Linear Algebra in Physics

(Summer Semester, 2006)

1 Introduction

The mathematical idea of a vector plays an important role in many areas of physics.

- Thinking about a particle traveling through space, we imagine that its speed and direction of travel can be represented by a vector $\mathbf{v}$ in 3-dimensional Euclidean space $\mathbb{R}^3$. Its path in time t might be given by a continuously varying line — perhaps with self-intersections — at each point of which we have the velocity vector $\mathbf{v}(t)$.

- A static structure such as a bridge has loads which must be calculated at various points. These are also vectors, giving the direction and magnitude of the force at those isolated points.

- In the theory of electromagnetism, Maxwell's equations deal with vector fields in 3-dimensional space which can change with time. Thus at each point of space and time, two vectors are specified, giving the electrical and the magnetic fields at that point.

- Given two different frames of reference in the theory of relativity, the transformation of the distances and times from one to the other is given by a linear mapping of vector spaces.

- In quantum mechanics, a given experiment is characterized by an abstract space of complex functions. Each function is thought of as being itself a kind of vector. So we have a vector space of functions, and the methods of linear algebra are used to analyze the experiment.

Looking at these five examples where linear algebra comes up in physics, we see that for the first three, involving "classical physics", we have vectors placed at different points in space and time. On the other hand, the fifth example is a vector space where the vectors are not to be thought of as being simple arrows in the normal, classical space of everyday life. In any case, it is clear that the theory of linear algebra is very basic to any study of physics.

But rather than thinking in terms of vectors as representing physical processes, it is best to begin these lectures by looking at things in a more mathematical, abstract way. Once we have gotten a feeling for the techniques involved, then we can apply them to the simple picture of vectors as being arrows located at different points of the classical 3-dimensional space.

2 Basic Definitions

Definition. *Let X and Y be sets. The Cartesian product $X \times Y$, of X with Y is the set of all possible pairs (x, y) such that $x \in X$ and $y \in Y$.*

Definition. *A group is a non-empty set G, together with an operation*[1]*, which is a mapping '$\cdot$': $G \times G \to G$, such that the following conditions are satisfied.*

1. *For all $a, b, c \in G$, we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$,*
2. *There exists a particular element (the "neutral" element), often called e in group theory, such that $e \cdot g = g \cdot e = g$, for all $g \in G$.*
3. *For each $g \in G$, there exists an inverse element $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.*

If, in addition, we have $a \cdot b = b \cdot a$ for all $a, b \in G$, then G is called an "Abelian" group.

Definition. *A field is a non-empty set F, having two arithmetical operations, denoted by '$+$' and '$\cdot$', that is, addition and multiplication*[2]*. Under addition, F is an Abelian group with a neutral element denoted by '0'. Furthermore, there is another element, denoted by '1', with $1 \neq 0$, such that $F \setminus \{0\}$ (that is, the set F, with the single element 0 removed) is an Abelian group, with neutral element 1, under multiplication. In addition, the distributive property holds:*

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{and} \quad (a + b) \cdot c = a \cdot c + b \cdot c,$$

for all $a, b, c \in F$.

The simplest example of a field is the set consisting of just two elements $\{0, 1\}$ with the obvious multiplication. This is the field $\mathbb{Z}/2\mathbb{Z}$. Also, as we have seen in the analysis lectures, for any prime number $p \in \mathbb{N}$, the set $\mathbb{Z}/p\mathbb{Z}$ of residues modulo p is a field.

The following theorem, which should be familiar from the analysis lectures, gives some elementary general properties of fields.

Theorem 1. *Let F be a field. Then for all $a, b \in F$, we have:*

1. $a \cdot 0 = 0 \cdot a = 0$,
2. $a \cdot (-b) = -(a \cdot b) = (-a) \cdot b$,
3. $-(-a) = a$,
4. $(a^{-1})^{-1} = a$, *if* $a \neq 0$,
5. $(-1) \cdot a = -a$,
6. $(-a) \cdot (-b) = a \cdot b$,
7. $a \cdot b = 0 \Rightarrow a = 0$ *or* $b = 0$.

[1] The operation is usually called "multiplication" in abstract group theory, but the sets we will deal with are also groups under "addition".

[2] Of course, when writing a multiplication, it is usual to simply leave the '$\cdot$' out, so that the expression $a \cdot b$ is simplified to ab.

Proof. An exercise (dealt with in the analysis lectures). □

So the theory of abstract vector spaces starts with the idea of a *field* as the underlying arithmetical system. But in physics, and in most of mathematics (at least the analysis part of it), we do not get carried away with such generalities. Instead we will usually be confining our attention to one of two very particular fields, namely either the field of real numbers $\mathbb{R}$, or else the field of complex numbers $\mathbb{C}$.

Despite this, let us adopt the usual generality in the definition of a *vector space*.

Definition. *A vector space* $\mathbf{V}$ *over a field* F *is an Abelian group — with vector addition denoted by* $\mathbf{v} + \mathbf{w}$*, for vectors* $\mathbf{v}, \mathbf{w} \in \mathbf{V}$*. The neutral element is the "zero vector"* $\mathbf{0}$*. Furthermore, there is a scalar multiplication* $F \times \mathbf{V} \to \mathbf{V}$ *satisfying (for arbitrary* $a, b \in F$ *and* $\mathbf{v}, \mathbf{w} \in \mathbf{V}$*):*

1. $a \cdot (\mathbf{v} + \mathbf{w}) = a \cdot \mathbf{v} + a \cdot \mathbf{w}$*,*
2. $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$*,*
3. $(a \cdot b) \cdot \mathbf{v} = a \cdot (b \cdot \mathbf{v})$*, and*
4. $1 \cdot \mathbf{v} = \mathbf{v}$ *for all* $\mathbf{v} \in \mathbf{V}$*.*

Examples

- Given any field F, then we can say that F is a vector space over itself. The vectors are just the elements of F. Vector addition is the addition in the field. Scalar multiplication is multiplication in the field.

- Let $\mathbb{R}^n$ be the set of n-tuples, for some $n \in \mathbb{N}$. That is, the set of ordered lists of n real numbers. One can also say that this is

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}},$$

the Cartesian product, defined recursively. Given two elements

$$(x_1, \cdots, x_n) \quad \text{and} \quad (y_1, \ldots, y_n)$$

in $\mathbb{R}^n$, then the vector sum is simply the new vector

$$(x_1 + y_1, \cdots, x_n + y_n).$$

Scalar multiplication is

$$a \cdot (x_1, \cdots, x_n) = (a \cdot x_1, \cdots, a \cdot x_n).$$

It is a trivial matter to verify that $\mathbb{R}^n$, with these operations, is a vector space over $\mathbb{R}$.

- Let $C_0([0,1],\mathbb{R})$ be the set of all continuous functions $f:[0,1]\to\mathbb{R}$. This is a vector space with vector addition

$$(f+g)(x) = f(x) + g(x),$$

for all $x \in [0,1]$, defining the new function $(f+g) \in C_0([0,1],\mathbb{R})$, for all $f, g \in C_0([0,1],\mathbb{R})$. Scalar multiplication is given by

$$(a \cdot f)(x) = a \cdot f(x)$$

for all $x \in [0,1]$.

3 Subspaces

Let $\mathbf{V}$ be a vector space over a field F and let $\mathbf{W} \subset \mathbf{V}$ be some subset. If $\mathbf{W}$ is itself a vector space over F, considered using the addition and scalar multiplication in $\mathbf{V}$, then we say that $\mathbf{W}$ is a *subspace* of $\mathbf{V}$. Analogously, a subset H of a group G, which is itself a group using the multiplication operation from G, is called a *subgroup* of G. Subfields are similarly defined.

Theorem 2. *Let $\mathbf{W} \subset \mathbf{V}$ be a subset of a vector space over the field F. Then*

$$\mathbf{W} \textit{ is a subspace of } \mathbf{V} \Leftrightarrow a \cdot \mathbf{v} + b \cdot \mathbf{w} \in \mathbf{W},$$

for all $\mathbf{v}, \mathbf{w} \in \mathbf{W}$ and $a, b \in F$.

Proof. The direction '$\Rightarrow$' is trivial.

For '$\Leftarrow$', begin by observing that $1 \cdot \mathbf{v} + 1 \cdot \mathbf{w} = \mathbf{v} + \mathbf{w} \in \mathbf{W}$, and $a \cdot \mathbf{v} + 0 \cdot \mathbf{w} = a \cdot \mathbf{v} \in \mathbf{W}$, for all $\mathbf{v}, \mathbf{w} \in \mathbf{W}$ and $a \in F$. Thus $\mathbf{W}$ is closed under vector addition and scalar multiplication.

Is $\mathbf{W}$ a group with respect to vector addition? We have $0 \cdot \mathbf{v} = \mathbf{0} \in \mathbf{W}$, for $\mathbf{v} \in \mathbf{W}$; therefore the neutral element $\mathbf{0}$ is contained in $\mathbf{W}$. For an arbitrary $\mathbf{v} \in \mathbf{W}$ we have

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

Therefore $(-1) \cdot \mathbf{v}$ is the inverse element to $\mathbf{v}$ under addition, and so we can simply write $(-1) \cdot \mathbf{v} = -\mathbf{v}$.

The other axioms for a vector space can be easily checked. $\square$

The method of this proof also shows that we have similar conditions for subsets of groups or fields to be subgroups, or subfields, respectively.

Theorem 3. *Let $H \subset G$ be a (non-empty) subset of the group G. Then H is a subgroup of $G \Leftrightarrow ab^{-1} \in H$, for all $a, b \in H$.*

Proof. The direction '$\Rightarrow$' is trivial. As for '$\Leftarrow$', let $a \in H$. Then $aa^{-1} = e \in H$. Thus the neutral element of the group multiplication is contained in H. Also $ea^{-1} = a^{-1} \in H$. Furthermore, for all $a, b \in H$, we have $a(b^{-1})^{-1} = ab \in H$. Thus H is closed under multiplication. The fact that the multiplication is associative ($a(bc) = (ab)c$, for all a, b and $c \in H$) follows since G itself is a group; thus the multiplication throughout G is associative. $\square$

Theorem 4. *Let $\mathbf{U}, \mathbf{W} \subset \mathbf{V}$ be subspaces of the vector space $\mathbf{V}$ over the field F. Then $\mathbf{U} \cap \mathbf{W}$ is also a subspace.*

Proof. Let $\mathbf{v}, \mathbf{w} \in \mathbf{U} \cap \mathbf{W}$ be arbitrary vectors in the intersection, and let $a, b \in F$ be arbitrary elements of the field F. Then, since $\mathbf{U}$ is a subspace of $\mathbf{V}$, we have $a \cdot \mathbf{v} + b \cdot \mathbf{w} \in \mathbf{U}$. This follows from theorem 2. Similarly $a \cdot \mathbf{v} + b \cdot \mathbf{w} \in \mathbf{W}$. Thus it is in the intersection, and so theorem 2 shows that $\mathbf{U} \cap \mathbf{W}$ is a subspace. $\square$

4 Linear Independence and Dimension

Definition. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbf{V}$ be finitely many vectors in the vector space $\mathbf{V}$ over the field F. We say that the vectors are* linearly dependent *if there exists an equation of the form*

$$a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0},$$

such that not all $a_i \in F$ are simply zero. If no such non-trivial equation exists, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ is said to be linearly independent.

This definition is undoubtedly the most important idea that there is in the theory of linear algebra!

Examples

- In $\mathbb{R}^2$ let $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (0, 1)$ and $\mathbf{v}_3 = (1, 1)$. Then the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly *dependent*, since we have

  $$\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_1 = \mathbf{0}.$$

  On the other hand, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly *in*dependent.

- In $C_0([0, 1], \mathbb{R})$, let $f_1 : [0, 1] \to \mathbb{R}$ be given by $f_1(x) = 1$ for all $x \in [0, 1]$. Similarly, let f_2 be given by $f_2(x) = x$, and f_3 is $f_3(x) = 1 - x$. Then the set $\{f_1, f_2, f_3\}$ is linearly dependent.

Now take some vector space $\mathbf{V}$ over a field F, and let $S \subset \mathbf{V}$ be some subset of $\mathbf{V}$. (The set S can be finite or infinite here, although we will usually be dealing with finite sets.) Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \subset S$ be some finite collection of vectors in S, and let $a_1, \ldots, a_n \in F$ be some arbitrary collection of elements of the field. Then the sum

$$a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$$

is a *linear combination* of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in S. The set of all possible linear combinations of vectors in S is denoted by *span*(S), and it is called the linear span of S. One also writes $[S]$. S is the *generating set* of $[S]$. Therefore if $[S] = \mathbf{V}$, then we say that S is a generating set for V. If S is finite, and it generates $\mathbf{V}$, then we say that the vector space $\mathbf{V}$ is *finitely generated.*

Theorem 5. *Given $S \subset \mathbf{V}$, then $[S]$ is a subspace of $\mathbf{V}$.*

Proof. A simple consequence of theorem 2. □

Examples

- For any $n \in \mathbb{N}$, let

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, \ldots, 0) \\ \mathbf{e}_2 &= (0, 1, \ldots, 0) \\ &\vdots \\ \mathbf{e}_n &= (0, 0, \ldots, 1) \end{aligned}$$

  Then $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a generating set for $\mathbb{R}^n$.

- On the other hand, the vector space $C_0([0,1], \mathbb{R})$ is clearly *not* finitely generated.[3]

So let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ be a *finite* set. From now on in these discussions, we will assume that such sets are finite unless stated otherwise.

Theorem 6. *Let $\mathbf{w} = a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n$ be some vector in $[S] \subset \mathbf{V}$, where $a_1, \ldots, a_n$ are arbitrarily given elements of the field F. We will say that this representation of $\mathbf{w}$ is* unique *if, given some other linear combination, $\mathbf{w} = b_1\mathbf{v}_1 + \cdots b_n\mathbf{v}_n$, then we must have $b_i = a_i$ for all $i = 1, \ldots, n$. Given this, then we have that the set S is linearly independent $\Leftrightarrow$ the representation of all vectors in the span of S as linear combinations of vectors in S is unique.*

Proof. '$\Leftarrow$' We certainly have $0 \cdot \mathbf{v}_1 + \cdots 0 \cdot \mathbf{v}_n = \mathbf{0}$. Since this representation of the zero vector is unique, it follows S is linearly independent.

'$\Rightarrow$' Can it be that S is linearly independent, and yet there exists a vector in the span of S which is not uniquely represented as a linear combination of the vectors in S? Assume that there exist elements $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ of the field F, where $a_j \neq b_j$, for at least one j between 1 and n, such that

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n.$$

But then

$$(a_1 - b_1)\mathbf{v}_1 + \cdots + \underbrace{(a_j - b_j)}_{\neq 0}\mathbf{v}_i + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}$$

shows that S cannot be a linearly independent set. □

[3]In general such function spaces — which play a big role in quantum field theory, and which are studied using the mathematical theory of *functional analysis* — are not finitely generated. However in this lecture, we will mostly be concerned with finitely generated vector spaces.

Definition. *Assume that $S \subset \mathbf{V}$ is a finite, linearly independent subset with $[S] = \mathbf{V}$. Then S is called a* basis *for $\mathbf{V}$.*

Lemma. *Assume that $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ is linearly dependent. Then there exists some $j \in \{1, \ldots, n\}$, and elements $a_i \in F$, for $i \neq j$, such that*

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

Proof. Since S is linearly dependent, there exists some non-trivial linear combination of the elements of S, summing to the zero vector,

$$\sum_{i=1}^{n} b_i \mathbf{v}_i = \mathbf{0},$$

such that $b_j \neq 0$, for at least one of the j. Take such a one. Then

$$b_j \mathbf{v}_j = -\sum_{i \neq j} b_i \mathbf{v}_i$$

and so

$$\mathbf{v}_j = \sum_{i \neq j} \left(-\frac{b_i}{b_j} \right) \mathbf{v}_j.$$

□

Corollary. *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ be linearly dependent, and let $\mathbf{v}_j$ be as in the lemma above. Let $S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n\}$ be S, with the element $\mathbf{v}_j$ removed. Then $[S] = [S']$.*

Theorem 7. *Assume that the vector space $\mathbf{V}$ is finitely generated. Then there exists a basis for $\mathbf{V}$.*

Proof. Since $\mathbf{V}$ is finitely generated, there exists a finite generating set. Let S be such a finite generating set which has as few elements as possible. If S were linearly dependent, then we could remove some element, as in the lemma, leaving us with a still smaller generating set for $\mathbf{V}$. This is a contradiction. Therefore S must be a basis for $\mathbf{V}$. □

Theorem 8. *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for the vector space $\mathbf{V}$, and take some arbitrary non-zero vector $\mathbf{w} \in \mathbf{V}$. Then there exists some $j \in \{1, \ldots, n\}$, such that*

$$S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n\}$$

is also a basis of $\mathbf{V}$.

Proof. Writing $\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, we see that since $\mathbf{w} \neq \mathbf{0}$, at least one $a_j \neq 0$. Taking that j, we write

$$\mathbf{v}_j = a_j^{-1}\mathbf{w} + \sum_{i \neq j} \left(-\frac{a_i}{a_j} \right) \mathbf{v}_i.$$

We now prove that $[S'] = \mathbf{V}$. For this, let $\mathbf{u} \in \mathbf{V}$ be an arbitrary vector. Since S is a basis for $\mathbf{V}$, there exists a linear combination $\mathbf{u} = b_1\mathbf{v}_1 + \cdots b_n\mathbf{v}_n$. Then we have

$$\begin{aligned}
\mathbf{u} &= b_j\mathbf{v}_j + \sum_{i\neq j} b_i\mathbf{v}_i \\
&= b_j\left(a_j^{-1}\mathbf{w} + \sum_{i\neq j}\left(-\frac{a_i}{a_j}\right)\mathbf{v}_i\right) + \sum_{i\neq j} b_i\mathbf{v}_i \\
&= b_j a_j^{-1}\mathbf{w} + \sum_{i\neq j}\left(b_i - \frac{b_j a_i}{a_j}\right)\mathbf{v}_i
\end{aligned}$$

This shows that $[S'] = \mathbf{V}$.

In order to show that S' is linearly independent, assume that we have

$$\begin{aligned}
\mathbf{0} &= c\mathbf{w} + \sum_{i\neq j} c_i\mathbf{v}_i \\
&= c\left(\sum_{i=1}^{n} a_i\mathbf{v}_i\right) + \sum_{i\neq j} c_i\mathbf{v}_i \\
&= \sum_{i=1}^{n}(ca_i + c_i)\mathbf{v}_i \qquad (\text{with } c_j = 0),
\end{aligned}$$

for some c, and $c_i \in F$. Since the original set S was assumed to be linearly independent, we must have $ca_i + c_i = 0$, for all i. In particular, since $c_j = 0$, we have $ca_j = 0$. But the assumption was that $a_j \neq 0$. Therefore we must conclude that $c = 0$. It follows that also $c_i = 0$, for all $i \neq j$. Therefore, S' must be linearly independent. $\square$

Theorem 9 (Steinitz Exchange Theorem)**.** *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbf{V}$ and let $T = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \subset \mathbf{V}$ be some linearly independent set of vectors in $\mathbf{V}$. Then we have $m \leq n$. By possibly re-ordering the elements of S, we may arrange things so that the set*

$$U = \{\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$$

is a basis for $\mathbf{V}$.

Proof. Use induction over the number m. If $m = 0$ then $U = S$ and there is nothing to prove. Therefore assume $m \geq 1$ and furthermore, the theorem is true for the case $m - 1$. So consider the linearly independent set $T' = \{\mathbf{w}_1, \ldots, \mathbf{w}_{m-1}\}$. After an appropriate re-ordering of S, we have $U' = \{\mathbf{w}_1, \ldots, \mathbf{w}_{m-1}, \mathbf{v}_m, \ldots, \mathbf{v}_n\}$ being a basis for $\mathbf{V}$. Note that if we were to have $n < m$, then T' would itself be a basis for $\mathbf{V}$. Thus we could express $\mathbf{w}_m$ as a linear combination of the vectors in T'. That would imply that T was not linearly independent, contradicting our assumption. Therefore, $m \leq n$.

