a map from F^c to F^r where F is the field of the underlying vector spaces. We can see a few consequences:

1. It's a requirement of the coordinate maps associated with matrices that the dimension of the domain column vector space and domain (actual) vector space are the same, and likewise for the codomain. This tells us $\dim(X) = c$ and $\dim(Y) = r$.
2. By constructing B as we did we guaranteed that all rows and columns were linearly independent. Hence, $\dim(\text{Im}(g)) = c$ as there are c linearly independent columns
3. From the dimension theorem we can easily show $\dim(\ker(g)) = 0$ and therefore $\ker(g) = \{\mathbf{0}\}$.
4. As $\text{Im}(g)$ is a subspace of Y we have $c \leq r$

We would like to show $\dim(\text{Im}(g)) = \dim(Y)$. We can use claim 31 which specifies the relationships of the orthogonal complement of a subspace with the whole space. We have also shown that the image of a matrix is the span of its columns. We can therefore ask the question "are there any vectors within Y that are orthogonal to all vectors in $\text{Im}(g)$?" Suppose there is one, say $\mathbf{y} \in Y$.

$$\langle \mathbf{y}, \mathbf{p} \rangle = 0$$

for an arbitrary $\mathbf{p} \in \text{Im}(g)$. As shown before the columns of B will span the image, so we can find an expression for $\mathbf{p}$:

$$\mathbf{p} = \sum_{j=1}^{c} \gamma_j \mathbf{B}^j$$

Suppose without loss of generality A, and therefore B, are written with respect to an orthonormal basis of Y $\mathbf{w}_1 \dots \mathbf{w}_r$.

$$\mathbf{B}^j = \sum_{k=1}^{r} B_{kj}\mathbf{w}_k$$
$$\mathbf{p} = \sum_{j=1}^{c}\sum_{k=1}^{r} \gamma_j B_{kj}\mathbf{w}_k$$
$$\mathbf{y} = \sum_{i=1}^{r} \alpha_i \mathbf{w}_i$$

Using our aforementioned orthogonality statement:

$$0 = \left\langle \sum_{i=1}^{r} \alpha_i \mathbf{w}_i, \sum_{j=1}^{c}\sum_{k=1}^{r} \gamma_j B_{kj}\mathbf{w}_k \right\rangle$$
$$0 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{r} \alpha_i^* \gamma_j B_{kj} \langle \mathbf{w}_i, \mathbf{w}_k \rangle$$
$$0 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{r} \alpha_i^* \gamma_j B_{kj} \delta_{ik}$$
$$0 = \sum_{i=1}^{r}\sum_{j=1}^{c} \alpha_i^* \gamma_j B_{ij}$$
$$0 = \sum_{i=1}^{r} \alpha_i^* \left(\sum_{j=1}^{c} \gamma_j B_{ij} \right)$$
$$0 = \sum_{i=1}^{r} \alpha_i^* (B(\boldsymbol{\gamma}))_i$$

Now notice $B(\boldsymbol{\gamma})$, the column vector with entries that are the arbitrary coefficients of $\mathbf{p}$ with respect to B's columns, is never $\mathbf{0}$. This is because if it was, $\boldsymbol{\gamma}$ would be in the kernel of g, but we showed that the kernel only contains the zero vector. Therefore, as $\boldsymbol{\gamma}$ is arbitrary, α_i^* is 0 for all i, which implies $\alpha = 0$ and thus that $\mathbf{y} = \mathbf{0}$. Hence, the only vector orthogonal to the whole image is the zero vector, so from claim 31 we have $\dim(\mathrm{Im}(g)) = \dim(Y)$. In other words, $c = r$.

□

4.3 Elementary Operations

There are a number of operations we can do to matrices that leave ranks unchanged. Note that they apply equally to rows, but columns are used in this example. These are:

1. Swap two columns
2. Add two columns
3. Multiply a column by a scalar.

It should be clear why 1 leaves rank unchanged, as the order of columns doesn't affect how many of them are linearly independent. Further, 2 and 3 amount to linearly combining columns, which will preserve the columns that are linearly independent. There is an algorithm for converting a matrix into a form through which the rank can be easily read off, known as 'upper echelon form', but it essentially boils down to fiddling around with operations until there are no elements below the lead diagonal. I encourage you to find your own algorithm for achieving this[1].

[1]This choice has three motivations. Firstly, finding your own methods helps you remember things. Second, the algorithm is difficult to convey in words without an example. Lastly, including an example would involve formatting more matrices than I have the patience for.

4.3.1 Inverting Matrices

It happens that these elementary operations can be used on a matrix by multiplying with a particular set of invertible[2] elementary operation matrices, which I will call P_i and point you in the direction of Andre Lukas' notes to find their specific definitions. This fact has the property that if a matrix A can be reduced to the identity matrix I through elementary operations, it is invertible:

$$P_i \dots P_j A = I$$
$$P_i \dots P_j I = A^{-1}$$

By applying A^{-1} to both sides. Hence we can find the inverse of an invertible matrix by carrying out a specific sequence of elementary operations on it until it is I, then carrying out the same sequence of operations in the same order on I itself.

4.4 Systems of Linear Equations

4.4.1 In Terms of Linear Maps

It should be clear a system of linear equations can be rewritten as a matrix problem. For example, say we have the following system:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

which is equivalent to

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

So as this is a matrix problem, it is equivalently a linear map problem, $f(\mathbf{x}) = \mathbf{b}$. The crux of the problem is finding $\mathbf{x}$, the value of which will solve the equation encompassing all specific solutions. Suppose we manage to guess or otherwise find a single specific solution $\mathbf{x}_0$ that solves the equation[3]. We have the following situation:

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{b} \\ f(\mathbf{x}_0) &= \mathbf{b} \\ f(\mathbf{x}) - f(\mathbf{x}_0) &= \mathbf{0} \\ f(\mathbf{x} - \mathbf{x}_0) &= \mathbf{0} \end{aligned}$$

Now, lets consider an alternative problem, the homogeneous case $f(\mathbf{x}_h) = \mathbf{0}$. Clearly this is solved by the entire kernel space as any vector $\mathbf{x}_h \in \ker(f)$ will solve the system. By comparing this equation with the previous one, we can clearly see

$$\begin{aligned} \mathbf{x} - \mathbf{x}_0 &= \mathbf{x}_h \\ \mathbf{x} &= \mathbf{x}_h + \mathbf{x}_0 \end{aligned}$$

This should seem familiar: the solution to this inhomogeneous system is equal to the sum of the solution of the homogeneous system and a specific solution. Perhaps you'll see a relation to solving ODEs[4] if I write this statement in a slightly different way:

$$x_{\text{General}} = x_{\text{CF}} + x_{\text{PI}}$$

Just to reiterate, the solution to this system is not a vector, but instead a set of vectors. In particular, the solution set is

$$X = \ker(f) + \mathbf{x}_0$$

where $\mathbf{x} \in X$ solves the system.

[2]There exists, for each operation, an inverse one - i.e. swap the columns back, etc.
[3]As a certain mechanics lecturer says, the easiest way to solve a problem is to already know the solution.
[4]The reason for this similarity will become clearer towards the end of your first year ODEs course.

4.4.2 Structure of Solutions

It is clear that $\mathbf{x}_0$ doesn't lie within the kernel space. This is because if it did, $f(\mathbf{x}_0)$ would equal $\mathbf{0}$ - but we know it instead equals $\mathbf{b}$. Hence $\mathbf{x}_0$ must be linearly independent from all vectors $\in \ker(f)$. It's useful to know the number of free parameters of the solutions. This is a way of measuring the number of ways of varying vectors within the solution set such that the result still solves the system. Any vector within the solution set can be constructed by

1. linearly combining vectors within the kernel space
2. adding on the constant vector $\mathbf{x}_0$.

