Tune-Up Lecture Notes

Linear Algebra I

One usually first encounters a *vector* depicted as a directed line segment in Euclidean space, or what amounts to the same thing, as an ordered n-tuple of numbers of the form $\{a_i\}_{i=1,\ldots,n}$. The former is a geometrical description of a vector, whereas the latter can be interpreted as an algebraic representation with respect to a given coordinate system. However, the interplay between geometry and algebra is not transparant in more general vector spaces that could be infinite dimensional.

Finite dimensional vector spaces are relatively easy to work with, and most people have a well developed intuition for them. Infinite dimensional vector spaces can be considerably more difficult to appreciate, but they arise naturally in applications as function spaces. A function space consists of all functions from a given set X to either the real or complex numbers, and is finite dimensional precisely when the set X is a finite set. Hence the Euclidean space mentioned above can be viewed as the function space with $X = i = 1, \ldots, n$. Infinite dimensional spaces require additional concepts in order to use them effectively, as we shall see in later chapters. But there are similarities between finite and infinite spaces, as for example, solving partial differential equations by Hilbert space methods is a close analog to a simpler method employed in the study finite dimensional spaces. The objective of this chapter is to explore the basic properties of finite dimensional vector spaces from an abstract point of view so as to draw similarities and distinctions between finite versus infinite dimensional spaces.

1 Basic concepts.

A vector space V is a set of objects (called *vectors*) in association with with two operations called addition and scalar multiplication. The addition property says two vectors can be added together to form another vector, and scalar multiplication is that a vector can be multiplied by a scalar to form a new vector. These operations must satisfy some rather transparent axioms that are listed below. By a *scalar*, we mean an element of a *scalar field* $\mathbb{F}$, which in itself is a set of objects with the operations of addition and multiplication that satisfy a set of axioms. The precise statement of the field axioms is rather long and tedious, and since we shall only consider scalar fields that are either the set of real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$, we shall not explicitly state these. Suffice to say that scalars enjoy the usual rules of addition and multiplication that are familiar to all. The rational numbers $\mathbb{Q}$ is also a scalar field, but $\mathbb{Q}$ is not generally encountered in applications since it is not *complete* (more on this below). It should be noted that by referring to a vector space V, it is not just the elements of V that are relevant, but also the field $\mathbb{F}$. The terminology that V is a *real* (respectively, *complex*) vector space is used if the underlying field $\mathbb{F} = R$ (respectively $\mathbb{F} = C$ is particularly important. The precise definition of a *vector space* follows.

Definition: *A vector space V over the scalar field $\mathbb{F}$ is a non-empty set on which addition (+) and scalar multiplication is defined. This means that for each $v_1 \in V$ and $v_2 \in V$, there is a vector denoted by $v_1 + v_2$ that also belongs to V; and for each $v \in V$ and $\alpha \in \mathbb{F}$, there is a vector denoted by αv that also belongs to V. There is also a special vector in V denoted by 0. These operations satisfy the following axioms, where $v_1, v_2, v_3 \in V$ and $\alpha_1, \alpha_2 \in \mathbb{F}$:*

(i)	$v_1 + v_2 = v_2 + v_1$	*(addition is cumulative)*
(ii)	$(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$	*(addition is associative)*
(iii)	$v_1 + 0 = v_1$	*(0 is the additive identity)*
(iv)	*there exists* $w \in V$ *so that* $v_1 + w = 0$	*(existence of additive inverse)*
(v)	$\alpha_1(v_1 + v_2) = \alpha_1 v_1 + \alpha_1 v_2$	*(distributive law)*
(vi)	$(\alpha_1 + \alpha_2)v_1 = \alpha_1 v_1 + \alpha_2 v_1$	*(distributive law)*
(vii)	$\alpha_1(\alpha_2 v_1) = (\alpha_1 \alpha_2) v_1$	*(scalar multiplication is associative)*
(viii)	$1 v_1 = v_1$	*(scalar identity)*

One should note that the symbol 0 can be used in two different ways: one usage as the additive identity in (iii) and the other as the zero element in the feld. In practice, this should not cause any confusion. It is useful, if for nothing else then for simplifying the nomenclature, to denote the element w that appears in (iv) as $-v$. One can deduce from the distributive property of scalar multiplication that $-v = (-1)v$.