Now since U' is a basis for $\mathbf{V}$, we can express $\mathbf{w}_m$ as a linear combination

$$\mathbf{w}_m = a_1\mathbf{w}_1 + \cdots + a_{m-1}\mathbf{w}_{m-1} + a_m\mathbf{v}_m + \cdots + a_n\mathbf{v}_n.$$

If we had all the coefficients of the vectors from S being zero, namely

$$a_m = a_{m+1} = \cdots = a_n = 0,$$

then we would have $\mathbf{w}_m$ being expressed as a linear combination of the other vectors in T. Therefore T would be linearly dependent, which is not true. Thus one of the $a_j \neq 0$, for $j \geq m$. Using theorem 8, we may exchange $\mathbf{w}_m$ for the vector $\mathbf{v}_j$ in U', thus giving us the basis U. ☐

Theorem 10 (Extension Theorem). *Assume that the vector space $\mathbf{V}$ is finitely generated and that we have a linearly independent subset $S \subset \mathbf{V}$. Then there exists a basis B of $\mathbf{V}$ with $S \subset B$.*

Proof. If $[S] = \mathbf{V}$ then we simply take $B = S$. Otherwise, start with some given basis $A \subset \mathbf{V}$ and apply theorem 9 successively. ☐

Theorem 11. *Let $\mathbf{U}$ be a subspace of the (finitely generated) vector space $\mathbf{V}$. Then $\mathbf{U}$ is also finitely generated, and each possible basis for $\mathbf{U}$ has no more elements than any basis for $\mathbf{V}$.*

Proof. Assume there is a basis B of $\mathbf{V}$ containing n vectors. Then, according to theorem 9, there cannot exist more than n linearly independent vectors in $\mathbf{U}$. Therefore $\mathbf{U}$ must be finitely generated, such that any basis for $\mathbf{U}$ has at most n elements. ☐

Theorem 12. *Assume the vector space $\mathbf{V}$ has a basis consisting of n elements. Then every basis of $\mathbf{V}$ also has precisely n elements.*

Proof. This follows directly from theorem 11, since any basis generates $\mathbf{V}$, which is a subspace of itself. ☐

Definition. *The number of vectors in a basis of the vector space $\mathbf{V}$ is called the* dimension *of $\mathbf{V}$, written $dim(\mathbf{V})$.*

Definition. *Let $\mathbf{V}$ be a vector space with subspaces $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$. The subspace $\mathbf{X} + \mathbf{Y} = [X \cup Y]$ is called the* sum *of $\mathbf{X}$ and $\mathbf{Y}$. If $\mathbf{X} \cap \mathbf{Y} = \{\mathbf{0}\}$, then it is the* direct sum*, written $\mathbf{X} \oplus \mathbf{Y}$.*

Theorem 13 (A Dimension Formula). *Let $\mathbf{V}$ be a finite dimensional vector space with subspaces $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$. Then we have*

$$dim(\mathbf{X} + \mathbf{Y}) = dim(\mathbf{X}) + dim(\mathbf{Y}) - dim(\mathbf{X} \cap \mathbf{Y}).$$

Corollary. $dim(\mathbf{X} \oplus \mathbf{Y}) = dim(\mathbf{X}) + dim(\mathbf{Y})$.

Proof of Theorem 13. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbf{X} \cap \mathbf{Y}$. According to theorem 10, there exist extensions $T = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ and $U = \{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$, such that $S \cup T$ is a basis for $\mathbf{X}$ and $S \cup U$ is a basis for $\mathbf{Y}$. We will now show that, in fact, $S \cup T \cup U$ is a basis for $\mathbf{X} + \mathbf{Y}$.

To begin with, it is clear that $\mathbf{X}+\mathbf{Y} = [S \cup T \cup U]$. Is the set $S \cup T \cup U$ linearly independent? Let

$$\begin{aligned} \mathbf{0} &= \sum_{i=1}^{n} a_i \mathbf{v}_i + \sum_{j=1}^{m} b_j \mathbf{x}_j + \sum_{k=1}^{r} c_k \mathbf{y}_k \\ &= \mathbf{v} + \mathbf{x} + \mathbf{y}, \quad \text{say.} \end{aligned}$$

Then we have $\mathbf{y} = -\mathbf{v} - \mathbf{x}$. Thus $\mathbf{y} \in \mathbf{X}$. But clearly we also have, $\mathbf{y} \in \mathbf{Y}$. Therefore $\mathbf{y} \in \mathbf{X} \cap \mathbf{Y}$. Thus $\mathbf{y}$ can be expressed as a linear combination of vectors in S alone, and since $S \cup U$ is is a basis for Y, we must have $c_k = 0$ for $k = 1, \ldots, r$. Similarly, looking at the vector $\mathbf{x}$ and applying the same argument, we conclude that all the b_j are zero. But then all the a_i must also be zero since the set S is linearly independent.

Putting this all together, we see that the $dim(\mathbf{X}) = n+m$, $dim(\mathbf{Y}) = n+r$ and $dim(\mathbf{X} \cap \mathbf{Y}) = n$. This gives the dimension formula. □

Theorem 14. *Let* $\mathbf{V}$ *be a finite dimensional vector space, and let* $\mathbf{X} \subset \mathbf{V}$ *be a subspace. Then there exists another subspace* $\mathbf{Y} \subset \mathbf{V}$*, such that* $\mathbf{V} = \mathbf{X} \oplus \mathbf{Y}$*.*

Proof. Take a basis S of $\mathbf{X}$. If $[S] = \mathbf{V}$ then we are finished. Otherwise, use the extension theorem (theorem 10) to find a basis B of $\mathbf{V}$, with $S \subset B$. Then[4] $\mathbf{Y} = [B \setminus S]$ satisfies the condition of the theorem. □

5 Linear Mappings

Definition. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be vector spaces, both over the field* F*. Let* $f : \mathbf{V} \to \mathbf{W}$ *be a mapping from the vector space* $\mathbf{V}$ *to the vector space* $\mathbf{W}$*. The mapping* f *is called a linear mapping if*

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v})$$

for all a*,* $b \in F$ *and all* $\mathbf{u}$*,* $\mathbf{v} \in \mathbf{V}$*.*

By choosing a and b to be either 0 or 1, we immediately see that a linear mapping always has both $f(a\mathbf{v}) = af(\mathbf{v})$ and $f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$, for all $a \in F$ and for all $\mathbf{u}$ and $\mathbf{v} \in \mathbf{V}$. Also it is obvious that $f(\mathbf{0}) = \mathbf{0}$ always.

Definition. *Let* $f : \mathbf{V} \to \mathbf{W}$ *be a linear mapping. The* kernel *of the mapping, denoted by* $ker(f)$*, is the set of vectors in* $\mathbf{V}$ *which are mapped by* f *into the zero vector in* $\mathbf{W}$*.*

Theorem 15. *If* $ker(f) = \{\mathbf{0}\}$*, that is, if the zero vector in* $\mathbf{V}$ *is the only vector which is mapped into the zero vector in* $\mathbf{W}$ *under* f*, then* f *is an injection (monomorphism). The converse is of course trivial.*

Proof. That is, we must show that if $\mathbf{u}$ and $\mathbf{v}$ are two vectors in $\mathbf{V}$ with the property that $f(\mathbf{u}) = f(\mathbf{v})$, then we must have $\mathbf{u} = \mathbf{v}$. But

$$f(\mathbf{u}) = f(\mathbf{v}) \quad \Rightarrow \quad \mathbf{0} = f(\mathbf{u}) - f(\mathbf{v}) = f(\mathbf{u} - \mathbf{v}).$$

[4] The notation $B \setminus S$ denotes the set of elements of B which are not in S

Thus the vector $\mathbf{u} - \mathbf{v}$ is mapped by f to the zero vector. Therefore we must have $\mathbf{u} - \mathbf{v} = \mathbf{0}$, or $\mathbf{u} = \mathbf{v}$.

Conversely, since $f(\mathbf{0}) = \mathbf{0}$ always holds, and since f is an injection, we must have $ker(f) = \{\mathbf{0}\}$. □

Theorem 16. *Let $f : \mathbf{V} \to \mathbf{W}$ be a linear mapping and let $A = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \subset \mathbf{W}$ be linearly independent. Assume that m vectors are given in $\mathbf{V}$, so that they form a set $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathbf{V}$ with $f(\mathbf{v}_i) = \mathbf{w}_i$, for all i. Then the set B is also linearly independent.*

Proof. Let $a_1, \ldots, a_m \in F$ be given such that $a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$. But then

$$\mathbf{0} = f(\mathbf{0}) = f(a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m) = a_1 f(\mathbf{v}_1) + \cdots + a_m f(\mathbf{v}_m) = a_1\mathbf{w}_1 + \cdots + a_m\mathbf{w}_m.$$

Since A is linearly independent, it follows that all the a_i's must be zero. But that implies that the set B is linearly independent. □

Remark. *If $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathbf{V}$ is linearly independent, and $f : \mathbf{V} \to \mathbf{W}$ is linear, still, it does not necessarily follow that $\{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_m)\}$ is linearly independent in $\mathbf{W}$. On the other hand, if f is an* injection, *then $\{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_m)\}$ is linearly independent. This follows since, if $a_1 f(\mathbf{v}_1) + \cdots + a_m f(\mathbf{v}_m) = \mathbf{0}$, then we have*

$$\mathbf{0} = a_1 f(\mathbf{v}_1) + \cdots + a_m f(\mathbf{v}_m) = f(a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m) = f(\mathbf{0}).$$

But since f is an injection, we must have $a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$. Thus $a_i = 0$ for all i.

On the other hand, what is the condition for $f : \mathbf{V} \to \mathbf{W}$ to be a surjection (epimorphism)? That is, $f(\mathbf{V}) = \mathbf{W}$. Or put another way, for every $\mathbf{w} \in \mathbf{W}$, can we find some vector $\mathbf{v} \in \mathbf{V}$ with $f(\mathbf{v}) = \mathbf{w}$? One way to think of this is to consider a basis $B \subset \mathbf{W}$. For each $\mathbf{w} \in B$, we take

$$f^{-1}(\mathbf{w}) = \{\mathbf{v} \in \mathbf{V} : f(\mathbf{v}) = \mathbf{w}\}.$$

Then f is a surjection if $f^{-1}(\mathbf{w}) \neq \emptyset$, for all $\mathbf{w} \in B$.

Definition. *A linear mapping which is a bijection (that is, an injection and a surjection) is called an* isomorphism. *Often one writes $\mathbf{V} \cong \mathbf{W}$ to say that there exists an isomorphism from $\mathbf{V}$ to $\mathbf{W}$.*

Theorem 17. *Let $f : \mathbf{V} \to \mathbf{W}$ be an isomorphism. Then the inverse mapping $f^{-1} : \mathbf{W} \to \mathbf{V}$ is also a linear mapping.*

Proof. To see this, let a, $b \in F$ and $\mathbf{x}$, $\mathbf{y} \in \mathbf{W}$ be arbitrary. Let $f^{-1}(\mathbf{x}) = \mathbf{u} \in \mathbf{V}$ and $f^{-1}(\mathbf{y}) = \mathbf{v} \in \mathbf{V}$, say. Then

$$f(a\mathbf{u} + b\mathbf{v}) = (f(af^{-1}(\mathbf{x}) + bf^{-1}(\mathbf{v})) = af(f^{-1}(\mathbf{x})) + bf(f^{-1}(\mathbf{v})) = a\mathbf{x} + b\mathbf{y}.$$

Therefore, since f is a bijection, we must have

$$f^{-1}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{u} + b\mathbf{v} = af^{-1}(\mathbf{x}) + bf^{-1}(\mathbf{y}).$$

□

Theorem 18. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be finite dimensional vector spaces over a field* F*, and let* $f : \mathbf{V} \to \mathbf{W}$ *be a linear mapping. Let* $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *be a basis for* $\mathbf{V}$*. Then* f *is uniquely determined by the* n *vectors* $\{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)\}$ *in* $\mathbf{W}$*.*

Proof. Let $\mathbf{v} \in \mathbf{V}$ be an arbitrary vector in $\mathbf{V}$. Since B is a basis for $\mathbf{V}$, we can uniquely write

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n,$$

with $a_i \in F$, for each i. Then, since the mapping f is linear, we have

$$\begin{aligned} f(\mathbf{v}) &= f(a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n) \\ &= f(a_1\mathbf{v}_1) + \cdots + f(a_n\mathbf{v}_n) \\ &= a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n). \end{aligned}$$

Therefore we see that if the values of $f(\mathbf{v}_1)$,..., $f(\mathbf{v}_n)$ are given, then the value of $f(\mathbf{v})$ is uniquely determined, for each $\mathbf{v} \in \mathbf{V}$.

On the other hand, let $A = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a set of n *arbitrarily given* vectors in $\mathbf{W}$. Then let a mapping $f : \mathbf{V} \to \mathbf{W}$ be *defined* by the rule

$$f(\mathbf{v}) = a_1\mathbf{u}_1 + \cdots a_n\mathbf{u}_n,$$

for each arbitrarily given vector $\mathbf{v} \in \mathbf{V}$, where $\mathbf{v} = a_1\mathbf{v}_1 + \cdots a_n\mathbf{v}_n$. Clearly the mapping is uniquely determined, since $\mathbf{v}$ is uniquely determined as a linear combination of the basis vectors B. It is a trivial matter to verify that the mapping which is so defined is also linear. We have $f(\mathbf{v}_i) = \mathbf{u}_i$ for all the basis vectors $\mathbf{v}_i \in B$. □

Theorem 19. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be two finite dimensional vector spaces over a field* F*. Then we have* $\mathbf{V} \cong \mathbf{W} \Leftrightarrow dim(\mathbf{V}) = dim(\mathbf{W})$*.*

Proof. "$\Rightarrow$" Let $f : \mathbf{V} \to \mathbf{W}$ be an isomorphism, and let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ be a basis for $\mathbf{V}$. Then, as shown in our Remark above, we have $A = \{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)\} \subset \mathbf{W}$ being linearly independent. Furthermore, since B is a basis of $\mathbf{V}$, we have $[B] = \mathbf{V}$. Thus $[A] = \mathbf{W}$ also. Therefore A is a basis of $\mathbf{W}$, and it contains precisely n elements; thus $dim(\mathbf{V}) = dim(\mathbf{W})$.

"$\Leftarrow$" Take $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{V}$ to again be a basis of $\mathbf{V}$ and let $A = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \subset \mathbf{W}$ be some basis of $\mathbf{W}$ (with n elements). Now define the mapping $f : \mathbf{V} \to \mathbf{W}$ by the rule $f(\mathbf{v}_i) = \mathbf{w}_i$, for all i. By theorem 18 we see that a linear mapping f is thus uniquely determined. Since A and B are both bases, it follows that f must be a bijection. □

This immediately gives us a complete classification of all finite-dimensional vector spaces. For let $\mathbf{V}$ be a vector space of dimension n over the field F. Then clearly F^n is also a vector space of dimension n over F. The canonical basis is the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, where

$$\mathbf{e}_i = (0, \cdots, 0, \underbrace{1}_{i\text{-th Position}}, 0, \ldots, 0\}$$

for each i. Therefore, when thinking about $\mathbf{V}$, we can think that it is "really" just F^n. On the other hand, the central idea in the theory of linear algebra is that we can look at things using different possible bases (or "frames of reference" in physics). The space F^n seems to have a preferred, fixed frame of reference, namely the canonical basis. Thus it is better to think about an abstract $\mathbf{V}$, with various possible bases.

Examples

For these examples, we will consider the 2-dimensional real vector space $\mathbb{R}^2$, together with its canonical basis $B = \{\mathbf{e}_1, \mathbf{e}_2\} = \{(1,0),(0,1)\}$.

- $f_1 : \mathbb{R}^2 \to \mathbb{R}^2$ with $f_1(\mathbf{e}_1) = (-1,0)$ and $f_1(\mathbf{e}_2) = (0,1)$. This is a *reflection* of the 2-dimensional plane into itself, with the axis of reflection being the second coordinate axis; that is the set of points $(x_1,x_2) \in \mathbb{R}^2$ with $x_1 = 0$.

- $f_2 : \mathbb{R}^2 \to \mathbb{R}^2$ with $f_2(\mathbf{e}_1) = \mathbf{e}_2$ and $f_1(\mathbf{e}_2) = \mathbf{e}_1$. This is a reflection of the 2-dimensional plane into itself, with the axis of reflection being the diagonal axis $x_1 = x_2$.

- $f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ with $f_3(\mathbf{e}_1) = (\cos\phi, \sin\phi)$ and $f_1(\mathbf{e}_2) = (-\sin\phi, \cos\phi)$, for some real number $\phi \in \mathbb{R}$. This is a *rotation* of the plane about its middle point, through an angle of ϕ.[5] For let $\mathbf{v} = (x_1,x_2)$ be some arbitrary point of the plane $\mathbb{R}^2$. Then we have

  $$\begin{aligned} f_3(\mathbf{v}) &= x_1 f_3(\mathbf{e}_1) + x_2 f(\mathbf{e}_2) \\ &= x_1(\cos\phi, \sin\phi) + x_2(-\sin\phi, \cos\phi) \\ &= (x_1\cos\phi - x_2\sin\phi, x_1\sin\phi + x_2\cos\phi). \end{aligned}$$

  Looking at this from the point of view of geometry, the question is, what happens to the vector $\mathbf{v}$ when it is rotated through the angle ϕ while preserving its length? Perhaps the best way to look at this is to think about $\mathbf{v}$ in *polar coordinates*. That is, given any two real numbers x_1 and x_2 then, assuming that they are not both zero, we find two *unique* real numbers $r \geq 0$ and $\theta \in [0, 2\pi)$, such that

  $$x_1 = r\cos\theta \quad \text{and} \quad x_2 = r\sin\theta,$$

  where $r = \sqrt{x_1^2 + x_2^2}$. Then $\mathbf{v} = (r\cos\theta, r\sin\theta)$. So a rotation of $\mathbf{v}$ through the angle ϕ must bring it to the new vector $(r\cos(\phi+\theta), r\sin(\phi+\theta))$ which, if we remember the formulas for cosines and sines of sums, turns out to be

  $$(r(\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)), r(\sin(\theta)\cos(\phi) - \cos(\theta)\sin(\phi)).$$

  But then, remembering that $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$, we see that the rotation brings the vector $\mathbf{v}$ into the new vector

  $$(x_1\cos\phi - x_2\sin\phi, x_1\sin\phi + x_2\cos\phi),$$

[5] In analysis, we learn about the formulas of trigonometry. In particular we have

$$\begin{aligned} \cos(\theta+\phi) &= \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi), \\ \sin(\theta+\phi) &= \sin(\theta)\cos(\phi) - \cos(\theta)\sin(\phi). \end{aligned}$$

Taking $\theta = \pi/2$, we note that $\cos(\phi + \pi/2) = -\sin(\phi)$ and $\sin(\phi+\pi/2) = \cos(\phi)$.

which was precisely the specification for $f_3(\mathbf{v})$.

6 Linear Mappings and Matrices

This last example of a linear mapping of $\mathbb{R}^2$ into itself — which should have been simple to describe — has brought with it long lines of lists of coordinates which are difficult to think about. In three and more dimensions, things become even worse! Thus it is obvious that we need a more sensible system for describing these linear mappings. The usual system is to use *matrices*.

Now, the most obvious problem with our previous notation for vectors was that the lists of the coordinates $(x_1, \cdots, x_n)$ run over the page, leaving hardly any room left over to describe symbolically what we want to do with the vector. The solution to this problem is to write vectors not as *horizontal* lists, but rather as *vertical* lists. We say that the horizontal lists are *row vectors*, and the vertical lists are *column vectors*. This is a great improvement! So whereas before, we wrote

$$\mathbf{v} = (x_1, \cdots, x_n),$$

now we will write

$$\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It is true that we use up lots of vertical space on the page in this way, but since the rest of the writing is horizontal, we can afford to waste this vertical space. In addition, we have a very nice system for writing down the coordinates of the vectors after they have been mapped by a linear mapping.