As the latter step doesn't involve any 'degrees of freedom' to vary, the number of ways of varying the vectors within the set depends on the number of ways of varying vectors within the kernel. In particular, this number of free parameters is equal to the dimension of the kernel, k; this is because there are k different 'directions[5]' that a vector could be scaled along to create a new vector that stays within the kernel.

4.4.3 Matrix Rank Considerations

Suppose we have an $m \times n$ matrix A, such that $A : F^n \to F^m$ with $A\mathbf{x} = \mathbf{b}$. Let's consider all the possibilities. Note the maximum rank of this matrix is m, as that's the dimension of the codomain.

1. $\text{Rank}(A) = m$:
   From the dimension theorem the number of free parameters will be $n - m$. Further, there is always guaranteed to be a solution as the image is the whole codomain, so there is no possibility of $\mathbf{b}$ being outside the image and therefore unreachable by the action of A.
2. $\text{Rank}(A) < m$:
   In this case there may not be a solution to the system as it is possible to choose a $\mathbf{b}$ that is within the codomain but outside the image. Assuming, however, that $\mathbf{b} \in \text{Im}(A)$, we have (again by the dimension theorem) the number of free parameters as $n - \text{Rank}(A)$.
3. $n = m$ and $\text{Rank}(A) = n$:
   Now A maps from a column space to itself, and as the dimension of the domain is the same as that of the codomain, A is bijective and thus invertible. Hence, we have $\mathbf{x} = A^{-1}\mathbf{b}$. Further, as the dimension of the kernel is zero for bijective maps, we have zero free parameters.
4. $n = m$ and $\text{Rank}(A) < n$:
   In this case there may not be solutions for the same reason as in number 2. However, if there are, the number of free parameters is $n - \text{Rank}(A)$ by the dimension theorem.

4.4.4 The Augmented Matrix

Suppose we have a system $A\mathbf{x} = \mathbf{b}$, as before. Let A have rank r. We can apply an elementary operation matrix, P, to both sides. As P is invertible, we can always multiply by P^{-1} on both sides to recover the original system. Hence, $PA\mathbf{x} = P\mathbf{b}$ should have the same set of solutions as $A\mathbf{x} = \mathbf{b}$. This inspires us to combine the matrix A and the vector $\mathbf{b}$ into one object, known as an augmented matrix. This is done by adding $\mathbf{b}$ in as an extra row to the right of A to form $A' = (A|\mathbf{b})$. Now, we can apply the elementary operations on one object *and* both A and $\mathbf{b}$ at the same time. This also gives us a convenient way of checking if solutions exist.

Claim 46. *$\mathbf{b} \in Im(A) \iff Rank(A) = Rank(A')$*

Proof.

($\Leftarrow$)

Adding $\mathbf{b}$ has no effect on rank. This implies $\mathbf{b}$ is a linear combination of the columns of A, which also span the image of A. Hence, we have $\mathbf{b} \in \text{Im}(A)$.

[5] It's more abstract than this, but an intuition from $\mathbb{R}^3$ is often useful.

(⇒)

This is the exact same argument but in reverse. Left as an exercise to the reader[6]. □

Hence, we want to take A' into a form where rank can be read off, namely the upper echelon form I helpfully described earlier. Then, we can check its rank to see when solutions exist - helpful in problems where the examiner has included unknowns in the coefficients of the matrix, and find the number of free parameters. At this stage, one usually gets tired of row reduction and uses the fact the system has been simplified, combined with the knowledge of the expected number of free parameters, to solve the system with explicit calculation (i.e. algebraically) in a slightly less tedious manner. If one desires to simply read off the solutions, one can make the upper left $r \times r$ matrix into the identity matrix by further row reduction.

4.5 Determinants

4.5.1 Definintion

The determinant is a map from n vectors $\mathbf{a}_1 \dots \mathbf{a}_n \in F^n$ to a number $\det(\mathbf{a}_1 \dots \mathbf{a}_n) \in F$ with the following conditions:

1. Linear in each argument
2. Antisymmetric: swapping two arguments introduces a minus sign
3. $\det(\mathbf{e}_1 \dots \mathbf{e}_n) = 1$ where $\mathbf{e}_i$ are the standard unit vectors, in ascending order.

Further, the determinant of a matrix is defined as the determinant of its columns, taken as column vectors.
As a consequence, there can be no repeated arguments in a determinant:

$$\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots) = -\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots)$$

Which is only satisfied by $\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots) = 0$.

4.5.2 Permutations

Definition

A permutation is a bijective map from a set of n objects (we will be considering integers) to itself: $\sigma : \{1 \dots n\} \to \{1 \dots n\}$. They are operations which change the order of the n objects. Two permutations σ_1 and σ_2 can be composed to form a third permutation equivalent to applying one and then the other: $\sigma_3 = \sigma_2 \circ \sigma_1$

Transpositions

A permutation wherein two numbers are swapped with all others unaffected is known as a transposition. It can be proven[7] that all permutations can be written as compositions of this most basic type of permutation. However, there isn't a unique way of making a given permutation from composing transpositions. As composing a transposition twice is the identity permutation (it flips and then un-flips again) we can write $\tau_2 \circ \tau_1 = \tau_2 \circ \tau_1 \circ \tau_1 \circ \tau_1$ as equivalent permutations, if τ_i is a transposition. Despite this non-uniqueness, it can be shown that what *is* unique to a permutation is the sign. The sign is defined as

$$\text{sgn}(\sigma) := (-1)^k$$

where k is the number of transpositions that are required to be composed to create σ. The number of transpositions compositions may vary for a given permutation, but if that number is even for one, it's even for all. Obviously the same applies for odd numbers. Using this definition, there are two clear consequences that can be proved by simply considering the number of transpositions:

[6] A reader who must in this case be bored beyond belief. Trust me, it is literally the same argument. It has taken me longer to add this pointless footnote than to type out the proof, which just goes to show how bored *I* am.

[7] See Andre Lukas' notes

1. $\text{sgn}(\sigma_2 \circ \sigma_1) = \text{sgn}(\sigma_2)\text{sgn}(\sigma_1)$
2. $\text{sgn}(\sigma) = \text{sgn}(\sigma^{-1})$

4.5.3 General Form Derivation

By definition it is only possible to take a determinant of a square matrix. If its columns are $\in F^n$, it has n rows. The map is also of n objects, so the matrix must have n columns. Let $A_{(ij)}$ be the element in the ith row and jth column of A.

$$\begin{aligned}\det(A) &= \det(\mathbf{A}^1, \ldots, \mathbf{A}^i, \ldots, \mathbf{A}^n) \\ &= \det\left(\sum_{j_1=1}^{n} A_{(j_1 1)}\mathbf{e}_{j_1}, \ldots, \sum_{j_n=1}^{n} A_{(j_n n)}\mathbf{e}_{j_n}\right) \\ &= \sum_{j_1 \ldots j_n}^{n} \prod_{k=1}^{n} A_{(j_k k)} \det(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_n})\end{aligned}$$

When two j_i are the same, the determinant vanishes. This means the sum is only nonzero when the set $\{j_1 \ldots j_n\}$ is a permutation of the set $\{1 \ldots n\}$ as this will ensure there are no repeat arguments. Let this permutation be σ, such that $j_i = \sigma(i)$. Also, let S_n be the set of all permutations of n integers, such that $\sum_{\sigma \in S_n}$ denotes a sum for each individual permutation in the set.

$$= \sum_{\sigma \in S_n} \prod_{k=1}^{n} A_{(\sigma(k)k)} \det(\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)})$$

We want to get rid of the determinant by converting it to $\det(\mathbf{e}_1 \ldots \mathbf{e}_n) = 1$. This can be done by permuting the columns with σ^{-1} so that the ith column is moved to the $\sigma^{-1}(i)$th column. This puts the columns in ascending order. However, this will require a certain number of flips of the arguments of the determinant[8], leading to a factor of $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$ being introduced within the sums:

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{k=1}^{n} A_{(\sigma(k)k)}$$

This can be understood as the following algorithm: choose entries such that no two entries lie in the same row, and take their product, multiplying by the sign of the permutation taken. Repeat for all possible choices and then add them up.