It is important to note that a vector space structure does not allow two vectors to be multiplied. At some later point, product operations will be introduced, but a general vector space does not have an inherent product. Also note that abstract vectors do not resemble row or column vectors, but rather a row or column vector description is an example or a notational convenience to represent a vector. We shall see this and more in the following examples.

2 Examples

We now list several examples of vector spaces. Throughout, the field $\mathbb{F}$ can be either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

1) 1) Let $V = \mathbb{F}$ with $\mathbb{F}$ also as the underlying field. Then V is a vector space under the field operations of addition and multiplication are interpreted as the vector addition and scalar multiplication. One can consider $\mathbb{C}$ as a either a real or complex vector space.

2) More generally, let N be a positive integer and V equal the set of an ordered N-tuple of elements of $\mathbb{F}$ written as a row. An element $v \in V$ has the form

$$v = (\alpha_1, \alpha_2, \dots, \alpha_N)$$

where each $\alpha_i \in \mathbb{F}$, $n = 1, \dots, N$. Addition in V is defined by adding the corresponding components: if $v_1 = (\alpha_1, \alpha_2, \dots, \alpha_N)$ and $v_2 = (\beta_1, \beta_2, \dots, \beta_N)$ belong to V, then

$$v_1 + v_2 = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_N + \beta_N)$$

Scalar multiplication is defined by multiplying each component by the scalar: if $v = (\alpha_1, \alpha_2, \dots, \alpha_N) \in V$ and $\alpha \in \mathbb{F}$, then

$$\alpha v = (\alpha\alpha_1, \alpha\alpha_2, \dots, \alpha\alpha_N).$$

In a similar manner, one can consider the vector space of N-tuples written as a column with the same operations. It should cause no confusion if we use the same notation $\mathbb{F}^N$ to denote either the row or column space, although usually it denotes the column space. When the scalar field is the set of real numbers, these vectors can be interpreted geometrically as points in the N-dimensional Euclidean space $\mathbb{R}^N$.

3) Let $V = \mathcal{M}_{MN}$ be the set of all $M \times N$ matrices, where M and N are positive integers and the elements come from the same scalar field as the scalars. The operations of addition and scalar multiplication are defined componentwise in a manner analogous to Example 2. Often,

elements $A \in \mathcal{M}_{MN}$ are thought of as maps $A : \mathbb{F}^N \to \mathbb{F}^M$ that take a column vector $v \in \mathbb{R}^N$ to the column vector $Av \in \mathbb{R}^M$ that is defined by the usual matrix multiplication (see below). Note also that $\mathcal{M}_{MN}$ reduces to the space of N-dimensional row vectors if $M = 1$, and the space of M-dimensional column vectors if $N = 1$.

4) Let $V = \mathcal{P}$ be the set of all polynomials P of arbitrary degree N, where N is nonnegative integer. These are of the form

$$P(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_N t^N,$$

where the coefficients $\alpha_1, \ldots, \alpha_N$ belong to $\mathbb{F}$. For fixed N, the space $\mathcal{P}_N$ consisting of all polynomials of degree less than or equal to N is also a vector space. The vector space operations are the usual ones of elementary algebra.

5) Suppose $1 \leq p < \infty$, and let V be the set of (infinite) sequences

$$v = \{\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots\}$$

where

$$\|v\|_p := \left[\sum_{n=1}^{\infty} |\alpha_n|^p\right]^{\frac{1}{p}} < \infty.$$

The addition of two such sequences is done coordinate-wise, and scalar multiplication is similarly defined. The verification that V is a vector space requires the fact that the sum of two sequences in V also belongs to V. The proof of this fact is a direct consequence of Minkowski's inequality, which we do not prove, but states that

$$\|v_1 + v_2\|_p \leq \|v_1\|_p + \|v_2\|_p.$$

The vector space in this example is denoted by l_p.

6) The space $V = l_\infty$ is defined as the set of bounded sequences. That is, l_∞ consists of those sequences $v = \{\alpha_n\}_n$ for which

$$\|v\|_\infty := \sup_{n \in \mathbb{N}} \{|\alpha_|\} < \infty.$$

The space c_0 (respectively c_{00}) consists of the subset of l_∞ for which $\alpha_n \to 0$ as $n \to \infty$ (respectively, $\alpha_n = 0$ for all large n). Clearly the sum of two sequences from, respectively, l_∞, c_0, and c_{00} again belongs to l_∞, c_0, and c_{00}.