To illustrate this system, consider the rotation of the plane through the angle ϕ, which was described in the last section. In terms of row vectors, we have (x_1, x_2) being rotated into the new vector $(x_1 \cos\phi - x_2 \sin\phi, x_1 \sin\phi + x_2 \cos\phi)$. But if we change into the column vector notation, we have

$$\mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

being rotated to

$$\begin{pmatrix} x_1 \cos\phi - x_2 \sin\phi \\ x_1 \sin\phi + x_2 \cos\phi \end{pmatrix}.$$

But then, remembering how we *multiplied* matrices, we see that this is just

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \cos\phi - x_2 \sin\phi \\ x_1 \sin\phi + x_2 \cos\phi \end{pmatrix}.$$

So we can say that the 2×2 matrix $A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$ represents the mapping $f_3 : \mathbb{R}^2 \to \mathbb{R}^2$, and the 2×1 matrix $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ represents the vector $\mathbf{v}$. Thus we have

$$A \cdot \mathbf{v} = f(\mathbf{v}).$$

That is, matrix multiplication gives the result of the linear mapping.

Expressing $f : \mathbf{V} \to \mathbf{W}$ in terms of bases for both $\mathbf{V}$ and $\mathbf{W}$

The example we have been thinking about up till now (a rotation of $\mathbb{R}^2$) is a linear mapping of $\mathbb{R}^2$ into itself. More generally, we have linear mappings from a vector space $\mathbf{V}$ to a *different* vector space $\mathbf{W}$ (although, of course, both $\mathbf{V}$ and $\mathbf{W}$ are vector spaces over the same field F).

So let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbf{V}$ and let $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be a basis for $\mathbf{W}$. Finally, let $f : \mathbf{V} \to \mathbf{W}$ be a linear mapping. An arbitrary vector $\mathbf{v} \in \mathbf{V}$ can be expressed in terms of the basis for $\mathbf{V}$ as

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \sum_{j=1}^{n} a_j\mathbf{v}_j.$$

The question is now, what is $f(\mathbf{v})$? As we have seen, $f(\mathbf{v})$ can be expressed in terms of the images $f(\mathbf{v}_j)$ of the basis vectors of $\mathbf{V}$. Namely

$$f(\mathbf{v}) = \sum_{j=1}^{n} a_j f(\mathbf{v}_j).$$

But then, each of these vectors $f(\mathbf{v}_j)$ in $\mathbf{W}$ can be expressed in terms of the basis vectors in $\mathbf{W}$, say

$$f(\mathbf{v}_j) = \sum_{i=1}^{m} c_{ij}\mathbf{w}_i,$$

for appropriate choices of the "numbers" $c_{ij} \in F$. Therefore, putting this all together, we have

$$f(\mathbf{v}) = \sum_{j=1}^{n} a_j f(\mathbf{v}_j) = \sum_{j=1}^{n}\sum_{i=1}^{m} a_j c_{ij}\mathbf{w}_i.$$

In the matrix notation, using column vectors relative to the two bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$, we can write this as

$$f(\mathbf{v}) = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_j c_{1j} \\ \vdots \\ \sum_{j=1}^{n} a_j c_{mj} \end{pmatrix}.$$

When looking at this $m \times n$ matrix which represents the linear mapping $f : \mathbf{V} \to \mathbf{W}$, we can imagine that the matrix consists of n columns. The i-th column is then

$$\mathbf{u}_i = \begin{pmatrix} c_{1i} \\ \vdots \\ c_{mi} \end{pmatrix} \in \mathbf{W}.$$

That is, it represents a vector in $\mathbf{W}$, namely the vector $\mathbf{u}_i = c_{1i}\mathbf{w}_1 + \cdots + c_{mi}\mathbf{w}_m$.

But what is this vector $\mathbf{u}_i$? In the matrix notation, we have

$$\mathbf{v}_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbf{V},$$

where the single non-zero element of this column matrix is a 1 in the i-th position from the top. But then we have

$$f(\mathbf{v}_i) = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} c_{1i} \\ \vdots \\ c_{mi} \end{pmatrix} = \mathbf{u}_i.$$

Therefore *the columns of the matrix representing the linear mapping $f : \mathbf{V} \to \mathbf{W}$ are the images of the basis vectors of $\mathbf{V}$.*

Two linear mappings, one after the other

Things become more interesting when we think about the following situation. Let $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{X}$ be vector spaces over a common field F. Assume that $f : \mathbf{V} \to \mathbf{W}$ and $g : \mathbf{W} \to \mathbf{X}$ are linear. Then the composition $f \circ g : \mathbf{V} \to \mathbf{X}$, given by

$$f \circ g(\mathbf{v}) = g(f(\mathbf{v}))$$

for all $\mathbf{v} \in \mathbf{V}$ is clearly a linear mapping. One can write this as

$$\mathbf{V} \xrightarrow{f} \mathbf{W} \xrightarrow{g} \mathbf{X}.$$

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbf{V}$, $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be a basis for $\mathbf{W}$, and $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ be a basis for $\mathbf{X}$. Assume that the linear mapping f is given by the matrix

$$A = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix},$$

and the linear mapping g is given by the matrix

$$B = \begin{pmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{r1} & \cdots & d_{rm} \end{pmatrix}.$$

Then, if $\mathbf{v} = \sum_{j=1}^{n} a_j v_j$ is some arbitrary vector in $\mathbf{V}$, we have

$$\begin{aligned} f \circ g(\mathbf{v}) &= g\left(f\left(\sum_{j=1}^{n} a_j \mathbf{v}_j\right)\right) \\ &= g\left(\sum_{j=1}^{n} a_j f(\mathbf{v}_j)\right) \\ &= g\left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_j c_{ij} \mathbf{w}_i\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_j c_{ij} g(\mathbf{w}_i) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r} a_j c_{ij} d_{ki} \mathbf{x}_k. \end{aligned}$$

There are so many summations here! How can we keep track of everything? The answer is to use the matrix notation. The composition of linear mappings is then simply represented by matrix *multiplication*. That is, if

$$\mathbf{v} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

then we have

$$f \circ g(\mathbf{v}) = g(f(\mathbf{v})) = \begin{pmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{r1} & \cdots & d_{rm} \end{pmatrix} \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = BA\mathbf{v}.$$

So this is the reason we have defined matrix multiplication in this way.[6]

7 Matrix Transformations

Matrices are used to describe linear mappings $f : \mathbf{V} \to \mathbf{W}$ with respect to particular bases of $\mathbf{V}$ and $\mathbf{W}$. But clearly, if we choose *different* bases than the ones we had been thinking about before, then we will have a different matrix for describing the *same* linear mapping. Later on in these lectures we will see how changing the bases changes the matrix, but for now, it is time to think about various systematic ways of changing matrices — in a purely abstract way.

[6]Recall that if $A = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix}$ is an $m \times n$ matrix and $B = \begin{pmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{r1} & \cdots & d_{rm} \end{pmatrix}$ is an $r \times m$ matrix, then the product BA is an $r \times n$ matrix whose kj-th element is $\sum_{i=1}^{m} d_{ki} c_{ij}$.

Elementary Column Operations

We begin with the elementary column operations. Let us denote the set of all $n \times m$ matrices of elements from the field F by $M(m \times n, F)$. Thus, if

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in M(m \times n, F)$$

then it contains n columns which, as we have seen, are the images of the basis vectors of the linear mapping which is being represented by the matrix. So The first elementary column operation is to exchange column i with column j, for $i \neq j$. We can write

$$\begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1m} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mm} \end{pmatrix} \xrightarrow{S_{ij}} \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1m} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mi} & \cdots & a_{mm} \end{pmatrix}$$

So this column operation is denoted by S_{ij}. It can be thought of as being a mapping $S_{ij} : M(m \times n, F) \to M(m \times n, F)$.

Another way to imagine this is to say that S is the set of column vectors in the matrix A considered as an ordered list. Thus $S \subset F^m$. Then S_{ij} is the same set of column vectors, but with the positions of the i-th and the j-th vectors interchanged. But obviously, as a subset of F^n, the order of the vectors makes no difference. Therefore we can say that the span of S is the same as the span of S_{ij}. That is $[S] = [S_{ij}]$.

The second elementary column operation, denoted $S_i(a)$, is that we form the scalar product of the element $a \neq 0$ in F with the i-th vector in S. So the i-th vector

$$\begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}$$

is changed to

$$a \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} = \begin{pmatrix} aa_{1i} \\ \vdots \\ aa_{mi} \end{pmatrix}.$$

All the other column vectors in the matrix remain unchanged.

The third elementary column operation, denoted $S_{ij}(c)$ is that we take the j-th column (where $j \neq i$) and multiply it with $c \neq 0$, then add it to the i-th column. Therefore the i-th column is changed to

$$\begin{pmatrix} a_{1i} + ca_{1j} \\ \vdots \\ a_{mi} + ca_{mj} \end{pmatrix}.$$

All the other columns — *including the j-th column* — remain unchanged.

Theorem 20. $[S] = [S_{ij}] = [S_i(a)] = [S_{ij}(c)]$, *where* $i \neq j$ *and* $a \neq 0 \neq c$.

Proof. Let us say that $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset F^m$. That is, $\mathbf{v}_i$ is the i-th column vector of the matrix A, for each i. We have already seen that $[S] = [S_{ij}]$ is trivially true. But also, say $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ is some arbitrary vector in $[S]$. Then, since $a \neq 0$, we can write

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + a^{-1}x_i(a\mathbf{v}_i) + \cdots + x_n\mathbf{v}_n.$$

Therefore $[S] \subset [S_i(a)]$. The other inclusion, $[S_i(a)] \subset [S]$ is also quite trivial so that we have $[S] = [S_i(a)]$.

Similarly we can write

$$\begin{aligned} \mathbf{v} &= x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n \\ &= x_1\mathbf{v}_1 + \cdots + x_i(\mathbf{v}_i + c\mathbf{v}_j) + \cdots + (x_j - x_ic)\mathbf{v}_j + \cdots + x_n\mathbf{v}_n. \end{aligned}$$

Therefore $[S] \subset [S_{ij}(c)]$, and again, the other inclusion is similar. $\square$

Let us call $[S]$ the *column space* (Spaltenraum), which is a subspace of F^m. Then we see that the column space remains invariant under the three types of elementary column operations. In particular, the *dimension* of the column space remains invariant.

Elementary Row Operations

Again, looking at the $m \times n$ matrix A in a purely abstract way, we can say that it is made up of m *row vectors*, which are just the rows of the matrix. Let us call them $\mathbf{w}_1, \ldots, \mathbf{w}_m \in F^n$. That is,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 & = & (a_{11} & \cdots & a_{1n}) \\ & \vdots & & & \\ \mathbf{w}_m & = & (a_{m1} & \cdots & a_{mn}) \end{pmatrix}.$$

Again, we define the three elementary row operations analogously to the way we defined the elementary column operations. Clearly we have the same results. Namely if $R = \{w_1, \ldots, w_m\}$ are the original rows, in their proper order, then we have $[R] = [R_{ij}] = [R_i(a)] = [R_{ij}(c)]$.

But it is perhaps easier to think about the row operations when changing a matrix into a form which is easier to think about. We would like to change the matrix into a step form (*Zeilenstufenform*).

Definition. *The* $m \times n$ *matrix* A *is in* step form *if there exists some* r *with* $0 \leq r \leq m$ *and indices* $1 \leq j_1 < j_2 < \cdots < j_r \leq m$ *with* $a_{ij_i} = 1$ *for all* $i = 1, \ldots, r$ *and* $a_{st} = 0$ *for all* s, t *with* $t < j_s$ *or* $s > j_r$. *That is:*

$$A = \begin{pmatrix} \cdots & 1 & a_{1j_1+1} & & \cdots & \cdots & & & a_{1n} \\ 0 & & 1 & a_{2j_2+1} & & \cdots & & & a_{2n} \\ 0 & & 0 & 1 & a_{3j_3+1} & \cdots & & & a_{2n} \\ & & & & \ddots & & & & \vdots \\ 0 & & \cdots & & 0 & 1 & a_{rj_r+1} & \cdots & a_{rn} \\ 0 & & & & 0 & 0 & 0 & & 0 \\ \vdots & & & & \vdots & & & & \vdots \end{pmatrix}.$$

Theorem 21. *By means of a finite sequence of elementary row operations, every matrix can be transformed into a matrix in step form.*

Proof. Induction on m, the number of rows in the matrix. We use the technique of "Gaussian elimination", which is simply the usual way anyone would go about solving a system of linear equations. This will be dealt with in the next section. The induction step in this proof, which uses a number of simple ideas which are easy to write on the blackboard, but overly tedious to compose here in TEX, will be described in the lecture. $\square$

Now it is obvious that the *row space* (Zeilenraum), that is $[R] \subset F^n$, has the dimension r, and in fact the non-zero row vectors of a matrix in step form provide us with a basis for the row space. But then, looking at the *column* vectors of this matrix in step form, we see that the columns j_1, j_2, and so on up to j_r are all linearly independent, and they generate the column space. (This is discussed more fully in the lecture!)

Definition. *Given an $m \times n$ matrix, the dimension of the column space is called the column rank; similarly the dimension of the row space is the row rank.*

So, using theorem 21 and exercise 6.3, we conclude that:

Theorem 22. *For any matrix A, the column rank is equal to the row rank. This common dimension is simply called the* rank *— written $Rank(A)$ — of the matrix.*

Definition. *Let A be a quadratic $n \times n$ matrix. Then A is called* regular *if $Rank(A) = n$, otherwise A is called* singular.

Theorem 23. *The $n \times n$ matrix A is regular $\Leftrightarrow$ the linear mapping $f : F^n \to F^n$, represented by the matrix A with respect to the canonical basis of F^n is an isomorphism.*

Proof. '$\Rightarrow$' If A is regular, then the rank of A — namely the dimension of the column space $[S]$ — is n. Since the dimension of F^n is n, we must therefore have $[S] = F^n$. The linear mapping $f : F^n \to F^n$ is then both an injection (since S must be linearly independent) and also a surjection.

'$\Leftarrow$' Since the set of column vectors S is the set of images of the canonical basis vectors of F^n under f, they must be linearly independent. There are n column vectors; thus the rank of A is n. $\square$

8 Systems of Linear Equations

We now take a small diversion from our idea of linear algebra as being a method of describing *geometry*, and instead we will consider simple linear equations. In particular, we consider a system of m equations in n unknowns.

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

We can also think about this as being a vector equation. That is, if

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix},$$

and $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in F^m$, then our system of linear equations is just the single vector equation

$$A \cdot \mathbf{x} = \mathbf{b}.$$

But what is the most obvious way to solve this system of equations? It is a simple matter to write down an algorithm, as follows. The numbers a_{ij} and b_k are given (as elements of F), and the problem is to find the numbers x_l.

1. Let $i := 1$ and $j := 1$.

2. if $a_{ij} = 0$ then if $a_{kj} = 0$ for all $i < k \leq m$, set $j := j + 1$. Otherwise find the smallest index $k > i$ such that $a_{kj} \neq 0$ and exchange the i-th equation with the k-th equation.

3. Multiply both sides of the (possibly new) i-th equation by a_{ij}^{-1}. Then for each $i < k \leq m$, subtract a_{kj} times the i-th equation from the k-th equation. Therefore, at this stage, after this operation has been carried out, we will have $a_{kj} = 0$, for all $k > i$.

4. Set $i := i + 1$. If $i \leq n$ then return to step 2.

So at this stage, we have transformed the system of linear equations into a system in step form.

The next thing is to solve the system of equations in step form. The problem is that perhaps there is no solution, or perhaps there are many solutions. The easiest way to decide which case we have is to reorder the variables — that is the various x_i — so that the steps start in the upper left-hand corner, and they are all one unit wide. That is, things then look like this:

$$\begin{array}{ccccccccccccc}
x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & \cdots & + & a_{1n}x_n & = & b_1 \\
 & & x_2 & + & a_{23}x_3 & + & a_{24}x_4 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
 & & & & x_3 & + & a_{34}x_4 & + & \cdots & + & a_{3n}x_n & = & b_3 \\
 & & & & & & & & \vdots & & & & \\
 & & & & & & x_k & + & \cdots & + & a_{kn}x_k & = & b_k \\
 & & & & & & & & & & 0 & = & b_{k+1} \\
 & & & & & & & & & & & \vdots & \\
 & & & & & & & & & & 0 & = & b_m
\end{array}$$

(Note that this reordering of the variables is like our first elementary column operation for matrices.)

So now we observe that:

- If $b_l \neq 0$ for some $k+1 \leq l \leq m$, then the system of equations has *no solution*.
- Otherwise, if $k = n$ then the system has precisely one single solution. It is obtained by working backwards through the equations. Namely, the last equation is simply $x_n = b_n$, so that is clear. But then, substitute b_n for x_n in the $n-1$-st equation, and we then have $x_{n-1} = b_{n-1} - a_{n-1\,n}b_n$. By this method, we progress back to the first equation and obtain values for all the x_j, for $1 \leq j \leq n$.
- Otherwise, $k < n$. In this case we can assign *arbitrary* values to the variables $x_{k+1}, \ldots, x_n$, and then that fixes the value of x_k. But then, as before, we progressively obtain the values of x_{k-1}, x_{k-2} and so on, back to x_1.

This algorithm for finding solutions of systems of linear equations is called "Gaussian Elimination".

All of this can be looked at in terms of our matrix notation. Let us call the following $m \times n+1$ matrix the augmented matrix for our system of linear equations:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

Then by means of elementary row and column operations, the matrix is transformed into the new matrix which is in *simple* step form

$$A' = \begin{pmatrix} 1 & a'_{12} & \cdots & \cdot & a'_{1\,k+1} & \cdots & a'_{1n} & b'_1 \\ 0 & 1 & a'_{23} & \cdot & a'_{2\,k+1} & \cdots & a'_{2n} & b'_2 \\ \vdots & & \ddots & \cdot & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & a'_{k\,k+1} & \cdots & a'_{kn} & b'_k \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & b'_{k+1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & b'_m \end{pmatrix}.$$

Finding the eigenvectors of linear mappings

Definition. *Let* $\mathbf{V}$ *be a vector space over a field* F*, and let* $f : \mathbf{V} \to \mathbf{V}$ *be a linear mapping of* $\mathbf{V}$ *into itself. An* eigenvector *of* f *is a non-zero vector* $\mathbf{v} \in \mathbf{V}$ *(so we have* $\mathbf{v} \neq \mathbf{0}$*) such that there exists some* $\lambda \in F$ *with* $f(\mathbf{v}) = \lambda\mathbf{v}$*. The scalar* λ *is then called the* eigenvalue *associated with this eigenvector.*

So if f is represented by the $n \times n$ matrix A (with respect to some given basis of $\mathbf{V}$), then the problem of finding eigenvectors and eigenvalues is simply the problem of solving the equation

$$A\mathbf{v} = \lambda\mathbf{v}.$$

But here *both* λ and $\mathbf{v}$ are variables. So how should we go about things? Well, as we will see, it is necessary to look at the *characteristic polynomial* of the matrix, in order to find an eigenvalue λ. Then, once an eigenvalue is found, we can consider it to be a constant in our system of linear equations. And they become the *homogeneous*[7] system

$$\begin{array}{ccccccccc}
(a_{11}-\lambda)x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\
a_{21}x_1 & + & (a_{22}-\lambda)x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\
 & & & & & & & \vdots & \\
a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & (a_{nn}-\lambda)x_n & = & 0
\end{array}$$

which can be easily solved to give us the (or one of the) eigenvector(s) whose eigenvalue is λ.