Another way of writing this is with the Levi-Civita symbol, which will be revisited later. This symbol includes the sign of the permutation. It is defined as follows:

$$\varepsilon_{j_1 \ldots j_n} := \begin{cases} 1 & \text{if } j_1 \ldots j_n \text{ is an even permutation of } 1 \ldots n \\ -1 & \text{if } j_1 \ldots j_n \text{ is an odd permutation of } 1 \ldots n \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

$$\det(A) = \sum_{j_1 \ldots j_n} \varepsilon_{j_1 \ldots j_n} \prod_{k=1}^{n} A_{(j_k k)}$$

4.5.4 Consequences

A quick one without a rigorous proof - the determinant of a triangular matrix, i.e. one with no entries below the lead diagonal, is the product of the diagonal entries. Think again about the algorithm for the determinant as choosing entries from rows without repeating the row chosen from and taking their product, then adding up with the required signs. The first choice will always be the top left element, (row 1), which cannot be chosen from again, so the second must be the element (2,2), but now elements from neither row 1 or two can be chosen and the only other choice to prevent the determinant vanishing is (3,3), ad nauseum.

[8] p flips, where p is a number of transpositions that could be composed to create σ

Claim 47. *det(A) = det($A^\dagger$)**

Proof. Consider again the definition of the determinant $\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{k=1}^{n} A_{(\sigma(k)k)}$. We may freely rearrange elements in the product, so lets do so so that they are in order with the permutations $\sigma(k)$ going from 1 to n. This means essentially applying the inverse permutation to k. For example, take the case $n = 2$. We have

$$\begin{aligned}
\sigma_1 : \{1,2\} \to \{1,2\} \quad \mathrm{sgn}(\sigma_1 = 1) \\
\sigma_2 : \{1,2\} \to \{2,1\} \quad \mathrm{sgn}(\sigma_2 = -1) \\
\det(A) = (A_{(11)}A_{(22)}) - (A_{21}A_{12})
\end{aligned}$$

Notice how we could also organise the products so the first index of A is in ascending order in all cases:

$$\det(A) = (A_{(11)}A_{(22)}) - (A_{12}A_{21})$$

And notice that, in doing this, we have organised the second index to be in the order of the first entry being $\sigma^{-1}(1)$ and the second being $\sigma^{-1}(2)$. To be precise, the first term is organised $A_{(1\sigma_1^{-1}(1))}A_{(2\sigma_1^{-1}(2))}$ and the second $A_{(1\sigma_2^{-1}(1))}A_{(2\sigma_1^{-1}(2))}$. So in general, an equivalent expression for the determinant involving inverse permutations is

$$\begin{aligned}
\det(A) &= \sum_{\sigma^{-1} \in S_n} \mathrm{sgn}(\sigma^{-1}) \prod_{k=1}^{n} A_{(k\sigma^{-1}(k))} \\
\det(A) &= \left(\sum_{\sigma^{-1} \in S_n} \mathrm{sgn}(\sigma^{-1}) \prod_{k=1}^{n} A^\dagger_{(\sigma^{-1}(k)k)} \right)^*
\end{aligned}$$

Notice as every permutation has an inverse, we can just call σ^{-1} another permutation within S_n, say ς.

$$\begin{aligned}
\det(A) &= \left(\sum_{\varsigma \in S_n} \mathrm{sgn}(\varsigma) \prod_{k=1}^{n} A^\dagger_{(\varsigma(k)k)} \right)^* \\
&= \det(A^\dagger)^*
\end{aligned}$$

□

Claim 48. *det(AB) = det(A)det(B)*

Proof. From the definition of matrix multiplication we can see

$$\begin{aligned}
(\mathbf{AB})^i &= \sum_{j=1}^{n} B_{ji}\mathbf{A}^j \\
\det(AB) &= \det\left(\sum_{j_1=1}^{n} B_{j_1 1}\mathbf{A}^{j_1}, \ldots, \sum_{j_n=1}^{n} B_{j_n n}\mathbf{A}^{j_n} \right) \\
&= \sum_{j_1 \ldots j_n} \prod_{k=1}^{n} B_{j_k k}\det(\mathbf{A}^{j_1}, \ldots, \mathbf{A}^{j_n})
\end{aligned}$$

And again the determinant vanishes for repeat arguments, so we can implement permutations and only consider the determinant of distinct j_i column numbers. We also take the step of ordering the permutations of j_i such that they are in ascending order, which will introduce a factor of $\mathrm{sgn}(\sigma)$.

$$\begin{aligned}
&= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{k=1}^{n} B_{j_k k}\det(\mathbf{A}^1, \ldots, \mathbf{A}^n) \\
&= \det(A)\det(B)
\end{aligned}$$

□

Claim 49. *A bijective (invertible)* $\iff$ *det(A)* $\neq 0$

Proof.

$(\Rightarrow)$

$1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1})$ so clearly neither $\det(A)$ nor $\det(A^{-1})$ can be zero. Also, $\det(A^{-1})= 1/\det(A)$.

$(\Leftarrow)$

Assume A is not bijective with $\det(A) \neq 0$. In the case A is not injective, we have dim(ker(A))> 0 which by the dimension theorem leads to rank(A)$< n$. In the case where A is not surjective we have straightaway that rank(A)$< n$. So if A is not bijective, then rank(A)$< n$. This implies at least one column in A is a linear combination of the others, which when the sums and coefficients are taken out of the determinant by its linear property will leave repeat columns inside. This will make the determinant zero, which leads to a contradiction.

□

Claim 50. *Determinant is independent of basis chosen.*

Proof. Let us change the basis of A to A' such that $A' = PAP^{-1}$.

$$\det(A') = \det(P)\det(P^{-1})\det(A) = \det(A)$$

□

Claim 51. *The determinant of a unitary matrix is* ± 1

Proof.

$$\begin{aligned} A^\dagger A &= I \\ \det(A^\dagger A) &= 1 \\ \det(A^\dagger)\det(A) &= 1 \\ \det(A)^2 &= 1 \\ \det(A) &= \pm 1 \end{aligned}$$

□

Try and convince yourself that the determinant of a reflection matrix is always -1 and that of a rotation matrix 1, perhaps using the matrix $\text{diag}(1, \ldots, -1, \ldots, 1)$. This implies that unitary matrices can always be written as the composition of a reflection and a rotation matrix.