7) Suppose $I \subseteq \mathbb{R}$ is an interval of the real line, and let $V = C[I]$ be the set of all continuous, $\mathbb{F}$-valued functions defined on I. This again is a vector space with the usual operations of addition and scalar multiplication. For $k \in \mathbb{N}$, the space $C^k(I)$ of k-times continuously differentiable functions defined on I is likewise a vector space. Further examples could be given where I is instead a subset of $\mathbb{R}^n$, and/or the range of the function belongs to $\mathbb{F}^m$.

3 Span, linear independence, and basis.

We assume throughout that V is a given vector space. By the definition of a vector space, the sum of two vectors is another vector. But it is also immediate that that the sum of any finite number of vectors is again a vector, and we give this a name. A linear combination is an expression of the form

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_N v_N$$

where the α_n's are scalars and the v_n's are vectors. It is immediate that for any given set of vectors $\mathcal{B} \subseteq V$, the set W consisting of all possible linear combinations of $\mathcal{B}$ again forms a vector space. The space W is called the *span* of $\mathcal{B}$, and denoted by $W = span\mathcal{B}$.

Let $\mathcal{B} \subseteq V$ be a set of vectors. Then $\mathcal{B}$ is called *linearly independent* if for any finite subset of N vectors $\{v_n\}$ of $\mathcal{B}$, the only linear combination of these vectors that equals zero is the one where the coefficients are all zero. In other words, if

$$v_n \in \mathcal{B},\ \alpha_n \in \mathbb{F},\ \sum_{n=1}^{N} \alpha_n v_n = 0 \ \Rightarrow\ \alpha_n = 0\ ,\ \forall\, n = 1, \ldots, N.$$

$\mathcal{B}$ is *linearly dependent* if there exists $N \in \mathbb{N}$, a set of N vectors $\{v_i\} \subseteq \mathcal{B}$, and N scalars α_n not all zero, such that

$$\sum_{n=1}^{N} \alpha_n v_n = 0.$$

It is clear that any set containing the vector 0 is dependent. It is also immediate that any nonempty subset of an independent set is independent.

Consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} \ , \ v_2 = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} \ , \ v_3 = \begin{pmatrix} 2 \\ 5 \\ 3 \end{pmatrix} .$$

To check if these are independent, suppose there are scalars α_1, α_2, and α_3 for which

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0.$$

Writing out the equations component-wise leads to the following system of equations

$$\begin{aligned} \alpha_1 - \alpha_2 + 2\alpha_3 &= 0 \\ 3\alpha_1 - \alpha_2 + 5\alpha_3 &= 0 \\ \alpha_1 - \alpha_2 + 3\alpha_3 &= 0 \end{aligned}$$

which can be readily solved to give the unique solution $\alpha_1 = \alpha_2 = \alpha_3 = 0$ showing that the vectors are linearly independent.

As another example, consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \ , \ v_2 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \ , \ v_3 = \begin{pmatrix} 1 \\ 0 \\ -5 \end{pmatrix} .$$

A similar analysis as above leads to a system that has infinitely many solutions of the form $(\alpha_1, \alpha_2, \alpha_3) = (s, -2s, -s)$ where s is a parameter. Thus, there are α_n's that can be chosen not all zero for which the linear combination is zero, and consequently, these vectors are linearly dependent.

The concepts of span and independence are brought together in the notion of a basis.

Definition: *A set $\mathcal{B} \subseteq V$ of vectors is a basis for V if (i) $V = \text{span}\mathcal{B}$ and (ii) $\mathcal{B}$ is linearly independent.* □

It can be shown (although we shall not do it here) that every vector space has a basis, which is true even if it is not possible to find a finite set of basis vectors. Our definition of basis is more accurately called a *Hamel basis* because all the linear combinations must be finite. However, the more accurate nomenclature is pertinent only in infinite dimensions, as we shall see later. It turns out in fact that Hamel bases in infinite dimensions are typically useless for applications.

The concept of the dimension of a vector space has already been used repeatedly, but we have yet to define what it means. We do so now.