Now the $n \times n$ *identity* matrix is

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

Thus we see that an eigenvalue is any scalar $\lambda \in F$ such that the vector equation

$$(A - \lambda E)\mathbf{v} = \mathbf{0}$$

has a solution vector $\mathbf{v} \in \mathbf{V}$, such that $\mathbf{v} \neq \mathbf{0}$.[8]

9 Invertible Matrices

Let $f : \mathbf{V} \to \mathbf{W}$ be a linear mapping, and let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbf{V}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\} \subset \mathbf{W}$ be bases for $\mathbf{V}$ and $\mathbf{W}$, respectively. Then, as we have seen, the mapping f can be uniquely described by specifying the values of $f(\mathbf{v}_j)$, for each $j = 1, \dots, n$. We have

$$f(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i,$$

And the resulting matrix $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ is the matrix describing f with respect to these *given bases*.

A particular case

This is the case that $\mathbf{V} = \mathbf{W}$. So we have the linear mapping $f : \mathbf{V} \to \mathbf{V}$. But now, we only need a *single* basis for $\mathbf{V}$. That is, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbf{V}$ is the only basis we

[7]That is, all the b_i are zero. Thus a homogeneous system with matrix A has the form $A\mathbf{v} = \mathbf{0}$.

[8]Given any solution vector $\mathbf{v}$, then clearly we can multiply it with any scalar $\kappa \in F$, and we have

$$(A - \lambda E)(\kappa\mathbf{v}) = \kappa(A - \lambda E)\mathbf{v} = \kappa\mathbf{0} = \mathbf{0}.$$

Therefore, as long as $\kappa \neq 0$, we can say that $\kappa\mathbf{v}$ is also an eigenvector whose eigenvalue is λ.

need. Thus the matrix for f with respect to this single basis is determined by the specifications

$$f(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{v}_i.$$

A trivial example

For example, one particular case is that we have the identity mapping

$$f = id : \mathbf{V} \to \mathbf{V}.$$

Thus $f(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathbf{V}$. In this case it is obvious that the matrix of the mapping is the $n \times n$ identity matrix I_n.

Regular matrices

Let us now assume that A is some regular $n \times n$ matrix. As we have seen in theorem 23, there is an *isomorphism* $f : \mathbf{V} \to \mathbf{V}$, such that A is the matrix representing f with respect to the given basis of $\mathbf{V}$. According to theorem 17, the inverse mapping f^{-1} is also linear, and we have $f^{-1} \circ f = id$. So let f^{-1} be represented by the matrix B (again with respect to the same basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$). Then we must have the matrix equation

$$B \cdot A = I_n.$$

Or, put another way, in the multiplication system of matrix algebra we must have $B = A^{-1}$. That is, the matrix A is *invertible*.

Theorem 24. *Every regular matrix is invertible.*

Definition. *The set of all regular $n \times n$ matrices over the field F is denoted $GL(n, F)$.*

Theorem 25. *$GL(n, F)$ is a group under matrix multiplication. The identity element is the identity matrix.*

Proof. We have already seen in an exercise that matrix multiplication is associative. The fact that the identity element in $GL(n, F)$ is the identity matrix is clear. By definition, all members of $GL(n, F)$ have an inverse. It only remains to see that $GL(n, F)$ is closed under matrix multiplication. So let $A, C \in GL(n, F)$. Then there exist A^{-1}, $C^{-1} \in GL(n, F)$, and we have that $C^{-1} \cdot A^{-1}$ is itself an $n \times n$ matrix. But then

$$\left(C^{-1}A^{-1}\right) AC = C^{-1}\left(A^{-1}A\right) C = C^{-1}I_nC = C^{-1}C = I_n.$$

Therefore, according to the definition of $GL(n, F)$, we must also have $AC \in GL(n, F)$. $\square$

Simplifying matrices using multiplication with regular matrices

Theorem 26. *Let A be an $m \times n$ matrix. Then there exist regular matrices $C \in GL(m, F)$ and $D \in GL(n, F)$ such that the matrix $A' = CAD^{-1}$ consists simply of zeros, except possibly for a block in the upper lefthand corner, which is an identity matrix. That is*

$$A' = \left(\begin{array}{ccc|ccc} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right)$$

(Note that A' is also an $m \times n$ matrix. That is, it is not necessarily square.)

Proof. A is the representation of a linear mapping $f : \mathbf{V} \to \mathbf{W}$, with respect to bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of $\mathbf{V}$ and $\mathbf{W}$, respectively. The idea of the proof is to now find *new* bases $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathbf{V}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subset \mathbf{W}$, such that the matrix of f with respect to these new bases is as simple as possible.

So to begin with, let us look at $ker(f) \subset \mathbf{V}$. It is a subspace of $\mathbf{V}$, so its dimension is at most n. In general, it might be less than n, so let us write $dim(ker(f)) = n-p$, for some integer $0 \le p \le n$. Therefore we choose a basis for $ker(f)$, and we call it

$$\{\mathbf{x}_{p+1}, \ldots, \mathbf{x}_n\} \subset ker(f) \subset \mathbf{V}.$$

Using the extension theorem (theorem 12), we extend this to a basis

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_n\}$$

for $\mathbf{V}$.

Now at this stage, we look at the images of the vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ under f in $\mathbf{W}$. We find that the set $\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_p)\} \subset \mathbf{W}$ is linearly independent. To see this, let us assume that we have the vector equation

$$\mathbf{0} = \sum_{i=1}^{p} a_i f(\mathbf{x}_i) = f\left(\sum_{i=1}^{p} a_i \mathbf{x}_i\right)$$

for some choice of the scalars a_i. But that means that $\sum_{i=1}^{p} a_i \mathbf{x}_i \in ker(f)$. However $\{\mathbf{x}_{p+1}, \ldots, \mathbf{x}_n\}$ is a basis for $ker(f)$. Thus we have

$$\sum_{i=1}^{p} a_i \mathbf{x}_i = \sum_{j=p+1}^{n} b_j \mathbf{x}_j$$

for appropriate choices of scalars b_j. But $\{\mathbf{x}_1, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_n\}$ is a basis for $\mathbf{V}$. Thus it is itself linearly independent and therefore we must have $a_i = 0$ and $b_j = 0$ for all possible i and j. In particular, since the a_i's are all zero, we must have the set $\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_p)\} \subset \mathbf{W}$ being linearly independent.

To simplify the notation, let us call $f(\mathbf{x}_i) = \mathbf{y}_i$, for each $i = 1, \ldots, p$. Then we can again use the extension theorem to find a basis

$$\{\mathbf{y}_1, \ldots, \mathbf{y}_p, \mathbf{y}_{p+1}, \ldots, \mathbf{y}_m\}$$

of $\mathbf{W}$.

So now we define the isomorphism $g : \mathbf{V} \to \mathbf{V}$ by the rule

$$g(\mathbf{x}_i) = \mathbf{v}_i, \quad \text{for all} \quad i = 1, \ldots, n.$$

Similarly the isomorphism $h : \mathbf{W} \to \mathbf{W}$ is defined by the rule

$$h(\mathbf{y}_j) = \mathbf{w}_j, \quad \text{for all} \quad j = 1, \ldots, m.$$

Let D be the matrix representing the mapping g with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbf{V}$, and also let C be the matrix representing the mapping h with respect to the basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of $\mathbf{W}$.

Let us now look at the mapping

$$h \cdot f \cdot g^{-1} : \mathbf{V} \to \mathbf{W}.$$

For the basis vector $\mathbf{v}_i \in \mathbf{V}$, we have

$$hfg^{-1}(\mathbf{v}_i) = hf(\mathbf{x}_i) = \begin{cases} h(\mathbf{y}_i) = \mathbf{w}_i, & \text{for } i \leq p \\ h(\mathbf{0}) = \mathbf{0}, & \text{otherwise.} \end{cases}$$

This mapping must therefore be represented by a matrix in our simple form, consisting of only zeros, except possibly for a block in the upper lefthand corner which is an identity matrix. Furthermore, the rule that the composition of linear mappings is represented by the product of the respective matrices leads to the conclusion that the matrix $A' = CAD^{-1}$ must be of the desired form. $\square$

10 Similar Matrices; Changing Bases

Definition. *Let A and A' be $n \times n$ matrices. If a matrix $C \in GL(n, F)$ exists, such that $A' = C^{-1}AC$ then we say that the matrices A and A' are* similar.

Theorem 27. *Let $f : \mathbf{V} \to \mathbf{V}$ be a linear mapping and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be two bases for $\mathbf{V}$. Assume that A is the matrix for f with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and furthermore A' is the matrix for f with respect to the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Let $\mathbf{u}_i = \sum_{j=1}^{n} c_{ji}\mathbf{v}_j$ for all i, and*

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

Then we have $A' = C^{-1}AC$.

Proof. From the definition of A', we have

$$f(\mathbf{u}_i) = \sum_{j=1}^{n} a'_{ji}\mathbf{u}_j$$

for all $i = 1, \ldots, n$. On the other hand we have

$$\begin{aligned} f(\mathbf{u}_i) &= f\left(\sum_{j=1}^{n} c_{ji}\mathbf{v}_j\right) \\ &= \sum_{j=1}^{n} c_{ji} f(\mathbf{v}_j) \\ &= \sum_{j=1}^{n} c_{ji}\left(\sum_{k=1}^{n} a_{kj}\mathbf{v}_k\right) \\ &= \sum_{k=1}^{n}\left(\sum_{j=1}^{n} c_{ji}a_{kj}\right)\mathbf{v}_k \\ &= \sum_{k=1}^{n}\left(\sum_{j=1}^{n} a_{kj}c_{ji}\right)\left(\sum_{l=1}^{n} c^*_{lk}\mathbf{u}_l\right) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(c^*_{lk}a_{kj}c_{ji}\right)\mathbf{u}_l. \end{aligned}$$

Here, the inverse matrix C^{-1} is denoted by

$$C^{-1} = \begin{pmatrix} c^*_{11} & \cdots & c^*_{1n} \\ \vdots & \ddots & \vdots \\ c^*_{n1} & \cdots & c^*_{nn} \end{pmatrix}.$$

Therefore we have $A' = C^{-1}AC$. $\square$

Note that we have written here $\mathbf{v}_k = \sum_{l=1}^{n} c^*_{lk}\mathbf{u}_l$, and then we have said that the resulting matrix (which we call C^*) is, in fact, C^{-1}. To see that this is true, we begin with the definition of C itself. We have

$$\mathbf{u}_l = \sum_{j=1}^{n} c_{jl}\mathbf{v}_j.$$

Therefore

$$\mathbf{v}_k = \sum_{l=1}^{n}\sum_{j=1}^{n} c_{jl}c^*_{lk}\mathbf{v}_j.$$

That is, $CC^* = I_n$, and therefore $C^* = C^{-1}$.

Which mapping does the matrix C represent? From the equations $\mathbf{u}_i = \sum_{j=1}^{n} c_{ji}\mathbf{v}_j$ we see that it represents a mapping $g : \mathbf{V} \to \mathbf{V}$ such that $g(\mathbf{v}_i) = \mathbf{u}_i$ for all i, expressed in terms of the original basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. So we see that a *similarity* *transformation*, taking a square matrix A to a similar matrix $A' = C^{-1}AC$ is always associated with a *change of basis* for the vector space V.

Much of the theory of linear algebra is concerned with finding a *simple* basis (with respect to a given linear mapping of the vector space into itself), such that the matrix of the mapping with respect to this simpler basis is itself simple — for example diagonal, or at least trigonal.

11 Eigenvalues, Eigenspaces, Matrices which can be Diagonalized

Definition. *Let $f : \mathbf{V} \to \mathbf{V}$ be a linear mapping of an n-dimensional vector space into itself. A subspace $\mathbf{U} \subset \mathbf{V}$ is called* invariant *with respect to f if $f(\mathbf{U}) \subset \mathbf{U}$. That is, $f(\mathbf{u}) \in \mathbf{U}$ for all $\mathbf{u} \in \mathbf{U}$.*

Theorem 28. *Assume that the r dimensional subspace $\mathbf{U} \subset \mathbf{V}$ is invariant with respect to $f : \mathbf{V} \to \mathbf{V}$. Let A be the matrix representing f with respect to a given basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbf{V}$. Then A is similar to a matrix A' which has the following form*

$$A' = \left(\begin{array}{ccc|ccc} a'_{11} & \ldots & a'_{1r} & a'_{1(r+1)} & \ldots & a'_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a'_{r1} & \ldots & a'_{rr} & a'_{r(r+1)} & \ldots & a'_{rn} \\ \hline & & & a'_{(r+1)(r+1)} & \ldots & a'_{(r+1)n} \\ & 0 & & \vdots & & \vdots \\ & & & a'_{n(r+1)} & \ldots & a'_{nn} \end{array}\right)$$

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis for the subspace $\mathbf{U}$. Then extend this to a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ of $\mathbf{V}$. The matrix of f with respect to this new basis has the desired form. □

Definition. *Let $\mathbf{U}_1, \ldots, \mathbf{U}_p \subset \mathbf{V}$ be subspaces. We say that $\mathbf{V}$ is the* direct sum *of these subspaces if $\mathbf{V} = \mathbf{U}_1 + \cdots + \mathbf{U}_p$, and furthermore if $\mathbf{v} = \mathbf{u}_1 + \cdots + \mathbf{u}_p$ such that $\mathbf{u}_i \in \mathbf{U}_i$, for each i, then this expression for $\mathbf{v}$ is* unique. *In other words, if $\mathbf{v} = \mathbf{u}_1 + \cdots + \mathbf{u}_p = \mathbf{u}'_1 + \cdots + \mathbf{u}'_p$ with $\mathbf{u}'_i \in \mathbf{U}_i$ for each i, then $\mathbf{u}_i = \mathbf{u}'_i$, for each i. In this case, one writes $\mathbf{V} = \mathbf{U}_1 \oplus \cdots \oplus \mathbf{U}_p$*

This immediately gives the following result:

Theorem 29. *Let $f : \mathbf{V} \to \mathbf{V}$ be such that there exist subspaces $\mathbf{U}_i \subset \mathbf{V}$, for $i = 1, \ldots, p$, such that $\mathbf{V} = \mathbf{U}_1 \oplus \cdots \oplus \mathbf{U}_p$ and also f is invariant with respect to each $\mathbf{U}_i$. Then there exists a basis of $\mathbf{V}$ such that the matrix of f with respect to this basis has the following* block form.

$$A = \begin{pmatrix} A_1 & 0 & \ldots & & 0 \\ 0 & A_2 & 0 & & \\ \vdots & 0 & \ddots & 0 & \vdots \\ & & 0 & A_{p-1} & 0 \\ 0 & & \ldots & 0 & A_p \end{pmatrix}$$

where each block A_i is a square matrix, representing the restriction of f to the subspace $\mathbf{U}_i$.

Proof. Choose the basis to be a union of bases for each of the $\mathbf{U}_i$. $\square$

A special case is when the invariant subspace is an eigenspace.

Definition. *Assume that $\lambda \in F$ is an eigenvalue of the mapping $f : \mathbf{V} \to \mathbf{V}$. The set $\{\mathbf{v} \in \mathbf{V} : f(\mathbf{v}) = \lambda\mathbf{v}\}$ is called the* eigenspace *of λ with respect to the mapping f. That is, the eigenspace is the set of all eigenvectors (and with the zero vector $\mathbf{0}$ included) with eigenvalue λ.*

Theorem 30. *Each eigenspace is a subspace of $\mathbf{V}$.*

Proof. Let $\mathbf{u}, \mathbf{w} \in \mathbf{V}$ be in the eigenspace of λ. Let $a, b \in F$ be arbitrary scalars. Then we have

$$f(a\mathbf{u} + b\mathbf{w}) = af(\mathbf{u}) + bf(\mathbf{w}) = a\lambda\mathbf{u} + b\lambda\mathbf{w} = \lambda(a\mathbf{u} + b\mathbf{w}).$$

$\square$

Obviously if λ_1 and λ_2 are two different ($\lambda_1 \neq \lambda_2$) eigenvalues, then the only common element of the eigenspaces is the zero vector $\mathbf{0}$. Thus if every vector in $\mathbf{V}$ is an eigenvector, then we have the situation of theorem 29. One very particular case is that we have n different eigenvalues, where n is the dimension of $\mathbf{V}$.

Theorem 31. *Let $\lambda_1, \dots, \lambda_n$ be eigenvalues of the linear mapping $f : \mathbf{V} \to \mathbf{V}$, where $\lambda_i \neq \lambda_j$ for $i \neq j$. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors to these eigenvalues. That is, $\mathbf{v}_i \neq \mathbf{0}$ and $f(\mathbf{v}_i) = \lambda_i\mathbf{v}_i$, for each $i = 1, \dots, n$. Then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent.*

Proof. Assume to the contrary that there exist $a_1, \dots, a_n$, not all zero, with

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

Assume further that as few of the a_i as possible are non-zero. Let a_p be the first non-zero scalar. That is, $a_i = 0$ for $i < p$, and $a_p \neq 0$. Obviously some other a_k is non-zero, for some $k \neq p$, for otherwise we would have the equation $\mathbf{0} = a_p\mathbf{v}_p$, which would imply that $\mathbf{v}_p = \mathbf{0}$, contrary to the assumption that $\mathbf{v}_p$ is an eigenvector. Therefore we have

$$\mathbf{0} = f(\mathbf{0}) = f\left(\sum_{i=1}^{n} a_i\mathbf{v}_i\right) = \sum_{i=1}^{n} a_i f(\mathbf{v}_i) = \sum_{i=1}^{n} a_i\lambda_i\mathbf{v}_i.$$

Also

$$\mathbf{0} = \lambda_p\mathbf{0} = \lambda_p\left(\sum_{i=1}^{n} a_i\mathbf{v}_i\right).$$

Therefore

$$\mathbf{0} = \mathbf{0} - \mathbf{0} = \lambda_p\left(\sum_{i=1}^{n} a_i\mathbf{v}_i\right) - \sum_{i=1}^{n} a_i\lambda_i\mathbf{v}_i = \sum_{i=1}^{n} a_i(\lambda_p - \lambda_i)\mathbf{v}_i.$$

But, remembering that $\lambda_i \neq \lambda_j$ for $i \neq j$, we see that the scalar term for $\mathbf{v}_p$ is zero, yet all other non-zero scalar terms remain non-zero. Thus we have found a new sum with fewer non-zero scalars than in the original sum with the a_is. This is a contradiction. $\square$

Therefore, in this particular case, the given set of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ form a basis for $\mathbf{V}$. With respect to this basis, the matrix of the mapping is *diagonal*, with the diagonal elements being the eigenvalues.

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

12 The Elementary Matrices

These are $n \times n$ matrices which we denote by S_{ij}, $S_i(a)$, and $S_{ij}(c)$. They are such that when any $n \times n$ matrix A is multiplied on the *right* by such an S, then the given elementary *column* operation is performed on the matrix A. Furthermore, if the matrix A is multiplied on the *left* by such an elementary matrix, then the given *row* operation on the matrix is performed. It is a simple matter to verify that the following matrices are the ones we are looking for.

$$S_{ij} = \begin{pmatrix}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & 0 & & \\
 & & 1 & & & & & & & & \\
 & & & 0 & & \xrightarrow{i\text{–th row}} & & 1 & & & \\
 & & & & 1 & & & & & & \\
 & & & \downarrow & & \ddots & & \uparrow & & & \\
 & & & & & & 1 & & & & \\
 & & & 1 & & \xleftarrow[j\text{–th row}]{} & & 0 & & & \\
 & & & & & & & & 1 & & \\
 & & 0 & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{pmatrix}$$

Here, everything is zero except for the two elements at the positions ij and ji, which have the value 1. Also the diagonal elements are all 1 except for the elements at ii and jj, which are zero.

Then we have

$$S_i(a) = \begin{pmatrix}
1 & & & & & & \\
 & \ddots & & & & 0 & \\
 & & 1 & & & & \\
 & & & a & & & \\
 & & & & 1 & & \\
 & 0 & & & & \ddots & \\
 & & & & & & 1
\end{pmatrix}$$

That is, $S_i(a)$ is a diagonal matrix, all of whose diagonal elements are 1 except for the single element at the position ii, which has the value a.