4.5.5 Laplace's Expansion of the Determinant

Let the associated matrix, denoted $\tilde{A}_{(i,j)}$, be the matrix A with the ith row and jth column replaced by zeros, except the element A_{ij} which becomes 1. We then carry out $i-1$ row swaps and $j-1$ column swaps to create the following matrix

$$B_{(i,j)} = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & A'_{(i,j)} & \\ \vdots & & \end{pmatrix}$$

Where $A'_{(i,j)}$ is A with the ith row and jth column removed. Recall that one way to think of the determinant is to take the product of elements from rows with no repeat rows, then add them up taking permutation signs into account. In this case, there is only one starting row that will not yield a zero product; the first one. So, we have $\det(B_{(i,j)}) = \det(A'_{(i,j)})$. But $\det(B_{(i,j)})$ is just $\det(\tilde{A}_{(i,j)})$ with $i+j-2$ minus signs introduced from the swaps. Hence $\det(\tilde{A}_{(i,j)}) = (-1)^{i+j}\det(A'_{(i,j)})$.
Now, let the cofactor matrix C with entries C_{ij} be defined as

$$C_{ij} := \det(\tilde{A}_{(i,j)}) = (-1)^{i+j}\det(A'_{(i,j)})$$

Claim 52. $C^T A = det(A)I$

Proof.

$$\begin{aligned}
(C^T A)_{ij} &= \sum_k C_{ki} A_{kj} \\
&= \sum_k \det(\tilde{A}_{(k,i)}) A_{kj} \\
&= \sum_k \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{e}_k, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) A_{kj} \\
&= \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \sum_k A_{kj}\mathbf{e}_k, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) \\
&= \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{A}^j, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) \\
&= \delta_{ij}\det(A) = (\det(A)I)_{ij}
\end{aligned}$$

□

Two consequences: first, $A^{-1} = (1/\det(A))C^T$ and second that we can write determinants in terms of sub-determinants, i.e. determinants of smaller matrices. This is known as Laplace's expansion of the determinant:

$$\begin{aligned}
\det(A) &= (C^T A)_{jj} \\
&= \sum_k C_{kj} A_{kj} \\
&= \sum_k A_{kj} (-1)^{k+j} \det(A'_{(k,j)})
\end{aligned}$$

4.5.6 Cramer's Rule

The determinant, rather sneakily, gives us a way of solving systems of linear equations. Suppose we can write the system as $A\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is an $n \times n$ matrix. Now, define $B_{(j)}$ as A with the jth column replaced by the vector $\mathbf{b}$. Notice:

$$\begin{aligned}
\det(B_{(j)}) &= \det(\mathbf{A}^1, \dots, \mathbf{b}, \dots, \mathbf{A}^n) \\
&= \det(\mathbf{A}^1, \dots, A\mathbf{x}, \dots, \mathbf{A}^n) \\
&= \det(\mathbf{A}^1, \dots, \sum_i x_i \mathbf{A}^i, \dots, \mathbf{A}^n) \\
&= \sum_i x_i \det(\mathbf{A}^1, \dots, \mathbf{A}^i, \dots, \mathbf{A}^n) \\
&= \sum_i x_i \delta_{ij} \det(A) \\
&= x_j \det(A) \\
\implies x_j &= \frac{\det(B_{(j)})}{\det(A)}
\end{aligned}$$

4.6 Rotation Matrices

4.6.1 2 Dimensions

We want to find the matrix that represents a rotation of a vector in 2D by a given angle. Considering real matrices only, we want a map from $F^2 \to F^2$ so we need a 2x2 matrix:

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

If a matrix rotates vectors, it should leave lengths and angles invariant, so it should leave the scalar product invariant. We can then impose the condition that R is unitary, and further that $\det(R) = 1$

$$\begin{aligned} R^T R &= \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &\implies \begin{cases} a^2 + c^2 = 1 \\ ab = -cd \\ b^2 + d^2 = 1 \\ ad - bc = 1 \end{cases} \end{aligned}$$

From which $a = d$ and $b = -c$ follows. We can guess the solution as $a = d = \cos\theta$ and $b = -c = -\sin\theta$, which satisfies all 4 equations. Hence, we can write the matrix as:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Now, we can use the formula for the angles between vectors to find the cosine of the angle of an arbitrary vector with that same vector, but rotated by R. It is trivial, if a little tedious, to show that the angle between them under the action of R is in fact θ as we would expect, confirming that R is a rotation matrix.

4.6.2 3 Dimensions

By considering the columns of matrices as images of the basis vectors, it should be clear that the following are rotation matrices around the x, y and z axes respectively:

$$\begin{aligned} R_x &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \\ R_y &= \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \\ R_z &= \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

These matrices don't commute in general.

Chapter 5

Eigenvalues and Eigenvectors

5.1 Definitions

for $f : V \to V$ $\mathbf{v} \in V$ is an eigenvector if it satisfies $f(\mathbf{v}) = \lambda\mathbf{v}$ with $\lambda \in F$ known as the corresponding eigenvalue.
Given this equation, we can define the eigenspace of a linear map f as follows. We have $\text{Id}(\mathbf{v}) = \mathbf{v}$ so from the eigenvector equation we have

$$\begin{aligned} f(\mathbf{v}) &= \lambda\text{Id}(\mathbf{v}) \\ f(\mathbf{v}) - \lambda\text{Id}(\mathbf{v}) &= \mathbf{0} \\ (f - \lambda\text{Id})(\mathbf{v}) &= \mathbf{0} \quad \text{by the linearity of the maps} \end{aligned}$$

Hence $\mathbf{v}$ is solved by the entire space $\ker(f - \lambda\text{Id})$, known as the eigenspace, $\text{Eig}_f(\lambda)$. This is the space the eigenvectors for a particular λ live in. If each λ has one associated eigenvector, then $\dim(\text{Eig}_f(\lambda)) = 1$: this is known as the non-degenerate case. If each λ has more than one associated eigenvector then $\dim(\text{Eig}_f(\lambda) > 1$, which is known as the degenerate case.

5.2 Finding Eigenvalues and Eigenvectors

We have seen that all eigenvectors for a particular eigenvaliue live in the kernel of the map $f - \lambda\text{Id}$. We know the kernel of any linear map always contains $\mathbf{0}$, but if eigenvevtors exist then the kernel contains not only the zero vector but also other vectors. By claim 18 we can see this means the map is not injective, and therefore not bijective, and hence not invertible. Then, by claim 49 we can see that $\det(f - \lambda\text{Id}) = 0$. This determinant is known as the characteristic polynomial of f.
Suppose we represent f by the matrix A. We have $\det(A - \lambda\text{Id}) = 0$. Suppose further we change the basis of A such that A becomes A':

$$\begin{aligned} \det(A' - \lambda\text{Id}) &= 0 \\ \det(PAP^{-1} - \lambda\text{Id}) &= 0 \\ \det(PAP^{-1} - \lambda P\text{Id}P^{-1}) &= 0 \\ \det(P(A - \lambda\text{Id})P^{-1}) &= 0 \\ \det(P)\det(P^{-1})\det(A - \lambda\text{Id}) &= 0 \\ \det(A - \lambda\text{Id}) &= 0 \end{aligned}$$

So the characteristic polynomial, and the resulting eigenvalues that come from it, is basis independent.

5.3 Diagonalisation

5.3.1 Linear Maps

Claim 53. *A linear map $f : V \to V$ can be diagonalised, i.e. written with a respect to a matrix A that has entries of only eigenvalues on the diagonal $\iff$ eigenvectors of f form a basis of V.*

Proof.

($\Rightarrow$)

We assume f is described by $\text{diag}(\lambda_1 \dots \lambda_n)$ with respect to a basis $\mathbf{v}_i$. By the definition of matrix columns as the images of basis vectors we have $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. Therefore the basis $\mathbf{v}_i$ satisfies the eigenvector equation and $\mathbf{v}_i$ are a basis of eigenvectors.