Definition: *A vector space is said to have dimension N if there exists a basis consisting of N vectors. If there is no basis consisting of finite many elements, then the space is said to to be infinite dimensional.* $\square$

For the dimension of a vector space to be well-defined, we must know that every basis has the same number of elements. This is the content of the next theorem.

Theorem 3.1. *Suppose $\{v_1, v_2, \cdots, v_N\}$ and $\{w_1, w_2, \cdots, w_M\}$ are both bases for the vector space V. Then $N = M$.*

Proof: Assume the result is false, and without loss of generality, that $M < N$. Consider now the set $\{w_1, v_1, v_2, \cdots, v_N\}$. Since the v_n's form a basis, we can write w_1 as a linear combination of these:

$$w_1 = \sum_{n=1}^{N} \alpha_n v_n.$$

Since w_1 is not the vector 0, it is immediate that not all the α_n's are zero. Reindexing the v_n's if necessary, assume that $\alpha_1 \neq 0$, then

$$v_1 = \frac{1}{\alpha_1} w_1 - \sum_{n=2}^{N} \frac{\alpha_n}{\alpha_1} v_n.$$

We conclude that any vector that can be written as a linear combination of $\{v_1, v_2, \cdots, v_N\}$ can equally be written as a linear combination of $\{w_1, v_2, \cdots, v_N\}$. and thus the latter set spans V. But this set is also independent: suppose there are scalars $\beta_1, \ldots, \beta_N$ satisfying

$$0 = \beta_1 w_1 + \sum_{n=2}^{N} \frac{\beta_n}{v\ n} = \beta_1 \alpha_1 v_1 + \sum_{n=2}^{N} \big(\beta_1 \alpha_n + \beta_n\big) v_n.$$

Then, since the v_n's are independent, we must have $\beta_1 \alpha_1 = \beta_2 \alpha_2 + \beta_2 = \cdots = \beta_1 \alpha_N + \beta_N = 0$. Now $\alpha_1 \neq 0$, so it follows that $\beta_1 = 0$, and then subsequently each $\beta_n = 0$ for $n = 2, \dots, N$. In other words, $\{w_1, v_2, \cdots, v_N\}$ is linearly independent, and so forms a basis for V.

We now want to repeat the argument and replace one of $v_2, \dots, v_n$ by w_2. There exist scalars $\alpha_1, \dots, \alpha_N$ so that

$$w_2 = \alpha_1 w_1 + \sum_{n=2}^{N} \alpha_n v_n,$$

and not all of $\alpha_2, \dots, \alpha_N$ can be zero, since otherwise we have $w_1 - \alpha_1 w_2 = 0$ which contradicts the fact that the set $\{w_1, w_2\}$ is independent. Again by reindexing if necessary, we can assume $\alpha_2 \neq 0$, and by rearranging terms, we can write

$$v_2 = \frac{1}{\alpha_2} w_2 - \frac{\alpha_1}{\alpha_2} - \sum_{n=3}^{N} \frac{\alpha_n}{\alpha_2} v_n.$$

Thus in a similar manner as the argument above, we can conclude that the set $\{w_1, w_2, v_3, \cdots, v_N\}$ is a basis for V. After repeating this argument M times, we have that $\{w_1, w_2, w_3, \cdots, w_M, v_{M+1}, \dots, v_N\}$ is then a basis. However, we are also assuming that $\{w_1, w_2, w_3, \cdots, w_M\}$ is a basis, and so v_{M+1} is a nontrivial linear combination of $\{w_1, w_2, w_3, \cdots, w_M\}$, which is a contradiction. $\Box$

The above argument also can be used in the infinite dimensional case, and one can conclude that if there exists one basis that is infinite, then every basis is infinite. It also follows from the previous theorem that in an N dimensional space, any set of N linearly independent vectors forms a basis. It is often easy to find at least one basis since the very description of the vector space actually invokes a particular coordinate system (this was the case in Examples 1-4 above). In such a case, it is therefore immediate to determine the dimension. However, in practice, one seeks a particular basis that has desirable properties that may differ from the obvious one.

We return to the previous examples of vector spaces and determine their dimensions.

1) The vector space $V = \mathbb{F}$ is one dimensional over the field $\mathbb{F}$, where a basis is just the multiplicative identity 1. However, $\mathbb{C}$ is two dimensional (with basis $\{1, i\}$) if considered as a real vector space.