Finally we have

$$S_{ij}(c) = \begin{pmatrix} 1 & & & & & & & & 0 \\ & \ddots & & & & & & & \\ & & 1 & & \xrightarrow{i-\text{th row}} & & c & & \\ & & & 1 & & & & & \\ & & & & \ddots & & \uparrow & j-\text{th column} & \\ & & & & & 1 & & & \\ & & & & & & 1 & & \\ & 0 & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix}$$

So this is again just the $n \times n$ identity matrix, but this time we have replaced the zero in the ij-th position with the scalar c. It is an elementary exercise to see that:

Theorem 32. *Each of the $n \times n$ elementary matrices are regular.*

And thus we can prove that these elementary matrices *generate* the group $GL(n, F)$. Furthermore, for every elementary matrix, the inverse matrix is again elementary.

Theorem 33. *Every matrix in $GL(n, F)$ can be represented as a product of elementary matrices.*

Proof. Let $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in GL(n, F)$ be some arbitrary regular matrix. We have already seen that A can be transformed into a matrix in step form by means of elementary row operations. That is, there is some sequence of elementary matrices: $S_1, \ldots, S_p$, such that the product

$$A^* = S_p \cdots S_1 A$$

is an $n \times n$ matrix in step form. However, since A was a *regular* matrix, the number of steps must be equal to n. That is, A^* must be a *triangular* matrix whose diagonal elements are all equal to 1.

$$A^* = \begin{pmatrix} 1 & a^*_{12} & a^*_{13} & \star & \cdots & \star & a^*_{1n} \\ 0 & 1 & a^*_{23} & \star & \cdots & \star & a^*_{2n} \\ 0 & 0 & 1 & \star & \cdots & \star & a^*_{3n} \\ \vdots & \vdots & & \ddots & \star & \star & \vdots \\ 0 & & \cdots & 0 & 1 & a^*_{(n-2)(n-1)} & a^*_{(n-2)n} \\ 0 & & & \cdots & 0 & 1 & a^*_{(n-1)n} \\ 0 & & & \cdots & & 0 & 1 \end{pmatrix}$$

But now it is obvious that the elements above the diagonal can all be reduced to zero by elementary row operations of type $S_{ij}(c)$. These row operations can again be realized by multiplication of A^* on the right by some further set of elementary matrices: $S_{p+1}, \ldots, S_q$. This gives us the matrix equation

$$S_q \cdots S_{p+1} S_p \cdots S_1 A = I_n$$

or

$$A = S_1^{-1} \cdot \cdots \cdot S_p^{-1} S_{p+1}^{-1} \cdots S_q^{-1}.$$

Since the inverse of each elementary matrix is itself elementary, we have thus expressed A as a product of elementary matrices. □

This proof also shows how we can go about programming a computer to calculate the inverse of an invertible matrix. Namely, through the process of Gauss elimination, we convert the given matrix into the identity matrix I_n. During this process, we keep multiplying together the elementary matrices which represent the respective row operations. In the end, we obtain the inverse matrix

$$A^{-1} = S_q \cdots S_{p+1} S_p \cdots S_1.$$

We also note that this is the method which can be used to obtain the value of the determinant function for the matrix. But first we must find out what the definition of determinants of matrices is!

13 The Determinant

Let $M(n \times n, F)$ be the set of all $n \times n$ matrices of elements of the field F.

Definition. *A mapping $det : M(n \times n, F) \to F$ is called a* determinant function *if it satisfies the following three conditions.*

1. *$det(I_n) = 1$, where I_n is the identity matrix.*
2. *If $A \in M(n \times n, F)$ is changed to the matrix A' by multiplying all the elements in a single row with the scalar $a \in F$, then $det(A') = a \cdot det(A)$. (This is our row operation $S_i(a)$.)*
3. *If A' is obtained from A by adding one row to a different row, then $det(A') = det(A)$. (This is our row operation $S_{ij}(1)$.)*

Simple consequences of this definition

Let $A \in M(n \times n, F)$ be an arbitrary $n \times n$ matrix, and let us say that A is transformed into the new matrix A' by an elementary row operation. Then we have:

- If A' is obtained by multiplying row i by the scalar $a \in F$, then $det(A') = a \cdot det(A)$. This is completely obvious! It is just part of the definition of "determinants".
- Therefore, if A' is obtained from A by multiplying a row with -1 then we have $det(A') = -det(A)$.

- Also, it follows that a matrix containing a row consisting of zeros must have zero as its determinant.

- If A has two identical rows, then its determinant must also be zero. For can we multiply one of these rows with -1, then add it to the other row, obtaining a matrix with a zero row.

- If A' is obtained by exchanging rows i and j, then $det(A') = -det(A)$. This is a bit more difficult to see. Let us say that $A = (\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_j, \dots \mathbf{u}_n)$, where $\mathbf{u}_k$ is the k-th row of the matrix, for each k. Then we can write

$$\begin{aligned}
det(A) &= det(\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_j, \dots \mathbf{u}_n) \\
&= det(\mathbf{u}_1, \dots, \mathbf{u}_i + \mathbf{u}_j, \dots, \mathbf{u}_j, \dots, \mathbf{u}_n) \\
&= -det(\mathbf{u}_1, \dots, -(\mathbf{u}_i + \mathbf{u}_j), \dots, \mathbf{u}_j, \dots, \mathbf{u}_n) \\
&= -det(\mathbf{u}_1, \dots, -(\mathbf{u}_i + \mathbf{u}_j), \dots, \mathbf{u}_j - (\mathbf{u}_i + \mathbf{u}_j), \dots, \mathbf{u}_n) \\
&= det(\mathbf{u}_1, \dots, \mathbf{u}_i + \mathbf{u}_j, \dots, -\mathbf{u}_i, \dots, \mathbf{u}_n) \\
&= det(\mathbf{u}_1, \dots, (\mathbf{u}_i + \mathbf{u}_j) - \mathbf{u}_i, \dots, -\mathbf{u}_i, \dots, \mathbf{u}_n) \\
&= det(\mathbf{u}_1, \dots, \mathbf{u}_j, \dots, -\mathbf{u}_i, \dots, \mathbf{u}_n) \\
&= -det(\mathbf{u}_1, \dots, \mathbf{u}_j, \dots, \mathbf{u}_i, \dots, \mathbf{u}_n)
\end{aligned}$$

  (This is the elementary row operation S_{ij}.)

- If A' is obtained from A by an elementary row operation of the form $S_{ij}(c)$, then $det(A') = det(A)$. For we have:

$$\begin{aligned}
det(A) &= det(\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_j, \dots, \mathbf{u}_n) \\
&= c^{-1}det(\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, c\mathbf{u}_j, \dots, \mathbf{u}_n) \\
&= c^{-1}det(\mathbf{u}_1, \dots, \mathbf{u}_i + c\mathbf{u}_j, \dots, c\mathbf{u}_j, \dots, \mathbf{u}_n) \\
&= det(\mathbf{u}_1, \dots, \mathbf{u}_i + c\mathbf{u}_j, \dots, \mathbf{u}_j, \dots, \mathbf{u}_n)
\end{aligned}$$

Therefore we see that each elementary row operation has a well-defined effect on the determinant of the matrix. This gives us the following algorithm for calculating the determinant of an arbitrary matrix in $M(n \times n, F)$.

How to find the determinant of a matrix

Given: An arbitrary matrix $A \in M(n \times n, F)$.
Find: $det(A)$.

Method:

1. Using elementary row operations, transform A into a matrix in step form, keeping track of the changes in the determinant at each stage.

2. If the bottom line of the matrix we obtain only consists of zeros, then the determinant is zero, and thus the determinant of the original matrix was zero.

3. Otherwise, the matrix has been transformed into an upper triangular matrix, all of whose diagonal elements are 1. But now we can transform this matrix into the identity matrix I_n by elementary row operations of the type $S_{ij}(c)$. Since we know that $det(I_n)$ must be 1, we then find a unique value for the determinant of the original matrix A. In particular, in this case $det(A) \neq 0$.

Note that in both this algorithm, as well as in the algorithm for finding the inverse of a regular matrix, the method of Gaussian elimination was used. Thus we can combine both ideas into a single algorithm, suitable for practical calculations in a computer, which yields both the matrix inverse (if it exists), and the determinant. This algorithm also proves the following theorem.

Theorem 34. *There is only one determinant function and it is uniquely given by our algorithm. Furthermore, a matrix $A \in M(n \times n, F)$ is regular if and only if $det(A) \neq 0$.*

In particular, using these methods it is easy to see that the following theorem is true.

Theorem 35. *Let A, $B \in M(n \times n, F)$. Then we have $det(A \cdot B) = det(A) \cdot det(B)$.*

Proof. If either A or B is singular, then $A \cdot B$ is singular. This can be seen by thinking about the linear mappings $\mathbf{V} \to \mathbf{V}$ which A and B represent. At least one of these mappings is singular. Thus the dimension of the image is less than n, so the dimension of the image of the composition of the two mappings must also be less than n. Therefore $A \cdot B$ must be singular. That means, on the one hand, that $det(A \cdot B) = 0$. And on the other hand, that either $det(A) = 0$ or else $det(B) = 0$. Either way, the theorem is true in this case.

If both A and B are regular, then they are both in $GL(n, F)$. Therefore, as we have seen, they can be written as products of elementary matrices. It suffices then to prove that $det(S_1)det(S_2) = det(S_1 S_2)$, where S_1 and S_2 are elementary matrices. But our arguments above show that this is, indeed, true. $\square$

Remembering that A is regular if and only if $A \in GL(n, F)$, we have:

Corollary. *If $A \in GL(n, F)$ then $det(A^{-1}) = (det(A))^{-1}$.*

In particular, if $det(A) = 1$ then we also have $det(A^{-1}) = 1$. The set of all such matrices must then form a group.

Another simple corollary is the following.

Corollary. *Assume that the matrix A is in block form, so that the linear mapping which it represents splits into a direct sum of invariant subspaces (see theorem 29). Then $det(A)$ is the product of the determinants of the blocks.*

Proof. If

$$A = \begin{pmatrix} A_1 & 0 & \dots & & 0 \\ 0 & A_2 & 0 & & \\ \vdots & 0 & \ddots & 0 & \vdots \\ & & 0 & A_{p-1} & 0 \\ 0 & & \dots & 0 & A_p \end{pmatrix}$$

then for each $i = 1, \dots, p$ let

$$A_i^* = \begin{pmatrix} 1 & 0 & \dots & & 0 \\ 0 & \ddots & 0 & & \\ \vdots & 0 & A_i & 0 & \vdots \\ & & 0 & \ddots & 0 \\ 0 & & \dots & 0 & 1 \end{pmatrix}.$$

That is, for the matrix A_i^*, all the blocks except the i-th block are replaced with identity-matrix blocks. Then $A = A_1^* \cdots A_p^*$, and it is easy to see that $det(A_i^*) = det(A_i)$ for each i. □

Definition. *The* special linear group *of order n is defined to be the set*

$$SL(n, F) = \{A \in GL(n, F) : det(A) = 1\}.$$

Theorem 36. *Let $A' = C^{-1}AC$. Then $det(A') = det(A)$.*

Proof. This follows, since $det(C^{-1}) = (det(C))^{-1}$. □

14 Leibniz Formula

Definition. *A* permutation *of the numbers $\{1, \dots, n\}$ is a bijection*

$$\sigma : \{1, \dots, n\} \to \{1, \dots, n\}.$$

The set of all permutations of the numbers $\{1, \dots, n\}$ is denoted S_n. In fact, S_n is a group: the symmetric group *of order n. Given a permutation $\sigma \in S_n$, we will say that a pair of numbers (i, j), with $i, j \in \{1, \dots, n\}$ is a "reversed pair" if $i < j$, yet $\sigma(i) > \sigma(j)$. Let $s(\sigma)$ be the total number of reversed pairs in σ. Then the* sign *of sigma is defined to be the number*

$$sign(\sigma) = (-1)^{s(\sigma)}.$$

Theorem 37 (Leibniz)**.** *Let the elements in the matrix A be a_{ij}, for i, j between 1 and n. Then we have*

$$det(A) = \sum_{\sigma \in S_n} sign(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i}.$$

As a consequence of this formula, the following theorems can be proved:

Theorem 38. *Let A be a diagonal matrix.*

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & \lambda_n \end{pmatrix}$$

Then $det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$.

Theorem 39. *Let A be a triangular matrix.*

$$\begin{pmatrix} a_{11} & a_{12} & \star & \cdots & \star \\ 0 & a_{22} & \star & \cdots & \star \\ 0 & 0 & \ddots & & \vdots \\ \vdots & & 0 & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & \cdots & 0 & 0 & a_{nn} \end{pmatrix}$$

Then $det(A) = a_{11}a_{22} \cdots a_{nn}$.

Leibniz formula also gives:

Definition. *Let $A \in M(n \times n, F)$. The* transpose *A^t of A is the matrix consisting of elements a^t_{ij} such that for all i and j we have $a^t_{ij} = a_{ji}$, where a_{ji} are the elements of the original matrix A.*

Theorem 40. $det(A^t) = det(A)$.

14.1 Special rules for 2×2 and 3×3 matrices

Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Then Leibniz formula reduces to the simple formula

$$det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

For 3×3 matrices, the formula is a little more complicated. Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Then we have

$$det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{33} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

14.2 A proof of Leibniz Formula

Let the rows of the $n \times n$ identity matrix be $\epsilon_1, \ldots, \epsilon_n$. Thus

$$\epsilon_1 = (1\ 0\ 0 \cdots 0), \quad \epsilon_2 = (0\ 1\ 0 \cdots 0), \ldots, \quad \epsilon_n = (0\ 0\ 0 \cdots 1).$$

Therefore, given that the i-th row in a matrix is

$$\xi_i = (a_{i1}\ a_{i2}\ \cdots\ a_{in}),$$

then we have

$$\xi_i = \sum_{j=1}^{n} a_{ij}\epsilon_j.$$

So let the matrix A be represented by its rows,

$$A = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}.$$

It was an exercise to show that the determinant function is additive. That is, if B and C are $n \times n$ matrices, then we have $det(B + C) = det(B) + det(C)$. Therefore we can write

$$\begin{aligned} det(A) &= det \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \\ &= \sum_{j_1=1}^{n} a_{1j_1} det \begin{pmatrix} \epsilon_{j_1} \\ \xi_2 \vdots \\ \xi_n \end{pmatrix} \\ &= \sum_{j_1=1}^{n} a_{1j_1} \sum_{j_2=1}^{n} a_{2j_2} det \begin{pmatrix} \epsilon_{j_1} \\ \epsilon_{j_2} \\ \xi_3 \\ \vdots \\ \xi_n \end{pmatrix} \\ &= \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} a_{1j_1} \cdots a_{nj_n} det \begin{pmatrix} \epsilon_{j_1} \\ \vdots \\ \epsilon_{j_n} \end{pmatrix}. \end{aligned}$$

But what is $det \begin{pmatrix} \epsilon_{j_1} \\ \vdots \\ \epsilon_{j_n} \end{pmatrix}$? To begin with, observe that if $\epsilon_{j_k} = \epsilon_{j_l}$ for some $j_k \neq j_l$, then two rows are identical, and therefore the determinant is zero. Thus we need only the sum over all possible *permutations* $(j_1, j_2, \ldots, j_n)$ of the numbers $(1, 2, \ldots, n)$. Then, given such a permutation, we have the matrix $\begin{pmatrix} \epsilon_{j_1} \\ \vdots \\ \epsilon_{j_n} \end{pmatrix}$. This can be transformed back into the identity matrix $\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$ by means of successively exchanging pairs of rows. Each time this is done, the determinant changes sign (from +1 to -1, or from -1 to +1). Finally, of course, we know that the determinant of the identity matrix is 1.

Therefore we obtain Leibniz formula

$$det(A) = \sum_{\sigma \in S_n} sign(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

15 Why is the Determinant Important?

I am sure there are many points which could be advanced in answer to this question. But here I will concentrate on only two special points.

- *The transformation formula for integrals in higher-dimensional spaces.*

  This is a theorem which is usually dealt with in the Analysis III lecture. Let $G \subset \mathbb{R}^n$ be some open region, and let $f : G \to \mathbb{R}$ be a continuous function. Then the integral

  $$\int_G f(\mathbf{x}) d^{(n)}\mathbf{x}$$

  has some particular value (assuming, of course, that the integral converges). Now assume that we have a continuously differentiable injective mapping $\phi : G \to \mathbb{R}^n$ and a continuous function $F : \phi(G) \to \mathbb{R}$. Then we have the formula

  $$\int_{\phi(G)} F(\mathbf{u}) d^{(n)}\mathbf{u} = \int_G F(\phi(\mathbf{x})) |det D(\phi(\mathbf{x}))| d^{(n)}\mathbf{x}.$$

  Here, $D(\phi(\mathbf{x}))$ is the Jacobi matrix of ϕ at the point $\mathbf{x}$.

  This formula reflects the geometric idea that the determinant measures the change of the volume of n-dimensional space under the mapping ϕ.

  If ϕ is a linear mapping, then take $Q \subset \mathbb{R}^n$ to be the unit cube: $Q = \{(x_1, \ldots, x_n) : 0 \leq x_i \leq 1, \forall i\}$. Then the volume of Q, which we can denote by $vol(Q)$ is simply 1. On the other hand, we have $vol(\phi(Q)) = det(A)$, where A is the matrix representing ϕ with respect to the canonical coordinates for $\mathbb{R}^n$. (A negative determinant — giving a negative volume — represents an orientation-reversing mapping.)

- *The characteristic polynomial.*

  Let $f : \mathbf{V} \to \mathbf{V}$ be a linear mapping, and let $\mathbf{v}$ be an eigenvector of f with $f(\mathbf{v}) = \lambda \mathbf{v}$. That means that $(f - \lambda id)(\mathbf{v}) = \mathbf{0}$; therefore the mapping $(f - \lambda id) : \mathbf{V} \to \mathbf{V}$ is singular. Now consider the matrix A, representing f with respect to some particular basis of $\mathbf{V}$. Since λI_n is the matrix representing the mapping λid, we must have that the difference $A - \lambda I_n$ is a singular matrix. In particular, we have $det(A - \lambda I_n) = 0$.

  Another way of looking at this is to take a "variable" x, and then calculate (for example, using the Leibniz formula) the polynomial in x

  $$P(x) = det(A - xI_n).$$

  This polynomial is called the *characteristic polynomial* for the matrix A. Therefore we have the theorem:

Theorem 41. *The zeros of the characteristic polynomial of A are the eigenvalues of the linear mapping $f : \mathbf{V} \to \mathbf{V}$ which A represents.*

Obviously the degree of the polynomial is n for an $n \times n$ matrix A. So let us write the characteristic polynomial in the standard form

$$P(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0.$$

The coefficients $c_0, \ldots, c_n$ are all elements of our field F.

Now the matrix A represents the mapping f with respect to a particular choice of basis for the vector space $\mathbf{V}$. With respect to some other basis, f is represented by some other matrix A', which is similar to A. That is, there exists some $C \in GL(n, F)$ with $A' = C^{-1}AC$. But we have

$$\begin{aligned} det(A' - xI_n) &= det(C^{-1}AC - xC^{-1}I_nC) \\ &= det(C^{-1}(A - xI_n)C) \\ &= det(C^{-1})det(A - xI_n)det(C) \\ &= det(A - xI_n) \\ &= P(x). \end{aligned}$$

Therefore we have:

Theorem 42. *The characteristic polynomial is invariant under a change of basis; that is, under a similarity transformation of the matrix.*

In particular, each of the coefficients c_i of the characteristic polynomial $P(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$ remains unchanged after a similarity transformation of the matrix A.

What is the coefficient c_n? Looking at the Leibniz formula, we see that the term x^n can only occur in the product

$$(a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x) = (-1)x^n - (a_{11} + a_{22} + \cdots + a_{nn})x^{n-1} + \cdots .$$

Therefore $c_n = 1$ if n is even, and $c_n = -1$ if n is odd. This is not particularly interesting.