($\Leftarrow$)

We assume eigenvectors of f form a basis of V. This implies we can write the columns of the matrix that represents f, i.e. A, with respect to this basis of eigenvectors:

$$\mathbf{A}^j = \sum_i A_{ij} \mathbf{v}_i$$

But from the definition of matrix columns as the images of basis vectors we have

$$\mathbf{A}^j = f(\mathbf{v}_j) = \lambda_j \mathbf{v}_j$$
$$\implies \sum_i A_{ij} \mathbf{v}_i = \lambda_j \mathbf{v}_j$$

Which is clearly satisfied by $A = \text{diag}(\lambda_1 \dots \lambda_n)$

□

5.3.2 Matrices

Claim 54. *The $n \times n$ matrix A can be diagonalised, and is done so by $A_d = P^{-1}AP$ where $P = (\boldsymbol{v}_1 \dots \boldsymbol{v}_n)$ and $\boldsymbol{v}_i$ are eigenvectors of A $\iff$ the eigenvectors $\boldsymbol{v}_1 \dots \boldsymbol{v}_n$ form a basis for F^n.*

Proof.

($\Rightarrow$)

P is given as invertible. This implies P is bijective, so $\text{Rank}(P) = n$ by the dimension theorem. This means the columns of P are n linearly independent column vectors, which implies $\mathbf{v}_i$ forms a basis of F^n by claim 11. We now just need to show that the columns of P are the eigenvectors of A.

We are given that $A_d = P^{-1}AP$. Let us reframe this by letting $P^{-1} = Q$. Hence, we have $A_d = QAQ^{-1}$. This describes a change of basis from an unprimed set, which we will call $\mathbf{w}_i$, to a primed set which is the 'diagonal basis' that A_d is written with respect to. We will call this primed set $\mathbf{w}'_i$. Recall from the section on priming matrices that Q^{-1} has column vectors that are the coordinates of the primed set with respect to the unprimed one. In other words:

$$(\mathbf{Q}^{-1})^i = \mathbf{w}'_i \quad \text{(with respect to } \mathbf{w}_i\text{)}$$

But from the definition of Q^{-1} we have

$$\mathbf{w}'_i = \mathbf{P}^i = \mathbf{v}_i$$

So the primed basis set is $\mathbf{v}_i$, the set created by the columns of P. This is the basis with respect to which A is diagonal. Recall again from the section on priming matrices that $\mathbf{e}_i$ can be thought of as the ith primed basis vector written with respect to its own primed set. It should then be clear why we would be motivated to write $Q^{-1}\mathbf{e}_i = P\mathbf{e}_i = \mathbf{v}_i$. This has nothing to do with standard basis - just confusingly writing vectors with respect to their own basis sets. Applying $\mathbf{e}_i$ to the given equation we have

$$\begin{aligned} A_d\mathbf{e}_i &= \lambda_i\mathbf{e}_i \\ P^{-1}AP\mathbf{e}_i &= P^{-1}A\mathbf{v}_i \\ P^{-1}A\mathbf{v}_i &= \lambda_i\mathbf{e}_i \end{aligned}$$

Applying P to both sides:

$$\begin{aligned} A\mathbf{v}_i &= \lambda_i P\mathbf{e}_i \\ A\mathbf{v}_i &= \lambda_i\mathbf{v}_i \end{aligned}$$

So $\mathbf{v}_i$ are eigenvectors.

($\Leftarrow$)

Define the matrix $P = (\mathbf{v}_1 \ldots \mathbf{v}_n)$ with columns being eigenvectors of A. As the columns form a basis of F^n we have Rank(P)= n so the dimension of the domain and codomain are the same, and P is invertible by claim 22. Now consider the following expression:

$$\begin{aligned} AP_{ij} &= \sum_k A_{ik}P_{kj} \\ &= \sum_k A_{ik}(\mathbf{v}_j)_k \\ &= (A\mathbf{v}_j)_i \\ &= (\lambda_j\mathbf{v}_j)_i \\ &= \lambda_j P_{ij} \end{aligned}$$

Using this result in the following expression:

$$\begin{aligned} (P^{-1}AP)_{ij} &= \sum_k (P^{-1})_{ik}(AP)_{kj} \\ &= \sum_k (P^{-1})_{ik}\lambda_j P_{kj} \\ &= \lambda_j \sum_k (P^{-1})_{ik}P_{kj} \\ &= \lambda_j (P^{-1}P)_{ij} \\ &= \lambda_j \delta_{ij} \end{aligned}$$

Which implies $P^{-1}AP = \text{diag}(\lambda_1 \ldots \lambda_n)$.

□

5.4 Consequences

A couple of useful results arise due to the basis independence of the trace and determinant. We have $\det(A) = \det(A_d) = \prod_i \lambda_i$ and $\text{Tr}(A) = \text{Tr}(A_d) = \sum_i \lambda_i$.

Claim 55. *$f : V \to V$ is self-adjoint $\implies$ all eigenvalues are real.*

Proof. Let $\mathbf{v}$ be an eigenvector of f with eigenvalue λ.

$$\begin{aligned}\lambda\langle\mathbf{v},\mathbf{v}\rangle &= \langle\mathbf{v},\lambda\mathbf{v}\rangle \\ &= \langle\mathbf{v},f(\mathbf{v})\rangle \\ &= \langle f(\mathbf{v}),\mathbf{v}\rangle \\ &= \langle\lambda\mathbf{v},\mathbf{v}\rangle \\ &= \lambda^*\langle\mathbf{v},\mathbf{v}\rangle \\ \implies \lambda &= \lambda^*\end{aligned}$$

□

Claim 56. *$f : V \to V$ is self-adjoint $\implies$ all eigenvectors are orthogonal.*

Proof.

We have $f(\mathbf{v}_1) = \lambda_1\mathbf{v}_1$ and $f(\mathbf{v}_2) = \lambda_2\mathbf{v}_2$ with $\lambda_1 \neq \lambda_2$. We can create the following expression:

$$(\lambda_1 - \lambda_2)\langle\mathbf{v}_1,\mathbf{v}_2\rangle = \lambda_1\langle\mathbf{v}_1,\mathbf{v}_2\rangle - \lambda_2\langle\mathbf{v}_1,\mathbf{v}_2\rangle$$

As eigenvalues are real:

$$\begin{aligned}&= \langle\lambda_1\mathbf{v}_1,\mathbf{v}_2\rangle - \langle\mathbf{v}_1,\lambda_2\mathbf{v}_2\rangle \\ &= \langle f(\mathbf{v}_1),\mathbf{v}_2\rangle - \langle\mathbf{v}_1,f(\mathbf{v}_2)\rangle \\ &= \langle\mathbf{v}_1,f(\mathbf{v}_2)\rangle - \langle\mathbf{v}_1,f(\mathbf{v}_2)\rangle = 0 \\ \implies \langle\mathbf{v}_1,\mathbf{v}_2\rangle &= 0\end{aligned}$$

As we have the condition $\lambda_1 \neq \lambda_2$.

□

Claim 57. *$f : V \to V$ is unitary $\implies$ all eigenvectors have unit magnitude.*

Proof.

Let $\mathbf{v}$ be the corresponding eigenvector for λ.

$$\begin{aligned}|\lambda|^2\langle\mathbf{v},\mathbf{v}\rangle &= \lambda\lambda^*\langle\mathbf{v},\mathbf{v}\rangle \\ &= \langle\lambda\mathbf{v},\lambda\mathbf{v}\rangle \\ &= \langle f(\mathbf{v}),f(\mathbf{v})\rangle \\ &= \langle\mathbf{v},\mathbf{v}\rangle\end{aligned}$$

Due to the positive definiteness property we can divide both sides by the scalar product, leaving the required result.