So let us go one term lower and look at the coefficient c_{n-1}. Where does x^{n-1} occur in the Leibniz formula? Well, as we have just seen, there certainly is the term

$$(-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})x^{n-1},$$

which comes from the product of the diagonal elements in the matrix $A - xI_n$. Do any other terms also involve the power x^{n-1}? Let us look at Leibniz formula more carefully in this situation. We have

$$\begin{aligned} det(A - xI_n) &= (a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x) \\ &\quad + \sum_{\substack{\sigma \in S_n \\ \sigma \neq id}} sign(\sigma) \prod_{i=1}^{n} \left(a_{\sigma(i)i} - x\delta_{\sigma(i)i} \right) \end{aligned}$$

Here, $\delta_{ij} = 1$ if $i = j$. Otherwise, $\delta_{ij} = 0$. Now if σ is a *non-trivial* permutation — not just the identity mapping — then obviously we must have two *different* numbers i_1 and i_2, with $\sigma(i_1) \neq i_1$ and *also* $\sigma(i_2) \neq i_2$. Therefore we see that these further terms in the sum can only contribute at most $n-2$ powers of x. So we conclude that the $(n-1)$-st coefficient is

$$c_{n-1} = (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn}).$$

Definition. *Let* $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ *be an $n \times n$ matrix. The* trace *of A (in German, the* spur *of A) is the sum of the diagonal elements:*

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Theorem 43. *$tr(A)$ remains unchanged under a similarity transformation.*

An example

Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation through the angle θ. Then, with respect to the canonical basis of $\mathbb{R}^2$, the matrix of f is

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

Therefore the characteristic polynomial of A is

$$\begin{aligned} det\left[\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} - x\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] &= det\begin{pmatrix} \cos\theta - x & -\sin\theta \\ \sin\theta & \cos\theta - x \end{pmatrix} \\ &= x^2 - 2x\cos\theta + 1. \end{aligned}$$

That is to say, if $\lambda \in \mathbb{R}$ is an eigenvalue of f, then λ must be a zero of the characteristic polynomial. That is,

$$\lambda^2 - 2\lambda\cos\theta + 1 = 0.$$

But, looking at the well-known formula for the roots of quadratic polynomials, we see that such a λ can only exist if $|\cos\theta| = 1$. That is, $\theta = 0$ or π. This reflects the obvious geometric fact that a rotation through any angle other than 0 or π rotates any vector away from its original axis. In any case, the two possible values of θ give the two possible eigenvalues for f, namely $+1$ and -1.

16 Complex Numbers

On the other hand, looking at the characteristic polynomial, namely $x^2 - 2x\cos\theta + 1$ in the previous example, we see that in the case $\theta = \pm\pi$ this reduces to $x^2 + 1$. And in the realm of the complex numbers $\mathbb{C}$, this equation *does* have zeros, namely $\pm i$. Therefore we have the seemingly bizarre situation that a "complex" rotation through a quarter of a circle has vectors which are mapped back onto themselves (multiplied by plus or minus the “imaginary” number i). But there is no need for panic here! We need not follow the example of numerous famous physicists of the past, declaring the physical world to be “paradoxical”, “beyond human understanding”, etc. No. What we have here is a purely *algebraic* result using the abstract mathematical construction of the complex numbers which, in this form, has nothing to do with rotations of *real* physical space!

So let us forget physical intuition and simply enjoy thinking about the artificial mathematical game of extending the system of real numbers to the complex numbers. I assume that you all know that the set of complex numbers $\mathbb{C}$ can be thought of as being the set of numbers of the form $x + yi$, where x and y are elements of the real numbers $\mathbb{R}$ and i is an abstract symbol, introduced as a “solution” to the equation $x^2 + 1 = 0$. Thus $i^2 = -1$. Furthermore, the set of numbers of the form $x + 0 \cdot i$ can be identified simply with x, and so we have an embedding $\mathbb{R} \subset \mathbb{C}$. The rules of addition and multiplication in $\mathbb{C}$ are

$$(x_1 + y_1 i) + (x_2 + y_2 i) = (x_1 + x_2) + (y_1 + y_2)i$$

and

$$(x_1 + y_1 i) \cdot (x_2 + y_2 i) = (x_1 x_2 - y_1 y_2) + (x_1 y_2 + x_2 y_1)i.$$

Let $z = x + yi$ be some complex number. Then the *absolute value* of z is defined to be the (non-negative) real number $|z| = \sqrt{x^2 + y^2}$. The *complex conjugate* of z is $\overline{z} = x - yi$. Therefore $|z| = \sqrt{z\overline{z}}$.

It is a simple exercise to show that $\mathbb{C}$ is a field. The main result — called (in German) the Hauptsatz der Algebra — is that $\mathbb{C}$ is an *algebraically closed* field. That is, let $\mathbb{C}[z]$ be the set of all polynomials with complex numbers as coefficients. Thus, for $P(z) \in \mathbb{C}[z]$ we can write $P(z) = c_n z^n + \cdots + c_1 z + c_0$, where $c_j \in \mathbb{C}$, for all $j = 0, \ldots, n$. Then we have:

Theorem 44 (Hauptsatz der Algebra)**.** *Let $P(z) \in \mathbb{C}[z]$ be an arbitrary polynomial with complex coefficients. Then P has a zero in $\mathbb{C}$. That is, there exists some $\lambda \in \mathbb{C}$ with $P(\lambda) = 0$.*

The theory of complex numbers (Funktionentheorie in German) is an extremely interesting and pleasant subject. Complex analysis is quite different from the real analysis which you are learning in the Analysis I and II lectures. If you are interested, you might like to have a look at my lecture notes on the subject (in English), or look at any of the many books in the library with the title “Funktionentheorie”.

Unfortunately, owing to a lack of time in this summer semester, I will not be able to describe the proof of theorem 44 here. Those who are interested can find a proof in my other lecture notes on linear algebra. In any case, the consequence is

Theorem 45. *Every complex polynomial can be completely factored into linear factors. That is, for each $P(z) \in \mathbb{C}[z]$ of degree n, there exist n complex numbers (perhaps not all different) $\lambda_1, \ldots, \lambda_n$, and a further complex number c, such that*

$$P(z) = c(\lambda_1 - z) \cdots (\lambda_n - z).$$

Proof. Given $P(z)$, theorem 44 tells us that there exists some $\lambda_1 \in \mathbb{C}$, such that $P(\lambda_1) = 0$. Let us therefore divide the polynomial $P(z)$ by the polynomial $(\lambda_1 - z)$. We obtain

$$P(z) = (\lambda_1 - z) \cdot Q(z) + R(z),$$

where both $Q(z)$ and $R(z)$ are polynomials in $\mathbb{C}[z]$. However, the degree of $R(z)$ is less than the degree of the divisor, namely $(\lambda_1 - z)$, which is 1. That is, $R(z)$ must be a polynomial of degree zero, i.e. $R(z) = r \in \mathbb{C}$, a constant. But what is r? If we put λ_1 into our equation, we obtain

$$0 = P(\lambda_1) = (\lambda_1 - \lambda_1)Q(z) + r = 0 + r.$$

Therefore $r = 0$, and so

$$P(z) = (\lambda_1 - z)Q(z),$$

where $Q(z)$ must be a polynomial of degree $n-1$. Therefore we apply our argument in turn to $Q(z)$, again reducing the degree, and in the end, we obtain our factorization into linear factors. □

So the consequence is: let $\mathbf{V}$ be a vector space over the field of complex numbers $\mathbb{C}$. Then *every* linear mapping $f : \mathbf{V} \to \mathbf{V}$ has at least one eigenvalue, and thus at least one eigenvector.

17 Scalar Products, Norms, etc.

So now we have arrived at the subject matter which is usually taught in the second semester of the beginning lectures in mathematics — that is in Linear Algebra II — namely, the properties of (finite dimensional) real and complex vector spaces. Finally now, we are talking about *geometry*. That is, about vector spaces which have a *distance function*. (The word "geometry" obviously has to do with the measurement of physical distances on the earth.)

So let $\mathbf{V}$ be some finite dimensional vector space over $\mathbb{R}$, or $\mathbb{C}$. Let $\mathbf{v} \in \mathbf{V}$ be some vector in $\mathbf{V}$. Then, since $\mathbf{V} \cong \mathbb{R}^n$, or $\mathbb{C}^n$, we can write $\mathbf{v} = \sum_{j=1}^n a_j \mathbf{e}_j$, where $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is the canonical basis for $\mathbb{R}^n$ or $\mathbb{C}^n$, and $a_j \in \mathbb{R}$ or $\mathbb{C}$, respectively, for all j. Then the *length* of $\mathbf{v}$ is defined to be the non-negative real number

$$\|\mathbf{v}\| = \sqrt{|a_1|^2 + \cdots + |a_n|^2}.$$

Of course, as these things always are, we will not simply confine ourselves to measurements of normal physical things on the earth. We have already seen that the idea of a complex vector space defies our normal powers of geometric visualization. Also, we will not always restrict things to finite dimensional vector spaces. For example, spaces of functions — which are almost always infinite dimensional — are also very important in theoretical physics. Therefore, rather than saying that $\|\mathbf{v}\|$ is the "length" of the vector $\mathbf{v}$, we use a new word, and we say that $\|\mathbf{v}\|$ is the *norm* of $\mathbf{v}$. In order to define this concept in a way which is suitable for further developments, we will start with the idea of a *scalar product* of vectors.

Definition. *Let $F = \mathbb{R}$ or $\mathbb{C}$ and let $\mathbf{V}$, $\mathbf{W}$ be two vector spaces over F. A* bilinear form *is a mapping $s : \mathbf{V} \times \mathbf{W} \to F$ satisfying the following conditions with respect to arbitrary elements $\mathbf{v}$, $\mathbf{v}_1$ and $\mathbf{v}_2 \in \mathbf{V}$, $\mathbf{w}$, $\mathbf{w}_1$ and $\mathbf{w}_2 \in \mathbf{W}$, and $a \in F$.*

1. $s(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) = s(\mathbf{v}_1, \mathbf{w}) + s(\mathbf{v}_2, \mathbf{w})$,
2. $s(a\mathbf{v}, \mathbf{w}) = as(\mathbf{v}, \mathbf{w})$,
3. $s(\mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2) = s(\mathbf{v}, \mathbf{w}_1) + s(\mathbf{v}, \mathbf{w}_2)$ *and*
4. $s(\mathbf{v}, a\mathbf{w}) = as(\mathbf{v}, \mathbf{w})$.

If $\mathbf{V} = \mathbf{W}$, then we say that a bilinear form $s : \mathbf{V} \times \mathbf{V} \to F$ is symmetric, *if we always have $s(\mathbf{v}_1, \mathbf{v}_2) = s(\mathbf{v}_2, \mathbf{v}_1)$. Also the form is called* positive definite *if $s(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \neq \mathbf{0}$.*

On the other hand, if $F = \mathbb{C}$ and $f : \mathbf{V} \to \mathbf{W}$ is such that we always have

1. $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_1)$ *and*
2. $f(a\mathbf{v}) = \overline{a}f(\mathbf{v})$

Then f is a semi-linear *(not a linear) mapping. (Note: if $F = \mathbb{R}$ then semi-linear is the same as linear.)*

A mapping $s : \mathbf{V} \times \mathbf{W} \to F$ such that

1. *The mapping given by $s(\cdot, \mathbf{w}) : \mathbf{V} \to F$, where $\mathbf{v} \to s(\mathbf{v}, \mathbf{w})$ is semi-linear for all $\mathbf{w} \in \mathbf{W}$, whereas*
2. *The mapping given by $s(\mathbf{v}, \cdot) : \mathbf{W} \to F$, where $\mathbf{w} \to s(\mathbf{v}, \mathbf{w})$ is linear for all $\mathbf{v} \in \mathbf{V}$*

is called a sesqui-linear *form.*

In the case $\mathbf{V} = \mathbf{W}$, we say that the sesqui-linear form is Hermitian *(or* Euclidean, *if we only have $F = \mathbb{R}$), if we always have $s(\mathbf{v}_1, \mathbf{v}_2) = \overline{s(\mathbf{v}_2, \mathbf{v}_1)}$. (Therefore, if $F = \mathbb{R}$, an Hermitian form is symmetric.)*

Finally, a scalar product *is a positive definite Hermitian form $s : \mathbf{V} \times \mathbf{V} \to F$. Normally, one writes $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$, rather than $s(\mathbf{v}_1, \mathbf{v}_2)$.*

Well, these are a lot of new words. To be more concrete, we have the *inner products*, which are examples of scalar products.

Inner products

Let $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{C}^n$. Thus, we are considering these vectors as column vectors, defined with respect to the canonical basis of $\mathbb{C}^n$. Then define (using matrix multiplication)

$$\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\mathbf{u}}^t \mathbf{v} = (\overline{u}_1 \quad \overline{u}_2 \quad \cdots \quad \overline{u}_n) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \sum_{j=1}^{n} \overline{u}_j v_j.$$

It is easy to check that this gives a scalar product on $\mathbb{C}^n$. This particular scalar product is called the *inner product.*

Remark. *One often writes* $\mathbf{u} \cdot \mathbf{v}$ *for the inner product. Thus, considering it to be a scalar product, we just have* $\mathbf{u} \cdot \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$.

This inner product notation is often used in classical physics; in particular in Maxwell's equations. Maxwell's equations also involve the "vector product" $\mathbf{u} \times \mathbf{v}$. *However the vector product of classical physics only makes sense in 3-dimensional space. Most physicists today prefer to imagine that physical space has 10, or even more — perhaps even a frothy, undefinable number of — dimensions. Therefore it appears to be the case that the vector product might have gone out of fashion in contemporary physics. Indeed, mathematicians can imagine many other possible vector-space structures as well. Thus I shall dismiss the vector product from further discussion here.*

Definition. *A real vector space (that is, over the field of the real numbers* $\mathbb{R}$*), together with a scalar product is called a* Euclidean *vector space. A complex vector space with scalar product is called a* unitary *vector space.*

Now, the basic reason for making all these definitions is that we want to define the length — that is the norm — of the vectors in $\mathbf{V}$. Given a scalar product, then the norm of $\mathbf{v} \in \mathbf{V}$ — with respect to this scalar product — is the non-negative real number

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

More generally, one defines a norm-function on a vector space in the following way.

Definition. *Let* $\mathbf{V}$ *be a vector space over* $\mathbb{C}$ *(and thus we automatically also include the case* $\mathbb{R} \subset \mathbb{C}$ *as well). A function* $\|\cdot\| : \mathbf{V} \to \mathbb{R}$ *is called a* norm *on* $\mathbf{V}$ *if it satisfies the following conditions.*

1. $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$ *for all* $\mathbf{v} \in \mathbf{V}$ *and for all* $a \in \mathbb{C}$,
2. $\|\mathbf{v}_1 + \mathbf{v}_2\| \le \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$ *for all* $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}$ *(the* triangle inequality*), and*
3. $\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$.

Theorem 46 (Cauchy-Schwarz inequality)**.** *Let* $\mathbf{V}$ *be a Euclidean or a unitary vector space, and let* $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ *for all* $\mathbf{v} \in \mathbf{V}$. *Then we have*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\| \cdot \|\mathbf{v}\|$$

for all $\mathbf{u}$ *and* $\mathbf{v} \in \mathbf{V}$. *Furthermore, the equality* $|\langle \mathbf{u}, \mathbf{v} \rangle| = \|\mathbf{u}\| \cdot \|\mathbf{v}\|$ *holds if, and only if, the set* $\{\mathbf{u}, \mathbf{v}\}$ *is linearly* dependent.

Proof. It suffices to show that $|\langle \mathbf{u}, \mathbf{v}\rangle|^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle$. Now, if $\mathbf{v} = \mathbf{0}$, then — using the properties of the scalar product — we have both $\langle \mathbf{u}, \mathbf{v}\rangle = 0$ and $\langle \mathbf{v}, \mathbf{v}\rangle = 0$. Therefore the theorem is true in this case, and we may assume that $\mathbf{v} \neq \mathbf{0}$. Thus $\langle \mathbf{v}, \mathbf{v}\rangle > 0$. Let

$$a = \frac{\overline{\langle \mathbf{u}, \mathbf{v}\rangle}}{\langle \mathbf{v}, \mathbf{v}\rangle} \in \mathbb{C}.$$

Then we have

$$\begin{aligned} 0 &\leq \langle \mathbf{u} - a\mathbf{v}, \mathbf{u} - a\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u} - a\mathbf{v}\rangle + \langle -a\mathbf{v}, \mathbf{u} - a\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, -a\mathbf{v}\rangle + \langle -a\mathbf{v}, \mathbf{u}\rangle + \langle -a\mathbf{v}, -a\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle - \underbrace{a\langle \mathbf{u}, \mathbf{v}\rangle}_{\frac{\overline{\langle \mathbf{u}, \mathbf{v}\rangle}\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}} - \underbrace{\overline{a}\overline{\langle \mathbf{u}, \mathbf{v}\rangle}}_{\frac{\langle \mathbf{u}, \mathbf{v}\rangle\overline{\langle \mathbf{u}, \mathbf{v}\rangle}}{\langle \mathbf{v}, \mathbf{v}\rangle}} + \underbrace{a\overline{a}\langle \mathbf{v}, \mathbf{v}\rangle}_{\frac{\langle \mathbf{u}, \mathbf{v}\rangle\overline{\langle \mathbf{u}, \mathbf{v}\rangle}}{\langle \mathbf{v}, \mathbf{v}\rangle}}. \end{aligned}$$

Therefore,

$$0 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle\overline{\langle \mathbf{u}, \mathbf{v}\rangle}.$$

But

$$\langle \mathbf{u}, \mathbf{v}\rangle\overline{\langle \mathbf{u}, \mathbf{v}\rangle} = |\langle \mathbf{u}, \mathbf{v}\rangle|^2,$$

which gives the Cauchy-Schwarz inequality. When do we have equality?

If $\mathbf{v} = \mathbf{0}$ then, as we have already seen, the equality $|\langle \mathbf{u}, \mathbf{v}\rangle| = \|\mathbf{u}\| \cdot \|\mathbf{v}\|$ is trivially true. On the other hand, when $\mathbf{v} \neq \mathbf{0}$, then equality holds when $\langle \mathbf{u} - a\mathbf{v}, \mathbf{u} - a\mathbf{v}\rangle = 0$. But since the scalar product is positive definite, this holds when $\mathbf{u} - a\mathbf{v} = \mathbf{0}$. So in this case as well, $\{\mathbf{u}, \mathbf{v}\}$ is linearly dependent. □

Theorem 47. *Let* $\mathbf{V}$ *be a vector space with scalar product, and define the non-negative function* $\|\cdot\| : \mathbf{V} \to \mathbb{R}$ *by* $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$. *Then* $\|\cdot\|$ *is a norm function on* $\mathbf{V}$.

Proof. The first and third properties in our definition of norms are obviously satisfied. As far as the triangle inequality is concerned, begin by observing that for arbitrary complex numbers $z = x + yi \in \mathbb{C}$ we have

$$z + \overline{z} = (x + yi) + (x - yi) = 2x \leq 2|x| \leq 2|z|.$$

Therefore, let $\mathbf{u}$ and $\mathbf{v} \in \mathbf{V}$ be chosen arbitrarily. Then we have

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \overline{\langle \mathbf{u}, \mathbf{v}\rangle} + \langle \mathbf{v}, \mathbf{v}\rangle \\ &\leq \langle \mathbf{u}, \mathbf{u}\rangle + 2|\langle \mathbf{u}, \mathbf{v}\rangle| + \langle \mathbf{v}, \mathbf{v}\rangle \\ &\leq \langle \mathbf{u}, \mathbf{u}\rangle + 2\|\mathbf{u}\| \cdot \|\mathbf{v}\| + \langle \mathbf{v}, \mathbf{v}\rangle \quad \text{(Cauchy-Schwarz inequality)} \\ &= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\| \cdot \|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2. \end{aligned}$$

Therefore $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$. □

18 Orthonormal Bases

Our vector space $\mathbf{V}$ is now assumed to be either Euclidean, or else unitary — that is, it is defined over either the real numbers $\mathbb{R}$, or else the complex numbers $\mathbb{C}$. In either case we have a scalar product $\langle \cdot, \cdot \rangle : \mathbf{V} \times \mathbf{V} \to F$ (here, $F = \mathbb{R}$ or $\mathbb{C}$).