□

5.5 Normal Maps

A normal map f is one that satisfies $f \circ f^\dagger = f^\dagger \circ f$, or equivalently one for which the commutator of f and $f^\dagger$, $[f, f^\dagger] = f \circ f^\dagger - f^\dagger \circ f$ is zero.

Claim 58. *Let V be a vector space over $\mathbb{C}$ equipped with a hermitian scalar product with $f : V \to V$ a normal linear map. If λ is the eigenvalue of f for $\boldsymbol{v}$ then λ^* is the eigenvalue of $f^\dagger$ for $\boldsymbol{v}$.*

Proof.

Convince yourself from the properties of adjoint maps that $(f - \lambda \mathrm{Id})^\dagger = (f^\dagger - \lambda^* \mathrm{Id})$. Then we have

$$\begin{aligned} \det(f - \lambda \mathrm{Id}) = 0 &\implies \det(f - \lambda \mathrm{Id})^\dagger = 0^* \\ &\implies \det(f^\dagger - \lambda^* \mathrm{Id}) = 0 \end{aligned}$$

□

Claim 59. *Let V be a vector space over $\mathbb{C}$ equipped with a hermitian scalar product. $f : V \to V$ is normal $\iff$ there exists an orthonormal basis of eigenvectors for V.*

Proof. I couldn't find a proof of this that I understood, or didn't have an obvious loss of generality somewhere. As a result of this I don't feel confident explaining the steps of such a proof and have decided to omit this particular proof from this text for the time being. If you are interested to find one, look into the spectral theorem. □

5.6 Simultaneous Diagonalisation

The proof I will present uses the simplification that one of the matrices involved doesn't have degenerate eigenvalues. For a good proof of the general case, consult the video on simultaneous diagonalisation from the youtube channel Dr Peyam.

Claim 60. *Let A and B be diagonalisable $n \times n$ matrices. A and B commute $\iff$ A, B are simultaneously diagonalisable by the matrix P such that $A_d = P^{-1}AP$ and $B_d = P^{-1}BP$.*

Proof.

($\Leftarrow$)

$$\begin{aligned} AB - BA &= PA_dP^{-1}PB_dP^{-1} - PB_dP^{-1}PA_dP^{-1} \\ &= PA_dB_dP^{-1} - PB_dA_dP^{-1} \end{aligned}$$

Diagonal matrices commute:

$$\begin{aligned} &= PA_dB_dP^{-1} - PA_dB_dP^{-1} \\ &= 0 \\ \implies AB &= BA \end{aligned}$$

($\Rightarrow$)

Since we are given that A and B be diagonalisable matrices, the eigenvectors of A form a basis of F^n. The eigenvectors $\mathbf{v}_i$ satisfy $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. We have

$$\begin{aligned} A(B\mathbf{v}_i) &= BA\mathbf{v}_i \\ &= \lambda_i(B\mathbf{v}_i) \end{aligned}$$

Which implies $B\mathbf{v}_i$ is an eigenvector of A with the eigenvalue λ_i. We assumed that A doesn't have degenerate eigenvalues, i.e. it doesn't have multiple eigenvectors for the same eigenvalue. This implies that $B\mathbf{v}_i$ must be the same eigenvector as $\mathbf{v}_i$ as they correspond to the same eigenvalue λ_i. Hence, $B\mathbf{v}_i$ must be a simple scaling of $\mathbf{v}_i$, i.e.

$$B\mathbf{v}_i = \mu_i \mathbf{v}_i$$

for some scalar μ_i. Clearly this shows that A and B share common eigenvectors $\mathbf{v}_i$ (albeit with different eigenvalues), and as these eigenvectors form a basis A and B will both be diagonalisable by the same matrix P.

□

5.7 Applications

5.7.1 Quadratic Forms

A quadratic form is an object $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ such that A is a real, symmetric, $n \times n$ matrix. We can use diagonalisation to rewrite $q(\mathbf{x})$:

$$A = PA_dP^T$$

with the transpose being used instead of the inverse as A is a normal matrix, which implies it has an orthonormal basis of eigenvectors. This implies P is unitary and therefore that $P^{-1} = P^\dagger$. To show that all eigenvectors have real coefficients, consider that they satisfy $(A-\lambda I)\mathbf{v} = \mathbf{0}$. If the coefficients of $\mathbf{v}$ are complex, i.e. $v_k = a_k + ib_k$, we have $\sum_k (A-\lambda I)_{ik} a_k + i\sum_k (A-\lambda I)_{ik} b_k = 0 \implies b_k = 0$. Hence $P = P^*$ so $P^\dagger = P^T$. Hence:

$$\begin{aligned} q(\mathbf{x}) &= \mathbf{x}^T PA_dP^T\mathbf{x} \\ &= (P^T\mathbf{x})^T A_d P^T\mathbf{x} \\ &= \mathbf{y}^T A_d \mathbf{y} \\ &= \sum_i \lambda_i y_i^2 \end{aligned}$$

So we have chosen a new coordinate system, y_i, such that cross terms have been eliminated.

Suppose we have a quadratic form $ax_1^2 + bx_1x_2 + cx_2^2 = d$. We may rewrite this in symmetric matrix form:

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \underbrace{\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = d$$

Suppose A has eigenvalues λ_1, λ_2. After diagonalisation, we have

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = d$$

Which is the form of an ellipse or hyperbola in y_i. In particular, when $\lambda_1 = \lambda_2$, we have a circle, when $\lambda_1 \neq \lambda_2$ with the same sign as d, the equation is that of an ellipse and finally when $\lambda_1 \neq \lambda_2$ and they have opposite signs it's that of a hyperbola.
We can also find the lengths of the semi axes in the case of an ellipse. By comparison to the general form of an ellipse, it is clear that $l_i = \sqrt{\frac{d}{\lambda_i}}$. We would like to find the direction of the axes in this case. Recall that this general form of an ellipse is centered at the origin, with the l_1 axis aligned with the y_1 axis. Hence, the directions of the axes are the directions of the y_i axes. As $\mathbf{y} = P^T\mathbf{x}$, we can apply P to both sides to get $\mathbf{x} = P\mathbf{y}$; we can interpret this as the relationship that gives the coordinates with respect to the x_i axes of a vector given with respect to the y_i axes. When in the y_i system, the semi axes are along $\mathbf{e}_i$, so to find these coordinates in the x_i system we can do $P\mathbf{e}_i = \mathbf{v}_i$. Hence in the x_i system the semi axes lie along the eigenvectors of A.

5.7.2 Solving Systems of Linear Differential Equations

It is possible to write a linear nth order ODE in terms of a system of first order differential equations. For example consider the homogeneous case;

$$a_0x_0 + a_1x_0' + a_2x_0'' + \cdots + a_nx_0^{(n)} = 0$$

$$\begin{cases} x_1 = x_0' \\ x_2 = x_1' = x_0'' \\ \vdots \\ x_{n-1} = x_{n-2}' = x_0^{(n-1)} \\ a_0x_0 + a_1x_1 + \cdots + a_{n-1}x_{n-1} + a_nx_{n-1}' = 0 \end{cases}$$

Which is a system of n first order coupled linear differential equations. We can represent this in matrix form:

$$\begin{pmatrix} x_0' \\ x_1' \\ \vdots \\ x_{n-2}' \\ x_{n-1}' \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & -\frac{a_3}{a_n} & \dots & -\frac{a_{n-1}}{a_n} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{pmatrix}$$

You can clearly see how diagonalising the system would decouple it. Let's consider the 2nd order case:

$$a_0x_0 + a_1x_0' + x_0'' = 0$$

$$\begin{cases} x_1 = x_0' \\ x_1' = -a_0x_0 - a_1x_1 \end{cases}$$

$$\implies \underbrace{\begin{pmatrix} x_0' \\ x_1' \end{pmatrix}}_{\mathbf{x}'} = \underbrace{\begin{pmatrix} 0 & 1 \\ -a_0 & -a_1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}}_{\mathbf{x}}$$

Now suppose we know A has eigenvalues $\lambda_{1,2}$. We may diagonalise it:

$$\begin{aligned} \mathbf{x}' &= PA_dP^{-1}\mathbf{x} \\ P^{-1}\mathbf{x}' &= A_dP^{-1}\mathbf{x} \end{aligned}$$

Now let $P^{-1}\mathbf{x} = \mathbf{y}$. As P^{-1} is a constant with respect to the independent variable (say t), we can take it inside the derivative operator:

$$\begin{aligned} \mathbf{y}' &= A_d\mathbf{y} \\ y_i' &= \lambda_i y_i \\ \implies y_i &= c_ie^{\lambda_i t} \\ \implies \mathbf{x} = P\mathbf{y} &= \sum_i c_ie^{\lambda_i t}\mathbf{v}_i \end{aligned}$$

Which recovers the standard exponential solution of this type of differential equation. A slightly more complicated scenario is the case of repeated eigenvalues. In this case one may add in a factor of t in the right place to generate a linearly independent solution, but I'll leave that to your ODE lecturer to explain.

5.7.3 Repeated Multiplication of Matrices

Consider a diagonalisable matrix A. If we want to compute A^n, we can diagonalise A:

$$\begin{aligned} A^n &= (PA^dP^{-1})^n \\ &= PA_dP^{-1}P\dots P^{-1}PA_dP^{-1} \end{aligned}$$

The P and P^{-1} matrices that are sandwiched together cancel out, leaving n lots of A_d in the middle.

$$= PA_d^nP^{-1}$$

And as A_d is diagonal, we can just raise its entries to the nth power. Recall its entries are its eigenvalues:

$$= P\text{diag}(\lambda_i^n)P^{-1}$$

This has useful consequences in computing functions of matrices. This is because we may, in some cases, write the functions as Taylor series, and then apply our diagonalisation technique to the polynomial of matrices.

Chapter 6

Vector Identities and Geometry

6.1 Vector Identities

6.1.1 Kronecker Delta Symbol

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

This symbol has the useful property of index replacement. Using the Einstein implied summation convention hereon we have $\delta_{ij} a_i = a_j$.

6.1.2 Levi-Civita Symbol

$$\varepsilon_{j_1 \dots j_n} := \begin{cases} 1 & \text{if } j_1 \dots j_n \text{ is an even permutation of } 1 \dots n \\ -1 & \text{if } j_1 \dots j_n \text{ is an odd permutation of } 1 \dots n \\ 0 & \text{otherwise} \end{cases}$$

We will now verify a useful property of the symbol.

We propose the relationship

$$\varepsilon_{ijk}\varepsilon_{lmn} = \det \begin{pmatrix} \delta_{il} & \delta_{im} & \delta_{in} \\ \delta_{jl} & \delta_{jm} & \delta_{jn} \\ \delta_{kl} & \delta_{km} & \delta_{kn} \end{pmatrix}$$

First take the case two of ijk or two of lmn are equal. By definition this makes the LHS = 0, and will create repeat rows or columns in the determinant which makes RHS = 0. Next, consider ijk and lmn both = 123. This makes LHS = 0 and RHS = $\det(I) = 1$. Now, the crucial step is noticing that any permutation of 123 can be achieved by repeatedly swapping any of ijk or any of lmn. Notice that both sides will be antisymmetric under these swaps (as the determinant is the same as the determinant of the transpose, and swapping columns introduces a minus sign); in a sense they start equal at the permutation 123 and the same effect 'happens' to them no matter how we manipulate the permutations thereafter - keeping them equal. Hence in all cases the equality holds.

From this we can derive a number of useful facts;

1. $\varepsilon_{ijk}\varepsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}$
2. $\varepsilon_{ijk}\varepsilon_{ijm} = 2\delta_{km}$
3. $\varepsilon_{ijk}\varepsilon_{ijk} = 6$

Where 1 follows from the expansion and subsequent contraction of the determinant and 2 follows from multiplying both sides of 1 by δ_{jl}, and replacing indices. Finally, 3 follows from 2 by substitution.

6.1.3 Identities and Proofs

We can use these symbols to define the cross product $(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk} a_j b_k$. This can be shown by comparison of the standard determinant definition of the cross product with the definition of the determinant that uses ε. Notice also that the dot product may be written $\mathbf{a} \cdot \mathbf{b} = \delta_{ij} a_i b_j$.

Claim 61. *The following statements hold for arbitrary vectors;*

1. $\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}$
2. $\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) + (\boldsymbol{a} \times \boldsymbol{c})$
3. $\boldsymbol{a} \times \beta \boldsymbol{b} = \beta(\boldsymbol{a} \times \boldsymbol{b})$
4. $\boldsymbol{e}_1 \times \boldsymbol{e}_2 = \boldsymbol{e}_3 \quad \boldsymbol{e}_2 \times \boldsymbol{e}_3 = \boldsymbol{e}_1 \quad \boldsymbol{e}_3 \times \boldsymbol{e}_1 = \boldsymbol{e}_2$
5. $\boldsymbol{a} \times \boldsymbol{a} = 0$
6. $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b} \cdot \boldsymbol{a})\boldsymbol{c}$
7. $(\boldsymbol{a} \times \boldsymbol{b}) \cdot (\boldsymbol{c} \times \boldsymbol{d}) = (\boldsymbol{a} \cdot \boldsymbol{c})(\boldsymbol{b} \cdot \boldsymbol{d}) - (\boldsymbol{a} \cdot \boldsymbol{d})(\boldsymbol{b} \cdot \boldsymbol{c})$
8. $|\boldsymbol{a} \times \boldsymbol{b}|^2 = |\boldsymbol{a}|^2 |\boldsymbol{b}|^2 - (\boldsymbol{a} \cdot \boldsymbol{b})^2$

Proof.