As always, we assume that $\mathbf{V}$ is finite dimensional, and thus it has a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Thinking about the *canonical* basis for $\mathbb{R}^n$ or $\mathbb{C}^n$, and the *inner product* as our scalar product, we see that it would be nice if we had

- $\langle \mathbf{v}_j, \mathbf{v}_j \rangle = 1$, for all j (that is, the basis vectors are *normalized*), and furthermore
- $\langle \mathbf{v}_j, \mathbf{v}_k \rangle = 0$, for all $j \neq k$ (that is, the basis vectors are an *orthogonal* set in $\mathbf{V}$).[9]

That is to say, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an *orthonormal basis* of $\mathbf{V}$. Unfortunately, most bases are not orthonormal. But this doesn't really matter. For, starting from any given basis, we can successively alter the vectors in it, gradually changing it into an orthonormal basis. This process is often called the Gram-Schmidt orthonormalization process. But first, to show you why orthonormal bases are good, we have the following theorem.

Theorem 48. *Let $\mathbf{V}$ have the orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, and let $\mathbf{x} \in \mathbf{V}$ be arbitrary. Then*

$$\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{v}_j, \mathbf{x} \rangle \mathbf{v}_j.$$

That is, the coefficients of $\mathbf{x}$, with respect to the orthonormal basis, are simply the scalar products with the respective basis vectors.

Proof. This follows simply because if $\mathbf{x} = \sum_{j=1}^{n} a_j \mathbf{v}_j$, then we have for each k,

$$\langle \mathbf{v}_k, \mathbf{x} \rangle = \langle \mathbf{v}_k, \sum_{j=1}^{n} a_j \mathbf{v}_j \rangle = \sum_{j=1}^{n} a_j \langle \mathbf{v}_k, \mathbf{v}_j \rangle = a_k.$$

□

So now to the Gram-Schmidt process. To begin with, if a non-zero vector $\mathbf{v} \in \mathbf{V}$ is not normalized — that is, its norm is not one — then it is easy to multiply it by a

[9] Note that *any* orthogonal set of non-zero vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ in $\mathbf{V}$ is linearly independent. This follows because if

$$\mathbf{0} = \sum_{j=1}^{m} a_j \mathbf{u}_j$$

then

$$0 = \langle \mathbf{u}_k, \mathbf{0} \rangle = \langle \mathbf{u}_k, \sum_{j=1}^{m} a_j \mathbf{u}_j \rangle = \sum_{j=1}^{m} a_j \langle \mathbf{u}_k, \mathbf{u}_j \rangle = a_k \langle \mathbf{u}_k, \mathbf{u}_k \rangle$$

since $\langle \mathbf{u}_k, \mathbf{u}_j \rangle = 0$ if $j \neq k$, and otherwise it is not zero. Thus, we must have $a_k = 0$. This is true for *all* the a_k.

scalar, changing it into a vector with norm one. For we have $\langle \mathbf{v}, \mathbf{v} \rangle > 0$. Therefore $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} > 0$ and we have

$$\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \sqrt{\left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle} = \sqrt{\frac{\langle \mathbf{v}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}} = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1.$$

In other words, we simply multiply the vector by the inverse of its norm.

Theorem 49. *Every finite dimensional vector space* $\mathbf{V}$ *which has a scalar product has an orthonormal basis.*

Proof. The proof proceeds by constructing an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ from a given, arbitrary basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. To describe the construction, we use induction on the dimension, n. If $n = 1$ then there is almost nothing to prove. Any non-zero vector is a basis for $\mathbf{V}$, and as we have seen, it can be normalized by dividing by the norm. (That is, scalar multiplication with the inverse of the norm.)

So now assume that $n \geq 2$, and furthermore assume that the Gram-Schmidt process can be constructed for any $n-1$ dimensional space. Let $\mathbf{U} \subset \mathbf{V}$ be the subspace spanned by the first $n-1$ basis vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$. Since $\mathbf{U}$ is only $n-1$ dimensional, our assumption is that there exists an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ for $\mathbf{U}$. Clearly[10], adding in $\mathbf{v}_n$ gives a new basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{v}_n\}$ for $\mathbf{V}$. Unfortunately, this last vector, $\mathbf{v}_n$, might disturb the nice orthonormal character of the other vectors. Therefore, we *replace* $\mathbf{v}_n$ with the new vector[11]

$$\mathbf{u}_n^* = \mathbf{v}_n - \sum_{j=1}^{n-1} \langle \mathbf{u}_j, \mathbf{v}_n \rangle \mathbf{u}_j.$$

Thus the new set $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{u}_n^*\}$ is a basis of $\mathbf{V}$. Also, for $k < n$, we have

$$\begin{aligned} \langle \mathbf{u}_k, \mathbf{u}_n^* \rangle &= \left\langle \mathbf{u}_k, \mathbf{v}_n - \sum_{j=1}^{n-1} \langle \mathbf{u}_j, \mathbf{v}_n \rangle \mathbf{u}_j \right\rangle \\ &= \langle \mathbf{u}_k, \mathbf{v}_n \rangle - \sum_{j=1}^{n-1} \langle \mathbf{u}_j, \mathbf{v}_n \rangle \langle \mathbf{u}_k, \mathbf{u}_j \rangle \\ &= \langle \mathbf{u}_k, \mathbf{v}_n \rangle - \langle \mathbf{u}_k, \mathbf{v}_n \rangle = 0. \end{aligned}$$

Thus the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{u}_n^*\}$ is orthogonal. Perhaps $\mathbf{u}_n^*$ is not normalized, but as we have seen, this can be easily changed by taking the normalized vector

$$\mathbf{u}_n = \frac{\mathbf{u}_n^*}{\|\mathbf{u}_n^*\|}.$$

□

[10]Since both $\{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ are bases for $\mathbf{U}$, we can write each $\mathbf{v}_j$ as a linear combination of the $\mathbf{u}_k$'s. Therefore $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{v}_n\}$ spans $\mathbf{V}$, and since the dimension is n, it must be a basis.

[11]A linearly independent set remains linearly independent if one of the vectors has some linear combination of the *other* vectors added on to it.

19 Some "Classical Groups" Often Seen in Physics

- The orthogonal group $O(n)$: This is the set of all linear mappings $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $\langle \mathbf{u}, \mathbf{v} \rangle = \langle f(\mathbf{u}), f(\mathbf{v}) \rangle$, for all $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$. We think of this as being all possible rotations and inversions (Spiegelungen) of n-dimensional Euclidean space.

- The special orthogonal group $SO(n)$: This is the subgroup of $O(n)$, containing all orthogonal mappings whose matrices have determinant $+1$.

- The unitary group $U(n)$: The analog of $O(n)$, where the vector space is n-dimensional complex space $\mathbb{C}^n$. That is, $\langle \mathbf{u}, \mathbf{v} \rangle = \langle f(\mathbf{u}), f(\mathbf{v}) \rangle$, for all $\mathbf{u}$, $\mathbf{v} \in \mathbb{C}^n$.

- The special unitary group $SU(n)$: Again, the subgroup of $U(n)$ with determinant $+1$.

Note that for orthogonal, or unitary mappings, all eigenvalues — if they exist — must have absolute value 1. To see this, let $\mathbf{v}$ be an eigenvector with eigenvalue λ. Then we have

$$\langle \mathbf{v}, \mathbf{v} \rangle = \langle f(\mathbf{v}), f(\mathbf{v}) \rangle = \langle \lambda \mathbf{v}, \lambda \mathbf{v} \rangle = \overline{\lambda} \lambda \langle \mathbf{v}, \mathbf{v} \rangle = |\lambda|^2 \langle \mathbf{v}, \mathbf{v} \rangle.$$

Since $\mathbf{v}$ is an eigenvector, and thus $\mathbf{v} \neq \mathbf{0}$, we must have $|\lambda| = 1$.

We will prove that all *unitary* matrices can be diagonalized. That is, for every unitary mapping $\mathbb{C}^n \to \mathbb{C}^n$, there exists a basis consisting of eigenvectors. On the other hand, as we have already seen in the case of simple rotations of 2-dimensional space, "most" orthogonal matrices cannot be diagonalized. On the other hand, we can prove that every orthogonal mapping $\mathbb{R}^n \to \mathbb{R}^n$, where n is an *odd* number, has at least one eigenvector.[12]

- The self-adjoint mappings f (of $\mathbb{R}^n \to \mathbb{R}^n$ or $\mathbb{C}^n \to \mathbb{C}^n$) are such that $\langle \mathbf{u}, f(\mathbf{v}) \rangle = \langle f(\mathbf{v}), \mathbf{u} \rangle$, for all $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, respectively. As we will see, the matrices for such mappings are symmetric in the real case, and *Hermitian* in the complex case. In either case, the matrices can be diagonalized. Examples of Hermitian matrices are the Pauli spin-matrices:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We also have the *Lorentz group*, which is important in the Theory of Relativity. Let us imagine that physical space is $\mathbb{R}^4$, and a typical point is $\mathbf{v} = (t_v, x_v, y_v, z_v)$. Physicists call this *Minkowski space*, which they often denote by M^4. A linear mapping $f : M^4 \to M^4$ is called a *Lorentz transformation* if, for $f(\mathbf{v}) = (t_v^*, x_v^*, y_v^*, z_v^*)$, we have

- $-(t_v^*)^2 + (x_v^*)^2 + (y_v^*)^2 + (z_v^*)^2 = -t_v^2 + x_v^2 + y_v^2 + z_v^2$, for all $\mathbf{v} \in M^4$, and also

[12] For example, in our normal 3-dimensional space of physical reality, any rotating object — for example the Earth rotating in space — has an axis of rotation, which is an eigenvector.

- the mapping is "time-preserving" in the sense that the unit vector in the time direction, $(1,0,0,0)$ is mapped to some vector (t^*, x^*, y^*, z^*), such that $t^* > 0$.

The *Poincare group* is obtained if we consider, in addition, translations of Minkowski space. But translations are not linear mappings, so I will not consider these things further in this lecture.

20 Characterizing Orthogonal, Unitary, and Hermitian Matrices

20.1 Orthogonal matrices

Let $\mathbf{V}$ be an n-dimensional real vector space (that is, over the real numbers $\mathbb{R}$), and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbf{V}$. Let $f : \mathbf{V} \to \mathbf{V}$ be an orthogonal mapping, and let A be its matrix with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Then we say that A is an *orthogonal matrix.*

Theorem 50. *The $n \times n$ matrix A is orthogonal $\Leftrightarrow A^{-1} = A^t$. (Recall that if a_{ij} is the ij-th element of A, then the ij-the element of A^t is a_{ji}. That is, everything is "flipped over" the main diagonal in A.)*

Proof. For an orthogonal mapping f, we have $\langle \mathbf{u}, \mathbf{w} \rangle = \langle f(\mathbf{u}), f(\mathbf{w}) \rangle$, for all j and k. But in the matrix notation, the scalar product becomes the inner product. That is, if

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix},$$

then

$$\langle \mathbf{u}, \mathbf{w} \rangle = \mathbf{u}^t \cdot \mathbf{w} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \sum_{j=1}^{n} u_j w_j.$$

In particular, taking $\mathbf{u} = \mathbf{v}_j$ and $\mathbf{w} = \mathbf{v}_k$, we have

$$\langle \mathbf{v}_j, \mathbf{v}_k \rangle = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

In other words, the matrix whose jk-th element is always $\langle \mathbf{v}_j, \mathbf{v}_k \rangle$ is the $n \times n$ identity matrix I_n. On the other hand,

$$f(\mathbf{v}_j) = A\mathbf{v}_j = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \cdot \mathbf{v}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

That is, we obtain the j-th column of the matrix A. Furthermore, since $\langle \mathbf{v}_j, \mathbf{v}_k \rangle = \langle f(\mathbf{v}_j), f(\mathbf{v}_k) \rangle$, we must have the matrix whose jk-th elements are $\langle f(\mathbf{v}_j), f(\mathbf{v}_k) \rangle$

being again the identity matrix. So

$$(a_{1j} \quad \cdots \quad a_{nj}) \begin{pmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix} = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

But now, if you think about it, you see that this is just one part of the matrix multiplication $A^t A$. All together, we have

$$A^t A = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = I_n.$$

Thus we conclude that $A^{-1} = A^t$. (Note: this was only the proof that f orthogonal $\Rightarrow A^{-1} = A^t$. The proof in the other direction, going backwards through our argument, is easy, and is left as an exercise for you.) □

20.2 Unitary matrices

Theorem 51. *The $n \times n$ matrix A is unitary $\Leftrightarrow A^{-1} = \overline{A}^t$. (The matrix $\overline{A}$ is obtained by taking the complex conjugates of all its elements.)*

Proof. Entirely analogous with the case of orthogonal matrices. One must note however, that the inner product in the complex case is

$$\langle \mathbf{u}, \mathbf{w} \rangle = \overline{\mathbf{u}}^t \cdot \mathbf{w} = (\overline{u}_1 \quad \cdots \quad \overline{u}_n) \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \sum_{j=1}^{n} \overline{u}_j w_j.$$

□

20.3 Hermitian and symmetric matrices

Theorem 52. *The $n \times n$ matrix A is Hermitian $\Leftrightarrow A = \overline{A}^t$.*

Proof. This is again a matter of translating the condition $\langle \mathbf{v}_j, f(\mathbf{v}_k) \rangle = \langle f(\mathbf{v}_j), \mathbf{v}_k \rangle$ into matrix notation, where f is the linear mapping which is represented by the matrix A, with respect to the orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. We have

$$\langle \mathbf{v}_j, f(\mathbf{v}_k) \rangle = \overline{\mathbf{v}}_j^t \cdot A\mathbf{v}_k = \overline{\mathbf{v}}_j^t \begin{pmatrix} a_{1k} \\ \vdots \\ a_{n_k} \end{pmatrix} = a_{jk}.$$

On the other hand

$$\langle f(\mathbf{v}_j), \mathbf{v}_k \rangle = \overline{A\mathbf{v}_j}^t \cdot \mathbf{v}_k = (\overline{a}_{1j} \quad \cdots \quad \overline{a}_{nj}) \cdot \mathbf{v}_k = \overline{a}_{kj}.$$

□

In particular, we see that in the real case, self-adjoint matrices are symmetric.

21 Which Matrices can be Diagonalized?

The complete answer to this question is a bit too complicated for me to explain to you in the short time we have in this semester. It all has to do with a thing called the “minimal polynomial”.

Now we have seen that not all *orthogonal* matrices can be diagonalized. (Think about the rotations of $\mathbb{R}^2$.) On the other hand, we can prove that all *unitary*, and also all *Hermitian* matrices can be diagonalized.

Of course, a matrix M is only a representation of a linear mapping $f : \mathbf{V} \to \mathbf{V}$ with respect to a given basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of the vector space $\mathbf{V}$. So the idea that the matrix can be diagonalized is that it is similar to a diagonal matrix. That is, there exists another matrix S, such that $S^{-1}MS$ is diagonal.

$$S^{-1}MS = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

But this means that there must be a basis for $\mathbf{V}$, consisting entirely of eigenvectors.

In this section we will consider complex vector spaces — that is, $\mathbf{V}$ is a vector space over the complex numbers $\mathbb{C}$. The vector space $\mathbf{V}$ will be assumed to have a scalar product associated with it, and the bases we consider will be orthonormal.

We begin with a definition.

Definition. *Let* $\mathbf{W} \subset \mathbf{V}$ *be a subspace of* $\mathbf{V}$. *Let*

$$\mathbf{W}^\perp = \{\mathbf{v} \in \mathbf{V} : \langle \mathbf{v}, \mathbf{w} \rangle = 0, \forall \mathbf{w} \in \mathbf{W}\}.$$

Then $\mathbf{W}^\perp$ *is called the* perpendicular space *to* $\mathbf{W}$.

It is a rather trivial matter to verify that $\mathbf{W}^\perp$ is itself a subspace of $\mathbf{V}$, and furthermore $\mathbf{W} \cap \mathbf{W}^\perp = \{\mathbf{0}\}$. In fact, we have:

Theorem 53. $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$.

Proof. Let $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ be some orthonormal basis for the vector space $\mathbf{W}$. This can be extended to a basis $\{\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{w}_{m+1}, \dots, \mathbf{w}_n\}$ of $\mathbf{V}$. Assuming the Gram-Schmidt process has been used, we may assume that this is an orthonormal basis. The claim is then that $\{\mathbf{w}_{m+1}, \dots, \mathbf{w}_n\}$ is a basis for $\mathbf{W}^\perp$.

Now clearly, since $\langle \mathbf{w}_j, \mathbf{w}_k \rangle = 0$, for $j \neq k$, we have that $\{\mathbf{w}_{m+1}, \dots, \mathbf{w}_n\} \subset \mathbf{W}^\perp$. If $\mathbf{u} \in \mathbf{W}^\perp$ is some arbitrary vector in $\mathbf{W}^\perp$, then we have

$$\mathbf{u} = \sum_{j=1}^{n} \langle \mathbf{w}_j, \mathbf{u} \rangle \mathbf{w}_j = \sum_{j=m+1}^{n} \langle \mathbf{w}_j, \mathbf{u} \rangle \mathbf{w}_j,$$

since $\langle \mathbf{w}_j, \mathbf{u} \rangle = 0$ if $j \leq m$. (Remember, $\mathbf{u} \in \mathbf{W}^\perp$.) Therefore, $\{\mathbf{w}_{m+1}, \dots, \mathbf{w}_n\}$ is a linearly independent, orthonormal set which generates $\mathbf{W}^\perp$, so it is a basis. And so we have $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$. □

Theorem 54. *Let $f : \mathbf{V} \to \mathbf{V}$ be a unitary mapping ($\mathbf{V}$ is a vector space over the complex numbers $\mathbb{C}$). Then there exists an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for $\mathbf{V}$ consisting of eigenvectors under f. That is to say, the matrix of f with respect to this basis is a diagonal matrix.*

Proof. If the dimension of $\mathbf{V}$ is zero or one, then obviously there is nothing to prove. So let us assume that the dimension n is at least two, and we prove things by induction on the number n. That is, we assume that the theorem is true for spaces of dimension less than n.

Now, according to the fundamental theorem of algebra, the characteristic polynomial of f has a zero, λ say, which is then an eigenvalue for f. So there must be some non-zero vector $\mathbf{v}_n \in \mathbf{V}$, with $f(\mathbf{v}_n) = \lambda \mathbf{v}_n$. By dividing by the norm of $\mathbf{v}_n$ if necessary, we may assume that $\|\mathbf{v}_n\| = 1$.

Let $\mathbf{W} \subset \mathbf{V}$ be the 1-dimensional subspace generated by the vector $\mathbf{v}_n$. Then $\mathbf{W}^\perp$ is an $n-1$ dimensional subspace. We have that $\mathbf{W}^\perp$ is invariant under f. That is, if $\mathbf{u} \in \mathbf{W}^\perp$ is some arbitrary vector, then $f(\mathbf{u}) \in \mathbf{W}^\perp$ as well. This follows since

$$\lambda\langle f(\mathbf{u}), \mathbf{v}_n\rangle = \langle f(\mathbf{u}), \lambda\mathbf{v}_n\rangle = \langle f(\mathbf{u}), f(\mathbf{v}_n)\rangle = \langle \mathbf{u}, \mathbf{v}_n\rangle = 0.$$

But we have already seen that for an eigenvalue λ of a unitary mapping, we must have $|\lambda| = 1$. Therefore we must have $\langle f(\mathbf{u}), \mathbf{v}_n\rangle = 0$.