(1)

$$\begin{aligned}(\mathbf{a} \times \mathbf{b})_i &= \varepsilon_{ijk} a_j b_k \\ &= -\varepsilon_{ikj} b_k a_j \\ &= -(\mathbf{b} \times \mathbf{a})_i\end{aligned}$$

(2)

$$\begin{aligned}(\mathbf{a} \times (\mathbf{b} + \mathbf{c}))_i &= \varepsilon_{ijk} a_i (b_j + c_j) \\ &= \varepsilon_{ijk} a_i b_j + \varepsilon_{ijk} a_i c_j \\ &= (\mathbf{a} \times \mathbf{b})_i + (\mathbf{a} \times \mathbf{c})_i\end{aligned}$$

(3)

$$\begin{aligned}(\mathbf{a} \times \beta \mathbf{b})_i &= \varepsilon_{ijk} a_j \beta b_k \\ &= \beta \varepsilon_{ijk} a_j b_k \\ &= \beta (\mathbf{a} \times \mathbf{b})_i\end{aligned}$$

(4)

$$\begin{aligned}(\mathbf{e}_i)_j &= \delta_{ij} \\ \implies (\mathbf{e}_p \times \mathbf{e}_q)_i &= \varepsilon_{ijk} (\mathbf{e}_p)_j (\mathbf{e}_q)_k \\ &= \varepsilon_{ijk} \delta_{pj} \delta_{qk} \\ &= \varepsilon_{ipq} \quad \text{from which the required results follow.}\end{aligned}$$

(5)

$$(\mathbf{a} \times \mathbf{a}) = -(\mathbf{a} \times \mathbf{a}) = 0 \quad \text{by the anticommutative property.}$$

(6)

$$\begin{aligned}(\mathbf{a} \times (\mathbf{b} \times \mathbf{c}))_i &= \varepsilon_{ijk} a_j (\mathbf{b} \times \mathbf{c})_k \\ &= \varepsilon_{ijk} a_j \varepsilon_{klm} b_l c_m \\ &= \varepsilon_{kij} \varepsilon_{klm} a_j b_l c_m \\ &= (\delta_{il} \delta_{jm} - \delta_{jl} \delta_{im}) a_j b_l c_m \\ &= a_m c_m b_i - a_l b_l c_i \\ &= ((\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c})_i\end{aligned}$$

(7)

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) &= (\mathbf{a}\times\mathbf{b})_i(\mathbf{b}\times\mathbf{d})_i \\ &= \varepsilon_{ijk}a_jb_k\varepsilon_{ilm}c_ld_m \\ &= (\delta_{jl}\delta_{km}-\delta_{kl}\delta_{jm})a_jb_kc_ld_m \\ &= a_lb_mc_ld_m - a_jb_lc_ld_j \\ &= (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{b})\end{aligned}$$

(8)

Use the result from (7) with $\mathbf{c}=\mathbf{a}$ and $\mathbf{d}=\mathbf{b}$.

□

6.2 Vector Geometry

6.2.1 Geometrical Meaning of the Cross Product

Consider the quantity $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$:

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= a_i(\mathbf{b}\times\mathbf{c})_i \\ &= \epsilon_{ijk}a_ib_jc_k \\ &= \det(\mathbf{a},\mathbf{b},\mathbf{c})\end{aligned}$$

Which will vanish if any of the vectors are equal. This implies that the cross product of any two vectors is orthogonal to both of those vectors. We can also find the magnitude of the cross product from claim 61:

$$|\mathbf{a}\times\mathbf{b}| = (|\mathbf{a}|^2|\mathbf{b}|^2 - (|\mathbf{a}||\mathbf{b}|\cos\theta)^2)^{\frac{1}{2}} = |\mathbf{a}||\mathbf{b}|\sin\theta$$

from which we can conclude that $\mathbf{a}\times\mathbf{b} = |\mathbf{a}||\mathbf{b}|\sin\theta\hat{\mathbf{n}}$ where $\hat{\mathbf{n}}$ is the unit vector normal to both $\mathbf{a}$ and $\mathbf{b}$. The direction of this vector can be found using the right hand rule or corkscrew rule.

6.2.2 Minimum Distance Between a Line and Point

Consider a line $\mathbf{r}(t)=\mathbf{a}+t\mathbf{b}$. We would like to find the minimum distance between $\mathbf{r}$ and a point $\mathbf{p}_0$. First we begin by defining the vector $\mathbf{p}:=\mathbf{a}-\mathbf{p}_0$ and the distance squared function, $D^2:=|\mathbf{r}-\mathbf{p}_0|^2$. We have

$$\begin{aligned}D^2 &= |\mathbf{p}+t\mathbf{b}|^2 \\ &= \langle\mathbf{p}+t\mathbf{b},\mathbf{p}+t\mathbf{b}\rangle \\ &= \langle\mathbf{p}+t\mathbf{b},\mathbf{p}\rangle + t\langle\mathbf{p}+t\mathbf{b},\mathbf{b}\rangle \\ &= |\mathbf{p}|^2 + 2t\langle\mathbf{b},\mathbf{p}\rangle + t^2|\mathbf{b}|^2 \quad \text{taking derivatives:} \\ 2D\frac{dD}{dt} &= 2\langle\mathbf{b},\mathbf{p}\rangle + 2t|\mathbf{b}|^2 = 0 \quad \text{for stationary point} \\ \implies t_{\min} &= -\frac{\langle\mathbf{b},\mathbf{p}\rangle}{|\mathbf{b}|^2}\end{aligned}$$

If you'd like, check second derivatives to ensure this is a minimum point. Now subbing back in:

$$\begin{aligned}D^2_{\min} &= |\mathbf{p}|^2 - \frac{2\langle\mathbf{b},\mathbf{p}\rangle^2}{|\mathbf{b}|^2} + \frac{\langle\mathbf{b},\mathbf{p}\rangle^2}{|\mathbf{b}|^2} \\ &= \frac{1}{|\mathbf{b}|^2}(|\mathbf{p}|^2|\mathbf{b}|^2 - \langle\mathbf{b},\mathbf{p}\rangle^2) \\ &= \frac{|\mathbf{p}\times\mathbf{b}|^2}{|\mathbf{b}|^2} \\ D_{\min} &= \frac{|\mathbf{p}\times\mathbf{b}|}{|\mathbf{b}|}\end{aligned}$$

6.2.3 Intersection of Line and Plane

Consider the line $\mathbf{r}_l = \mathbf{p} + t\mathbf{q}$ and the plane $\mathbf{r}_p = \mathbf{a} + \mathbf{b}t_1 + \mathbf{c}t_2$. To find the intersection set them equal:

$$\begin{aligned}\mathbf{p} + t\mathbf{q} &= \mathbf{a} + \mathbf{b}t_1 + \mathbf{c}t_2 \\ \mathbf{p} - \mathbf{a} &= \mathbf{b}t_1 + \mathbf{c}t_2 - t\mathbf{q}\end{aligned}$$

and dot both sides with $(\mathbf{b} \times \mathbf{c})$:

$$\begin{aligned}(\mathbf{p} - \mathbf{a}) \cdot (\mathbf{b} \times \mathbf{c}) &= -t\mathbf{q} \cdot (\mathbf{b} \times \mathbf{c}) \\ t &= \frac{(\mathbf{a} - \mathbf{p}) \cdot (\mathbf{b} \times \mathbf{c})}{\mathbf{q} \cdot (\mathbf{b} \times \mathbf{c})}\end{aligned}$$

6.2.4 Minimum Distance Between Lines

For $i = 1, 2$ we have the lines $\mathbf{r}_i = \mathbf{p}_i + t_i\mathbf{q}_i$. Introduce the unit vector perpendicular to both, $\mathbf{n} = \frac{\mathbf{q}_1 \times \mathbf{q}_2}{|\mathbf{q}_1 \times \mathbf{q}_2|}$. We have the distance between the lines at some value of $t_i = t_i'$ where said distance is minimised:

$$\begin{aligned}\mathbf{r}_1 - \mathbf{r}_2 &= \mathbf{p}_1 - \mathbf{p}_2 + \mathbf{q}_1 t_1' - \mathbf{q}_2 t_2' \\ &= \pm d_{\min}\mathbf{n}\end{aligned}$$

Now dot both sides with $\mathbf{n}$:

$$\begin{aligned}(\mathbf{p}_1 - \mathbf{p}_2) \cdot \mathbf{n} &= \pm d_{\min} \\ d_{\min} &= \left| \frac{(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{q}_1 \times \mathbf{q}_2)}{|\mathbf{q}_1 \times \mathbf{q}_2|} \right|\end{aligned}$$