So we can consider f, restricted to $\mathbf{W}^\perp$, and using the inductive hypothesis, we obtain an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$ for $\mathbf{W}^\perp$. Therefore, adding in the last vector $\mathbf{v}_n$, we have an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for $\mathbf{V}$. □

Theorem 55. *All Hermitian matrices can be diagonalized.*

Proof. This is similar to the last one. Again, we use induction on n, the dimension of the vector space $\mathbf{V}$. We have a self-adjoint mapping $f : \mathbf{V} \to \mathbf{V}$. If n is zero or one, then we are finished. Therefore we assume that $n \geq 2$.

Again, we observe that the characteristic polynomial of f must have a zero, hence there exists some eigenvalue λ, and an eigenvector $\mathbf{v}_n$ of f, which has norm equal to one, where $f(\mathbf{v}_n) = \lambda\mathbf{v}_n$. Again take $\mathbf{W}$ to be the one dimensional subspace of $\mathbf{V}$ generated by $\mathbf{v}_n$. Let $\mathbf{W}^\perp$ be the perpendicular subspace. It is only necessary to show that, again, $\mathbf{W}^\perp$ is invariant under f. But this is easy. Let $\mathbf{u} \in \mathbf{W}^\perp$ be given. Then we have

$$\langle f(\mathbf{u}), \mathbf{v}_n\rangle = \langle \mathbf{u}, f(\mathbf{v}_n)\rangle = \langle \mathbf{u}, \lambda\mathbf{v}_n\rangle = \lambda\langle \mathbf{u}, \mathbf{v}_n\rangle = \lambda \cdot 0 = 0.$$

The rest of the proof follows as before. □

In the particular case where we have only real numbers (which of course are a subset of the complex numbers), then we have a symmetric matrix.

Corollary. *All real symmetric matrices can be diagonalized.*

Note furthermore, that even in the case of a unitary matrix, the symmetry condition, namely $a_{jk} = \overline{a}_{kj}$, implies that on the diagonal, we have $a_{jj} = \overline{a}_{jj}$ for all j. That is, the diagonal elements are all real numbers. But these are the eigenvalues. Therefore we have:

Corollary. *The eigenvalues of a self-adjoint matrix — that is, a symmetric or a Hermitian matrix — are all real numbers.*

Orthogonal matrices revisited

Let A be an $n \times n$ orthogonal matrix. That is, it consists of real numbers, and we have $A^t = A^{-1}$. In general, it cannot be diagonalized. But on the other hand, it can be brought into the following form by means of similarity transformations.

$$A'' = \begin{pmatrix} \pm 1 & & & & & \\ & \ddots & & & 0 & \\ & & \pm 1 & & & \\ & & & R_1 & & \\ & 0 & & & \ddots & \\ & & & & & R_p \end{pmatrix},$$

where each R_j is a 2×2 block of the form

$$\begin{pmatrix} \cos\theta & \pm\sin\theta \\ \sin\theta & \mp\cos\theta \end{pmatrix}.$$

To see this, start by imagining that A represents the orthogonal mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ with respect to the canonical basis of $\mathbb{R}^n$. Now consider the *symmetric* matrix

$$B = A + A^t = A + A^{-1}.$$

This matrix represents another linear mapping, call it $g : \mathbb{R}^n \to \mathbb{R}^n$, again with respect to the canonical basis of $\mathbb{R}^n$.

But, as we have just seen, B *can* be diagonalized. In particular, there exists some vector $\mathbf{v} \in \mathbb{R}^n$ with $g(\mathbf{v}) = \lambda g(\mathbf{v})$, for some $\lambda \in \mathbb{R}$. We now proceed by induction on the number n. There are two cases to consider:

- $\mathbf{v}$ is also an eigenvector for f, or
- it isn't.

The first case is easy. Let $\mathbf{W} \subset \mathbf{V}$ be simply $\mathbf{W} = [\mathbf{v}]$. i.e. this is just the set of all scalar multiples of $\mathbf{v}$. Let $\mathbf{W}^\perp$ be the perpendicular space to $\mathbf{W}$. (That is, $\mathbf{w} \in \mathbf{W}^\perp$ means that $\langle \mathbf{w}, \mathbf{v} \rangle = 0$.) But it is easy to see that $\mathbf{W}^\perp$ is also invarient under f. This follows by observing first of all that $f(\mathbf{v}) = \alpha\mathbf{v}$, with $\alpha = \pm 1$. (Remember that the eigenvalues of orthogonal mappings have absolute value 1.) Now take $\mathbf{w} \in \mathbf{W}^\perp$. Then $\langle f(\mathbf{w}), \mathbf{v} \rangle = \alpha^{-1}\langle f(\mathbf{w}), \alpha\mathbf{v} \rangle = \alpha^{-1}\langle f(\mathbf{w}), f(\mathbf{v}) \rangle = \alpha^{-1}\langle \mathbf{w}, \mathbf{v} \rangle = \alpha^{-1} \cdot 0 = 0$. Thus, by changing the basis of $\mathbb{R}^n$ to being an orthonormal basis, starting with $\mathbf{v}$ (which we can assume has been normalized), we obtain that the original matrix is similar to the matrix

$$\begin{pmatrix} \alpha & 0 \\ 0 & A^* \end{pmatrix},$$

where A^* is an $(n-1) \times (n-1)$ orthogonal matrix, which, according to the inductive hypothesis, can be transformed into the required form.

If $\mathbf{v}$ is *not* an eigenvector of f, then, still, we know it is an eigenvector of g, and furthermore $g = f + f^{-1}$. In particular, $g(\mathbf{v}) = \lambda\mathbf{v} = f(\mathbf{v}) + f^{-1}(\mathbf{v})$. That is,

$$f(f(\mathbf{v})) = \lambda f(\mathbf{v}) - \mathbf{v}.$$

So this time, let $\mathbf{W} = [\mathbf{v}, f(\mathbf{v})]$. This is a 2-dimensional subspace of $\mathbf{V}$. Again, consider $\mathbf{W}^\perp$. We have $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$. So we must show that $\mathbf{W}^\perp$ is invarient under f. Now we have another two cases to consider:

- $\lambda = 0$, and
- $\lambda \neq 0$.

So if $\lambda = 0$ then we have $f(f(\mathbf{v})) = -\mathbf{v}$. Therefore, again taking $\mathbf{w} \in \mathbf{W}^\perp$, we have $\langle f(\mathbf{w}), \mathbf{v}\rangle = \langle f(\mathbf{w}), -f(f(\mathbf{v}))\rangle = -\langle \mathbf{w}, f(\mathbf{v})\rangle = 0$. (Remember that $\mathbf{w} \in \mathbf{W}^\perp$, so that $\langle \mathbf{w}, f(\mathbf{v})\rangle = 0$.) Of course we also have $\langle f(\mathbf{w}), f(\mathbf{v})\rangle = \langle \mathbf{w}, \mathbf{v}\rangle = 0$.

On the other hand, if $\lambda \neq 0$ then we have $\mathbf{v} = \lambda f(\mathbf{v}) - f(f(\mathbf{v}))$ so that $\langle f(\mathbf{w}), \mathbf{v}\rangle = \langle f(\mathbf{w}), \lambda f(\mathbf{v}) - f(f(\mathbf{v}))\rangle = \lambda\langle f(\mathbf{w}), f(\mathbf{v})\rangle - \langle f(\mathbf{w}), f(f(\mathbf{v}))\rangle$, and we have seen that both of these scalar products are zero. Finally, we again have $\langle f(\mathbf{w}), f(\mathbf{v})\rangle = \langle \mathbf{w}, \mathbf{v}\rangle = 0$.

Therefore we have shown that $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$, where both of these subspaces are invariant under the orthogonal mapping f. By our inductive hypothesis, there is an orthonormal basis for f restricted to the $n-2$ dimensional subspace $\mathbf{W}^\perp$ such that the matrix has the required form. As far as $\mathbf{W}$ is concerned, we are back in the simple situation of an orthogonal mapping $\mathbb{R}^2 \to \mathbb{R}^2$, and the matrix for this has the form of one of our 2×2 blocks.

22 Dual Spaces

Again let $\mathbf{V}$ be a vector space over a field F (and, although its not really necessary here, we continue to take $F = \mathbb{R}$ or $\mathbb{C}$).

Definition. *The* dual space *to* $\mathbf{V}$ *is the set of all linear mappings* $f : \mathbf{V} \to F$. *We denote the dual space by* $\mathbf{V}^*$.

Examples

- Let $V = \mathbb{R}^n$. Then let f_i be the projection onto the i-th coordinat. That is, if $\mathbf{e}_j$ is the j-th canonical basis vector, then

  $$f_i(e_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

  So each f_i is a member of $\mathbf{V}^*$, for $i = 1, \ldots, n$, and as we will see, these *dual vectors* form a basis for the dual space.

- More generally, let $\mathbf{V}$ be any finite dimensional vector space, with some basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Let $f_i : \mathbf{V} \to F$ be defined as follows. For an arbitrary vector $\mathbf{v} \in \mathbf{V}$ there is a unique linear combination

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n.$$

Then let $f_i(\mathbf{v}_i) = a_i$. Again, $f_i \in \mathbf{V}^*$, and we will see that the n vectors, $f_1, \ldots, f_n$ form a basis of the dual space.

- Let $C_0([0,1])$ be the space of continuous functions $f : [0,1] \to \mathbb{R}$. As we have seen, this is a real vector space, and it is not finite dimensional. For each $f \in C_0([0,1])$ let

$$\Lambda(f) = \int_0^1 f(x)dx.$$

This gives us a linear mapping $\Lambda : C_0([0,1]) \to \mathbb{R}$. Thus it belongs to the dual space of $C_0([0,1])$.

- Another vector in the dual space to $C_0([0,1])$ is given as follows. Let $x \in [0,1]$ be some fixed point. Then let $\Gamma_x : C_0([0,1]) \to \mathbb{R}$ is defined to be $\Gamma(f) = f(x)$, for all $f \in C_0([0,1])$.

- For this last example, let us assume that $\mathbf{V}$ is a vector space with scalar product. (Thus $F = \mathbb{R}$ or $\mathbb{C}$.) For each $\mathbf{v} \in \mathbf{V}$, let $\phi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$. Then $\phi_{\mathbf{v}} \in \mathbf{V}^*$.

Theorem 56. *Let $\mathbf{V}$ be a finite dimensional vector space (over $\mathbb{C}$) and let $\mathbf{V}^*$ be the dual space. For each $\mathbf{v} \in \mathbf{V}$, let $\phi_{\mathbf{v}} : \mathbf{V} \to \mathbb{C}$ be given by $\phi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$. Then given an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbf{V}$, we have that $\{\phi_{\mathbf{v}_1}, \ldots, \phi_{\mathbf{v}_n}\}$ is a basis of $\mathbf{V}^*$. This is called the* dual basis *to $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.*

Proof. Let $\phi \in \mathbf{V}^*$ be an arbitrary linear mapping $\phi : \mathbf{V} \to \mathbb{C}$. But, as always, we remember that ϕ is uniquely determined by vectors (which in this case are simply complex numbers) $\phi(\mathbf{v}_1), \ldots, \phi(\mathbf{v}_n)$. Say $\phi(\mathbf{v}_j) = c_j \in \mathbb{C}$, for each j. Now take some arbitrary vector $\mathbf{v} \in \mathbf{V}$. There is the unique expression

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n.$$

But then we have

$$\begin{aligned}
\phi(\mathbf{v}) &= \phi(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) \\
&= a_1\phi(\mathbf{v}_1) + \cdots + a_n\phi(\mathbf{v}_n) \\
&= a_1c_1 + \cdots + a_nc_n \\
&= c_1\phi_{\mathbf{v}_1}(\mathbf{v}) + \cdots + c_n\phi_{\mathbf{v}_n}(\mathbf{v}) \\
&= (c_1\phi_{\mathbf{v}_1} + \cdots + c_n\phi_{\mathbf{v}_n})(\mathbf{v}).
\end{aligned}$$

Therefore, $\phi = c_1\phi_{\mathbf{v}_1} + \cdots + c_n\phi_{\mathbf{v}_n}$, and so $\{\phi_{\mathbf{v}_1}, \ldots, \phi_{\mathbf{v}_n}\}$ generates $\mathbf{V}^*$.

To show that $\{\phi_{\mathbf{v}_1}, \ldots, \phi_{\mathbf{v}_n}\}$ is linearly independent, let $\phi = c_1\phi_{\mathbf{v}_1} + \cdots + c_n\phi_{\mathbf{v}_n}$ be some linear combination, where $c_j \neq 0$, for at least one j. But then $\phi(\mathbf{v}_j) = c_j \neq 0$, and thus $\phi \neq 0$ in $\mathbf{V}^*$. □

Corollary. *$dim(\mathbf{V}^*) = dim(\mathbf{V})$.*

Corollary. *More specifically, we have an isomorphism $\mathbf{V} \to \mathbf{V}^*$, such that $\mathbf{v} \to \phi_{\mathbf{v}}$ for each $\mathbf{v} \in \mathbf{V}$.*

But somehow, this isomorphism doesn't seem to be very "natural". It is defined in terms of some specific basis of $\mathbf{V}$. What if $\mathbf{V}$ is not finite dimensional so that we have no basis to work with? For this reason, we do not think of $\mathbf{V}$ and $\mathbf{V}^*$ as being "really" just the same vector space.[13]

On the other hand, let us look at the dual space of the dual space $(\mathbf{V}^*)^*$. (Perhaps this is a slightly mind-boggling concept at first sight!) We imagine that "really" we just have $(\mathbf{V}^*)^* = \mathbf{V}$. For let $\Phi \in (\mathbf{V}^*)^*$. That means, for each $\phi \in \mathbf{V}^*$ we have $\Phi(\phi)$ being some complex number. On the other hand, we also have $\phi(\mathbf{v})$ being some complex number, for each $\mathbf{V} \in \mathbf{V}$. Can we uniquely identify each $\mathbf{V} \in \mathbf{V}$ with some $\Phi \in (\mathbf{V}^*)^*$, in the sense that both always give the same complex numbers, for all possible $\phi \in \mathbf{V}^*$?

Let us say that there exists a $\mathbf{v} \in \mathbf{V}$ such that $\Phi(\phi) = \phi(\mathbf{v})$, for all $\phi \in \mathbf{V}^*$. In fact, if we *define* $\Phi_{\mathbf{v}}$ to be $\Phi(\phi) = \phi(\mathbf{v})$, for each $\phi \in \mathbf{V}^*$, then we certainly have a linear mapping, $\mathbf{V}^* \to \mathbb{C}$. On the other hand, given some *arbitrary* $\Phi \in (\mathbf{V}^*)^*$, do we have a unique $\mathbf{v} \in \mathbf{V}$ such that $\Phi(\phi) = \phi(\mathbf{v})$, for all $\phi \in \mathbf{V}^*$? At least in the case where $\mathbf{V}$ is finite dimensional, we can affirm that it is true by looking at the dual basis.

Dual mappings

Let $\mathbf{V}$ and $\mathbf{W}$ be two vector spaces (where we again assume that the field is $\mathbb{C}$). Assume that we have a linear mapping $f : \mathbf{V} \to \mathbf{W}$. Then we can define a linear mapping $f^* : \mathbf{W}^* \to \mathbf{V}^*$ in a natural way as follows. For each $\phi \in \mathbf{W}^*$, let $f^*(\phi) = \phi \circ f$. So it is obvious that $f^*(\phi) : \mathbf{V} \to \mathbb{C}$ is a linear mapping. Now assume that $\mathbf{V}$ and $\mathbf{W}$ have scalar products, giving us the mappings $s : \mathbf{V} \to \mathbf{V}^*$ and $t : \mathbf{W} \to \mathbf{W}^*$. So we can draw a little "diagram" to describe the situation.

$$\begin{array}{ccc} \mathbf{V} & \xrightarrow{f} & \mathbf{W} \\ s \downarrow & & \downarrow t \\ \mathbf{V}^* & \xleftarrow{f^*} & \mathbf{W}^* \end{array}$$

The mappings s and t are isomorphisms, so we can go around the diagram, using the mapping $f^{adj} = s^{-1} \circ f^* \circ t : \mathbf{W} \to \mathbf{V}$. This is the adjoint mapping to f. So we see that in the case $\mathbf{V} = \mathbf{W}$, we have that a self-adjoint mapping $f : \mathbf{V} \to \mathbf{V}$ is such that $f^{adj} = f$.

Does this correspond with our earlier definition, namely that $\langle \mathbf{u}, f(\mathbf{v}) \rangle = \langle f(\mathbf{u}), \mathbf{v} \rangle$ for all $\mathbf{u}$ and $\mathbf{v} \in \mathbf{V}$? To answer this question, look at the diagram, which now has the form

$$\begin{array}{ccc} \mathbf{V} & \xrightarrow{f} & \mathbf{V} \\ s \downarrow & & \downarrow s \\ \mathbf{V}^* & \xleftarrow{f^*} & \mathbf{V}^* \end{array}$$

[13] In case we have a scalar product, then there is a "natural" mapping $\mathbf{V} \to \mathbf{V}^*$, where $\mathbf{v} \to \phi_{\mathbf{v}}$, such that $\phi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$, for all $\mathbf{u} \in \mathbf{V}$.

where $s(\mathbf{v}) \in \mathbf{V}^*$ is such that $s(\mathbf{v})(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$, for all $\mathbf{u} \in \mathbf{V}$. Now $f^{adj} = s^{-1} \circ f^* \circ s$; that is, the condition $f^{adj} = f$ becomes $s^{-1} \circ f^* \circ s = f$. Since s is an isomorphism, we can equally say that the condition is that $f^* \circ s = s \circ f$. So let $\mathbf{v}$ be some arbitrary vector in $\mathbf{V}$. We have $s \circ f(\mathbf{v}) = f^* \circ s(\mathbf{v})$. However, remembering that this is an element of $\mathbf{V}^*$, we see that this means

$$(s \circ f(\mathbf{v}))(\mathbf{u}) = (f^* \circ s)(\mathbf{v})(\mathbf{u}),$$

for all $\mathbf{u} \in \mathbf{V}$. But $(s \circ f(\mathbf{v}))(\mathbf{u}) = \langle f(\mathbf{v}), \mathbf{u} \rangle$ and $(f^* \circ s)(\mathbf{v})(\mathbf{u}) = \langle \mathbf{v}, f(\mathbf{u}) \rangle$. Therefore we have

$$\langle f(\mathbf{v}), \mathbf{u} \rangle = \langle \mathbf{v}, f(\mathbf{u}) \rangle$$

for all $\mathbf{v}$ and $\mathbf{u} \in \mathbf{V}$, as expected.

23 The End

This is the end of the semester, and thus the end of what I have to say about "linear algebra in physics" here. But that is not to say that there is nothing more that you have to know about the subject. For example, when studying the theory of relativity you will encounter tensors, which are combinations of linear mappings and dual mappings. One speaks of "covariant" and "contravariant" tensors. That is, linear mappings and dual mappings.

But then, proceeding to the general theory of relativity, these tensors are used to describe differential geometry. That is, we no longer have a linear (that is, a vector) space. Instead, we imagine that space is curved, and in order to describe this curvature, we define a thing called the tangent vector space which you can think of as being a kind of linear approximation to the spacial structure near a given point. And so it goes on, leading to more and more complicated mathematical constructions, taking us away from the simple "linear" mathematics which we have seen in this semester.

After a few years of learning the mathematics of contemporary theoretical physics, perhaps you will begin to ask yourselves whether it really makes so much sense after all. Can it be that the physical world is best described by using all of the latest techniques which pure mathematicians happen to have been playing around with in the last few years — in algebraic topology, functional analysis, the theory of complex functions, and so on and so forth? Or, on the other hand, could it be that physics has been loosing touch with reality, making constructions similar to the theory of epicycles of the medieval period, whose conclusions can never be verified using practical experiments in the real world?

In his book "The Meaning of Relativity", Albert Einstein wrote

> "One can give good reasons why reality cannot at all be represented by a continuous field. From the quantum phenomena it apears to follow with certainty that a finite system of finite energy can be completely described by a finite set of numbers (quantum numbers). This does not seem to be in accordance with a continuum theory, and must lead to an attempt to find a purely algebraic theory for the description of reality. But nobody knows how to obtain the basis of such a theory."