2) The space of ordered N-tuples (either considered as rows or columns) N-dimensional, which can be easily seen by noting that the set $\{e_1, \dots, e_N\}$

is a basis, where the vector e_n is the element with 0's in all the positions except the n^{th} where there is a 1. This basis is called the *canonical basis* or *usual set of basis vectors*. Note that there are many other bases, for example $\{e_1, e_1 + e_2, e_1 + e_2 + e_3, \ldots, e_1 + e_2 + \cdots + e_N\}$ is another basis.

3) The vector space $\mathcal{M}_{MN}$ of $M \times N$ matrices has dimension MN, since one can easily verify that the set of matrices $\{E_{mn} : m = 1, \ldots, M, n = 1, \ldots N\}$ is a basis, where E_{mn} denotes the matrix with 1 in the m^{th} row and n^{th} column and zeros elsewhere.

4) The set of polynomials $\mathcal{P}$ is infinite dimensional, since any finite set of polynomials would have a maximum degree. In fact, $\mathcal{P}$ has a basis given by $\{1, t, t^2, \ldots, t^n, \ldots\}$. The subcollection of polynomials $\mathcal{P}_N$ that have degree less than or equal to N has dimension $N + 1$, since the set of polynomials $\{1, t, t^2, \ldots, t^N\}$ forms a basis.

5-7) The sequence and function spaces in these examples are all infinite dimensional, but unlike the polynomial example, it is impossible to explicitly give a (Hamel) basis.

The following result reveals some properties of a set of linear indpendent vectors if the dimension of the space is a priori known.

Theorem 3.2. *Suppose V has dimension N and $\mathcal{B} = \{w_1, w_2, \cdots, w_M\}$ is an independent set. Then $M \leq N$. If $M = N$, then $\mathcal{B}$ is a basis. If $M < N$, then there exist vectors $v_{M+1}, \ldots, v_N$ in V so that $\mathcal{B} \cup \{v_{M+1}, \ldots, v_N\}$ is a basis.*

Proof: We refer back to the steps of the proof of the previous theorem, and use the consequences of related arguments. Suppose $\{v_1, \ldots, v_N\}$ is a basis for V. If $M > N$, then the above shows that after possibly reindexing the v_n's, the set $\{w_1, v_2, \ldots, v_N\}$ is also a basis. After carrying out this replacement $N-1$ more times (which only used the property on $\{mathcalB\}$ that it was an independent set), we conclude that $\{w_1, \ldots, w_N\}$ is a basis. This contradicts that w_{N+1} is in the span while at the same time being independent of $\{w_1, \ldots, w_N\}$, and hence $M \leq N$.

If $M = N$ and $\mathcal{B}$ was not a basis, then $\mathcal{B}$ must not span V, and so there exists a vector v not in $span\mathcal{B}$. We claim that $\mathcal{B} \cup \{v\}$ is linear independent. Indeed, suppose $\sum_{n=1}^{N} \alpha_n w_n + \alpha_{N+1} v = 0$. If $\alpha_{N+1} \neq 0$, then $v = \sum_{n=1}^{N} \frac{\alpha_n}{\alpha_{N+1}} w_n$ belongs to $span\mathcal{B}$, and so we must have $\alpha_{N+1} = 0$. However, since $\mathcal{B}$ is independent, it now follows that all the α_n's are zero, and the claim is proven. So we have that $\mathcal{B} \cup \{v\}$ is an independent set of size $N+1$, which contradicts the previous assertion.

Finally, if $M < N$, by the previous theorem $\mathcal{B}$ is not a basis. Thus since it is independent, it must not span V, and there exists a vector v_{M+1} in V that is not in $span\mathcal{B}$. As in the previous paragraph, we see that $\mathcal{B} \cup \{v_{M+1}\}$ is linearly independent. This argumaent can be repeated a total of $N - M$ times to select vectors $v_{M+1}, \dots v_N$, and we arrive at a set $\mathcal{B} \cup \{v_{M+1}, \dots, v_N\}$ of size N that is independent and therefore must form a basis. $\square$

3.1 Coordinates.

If $\mathcal{B} = \{v_1, \dots, v_N\}$ is a basis for a vector space V, then any vector can be written as a linear combination of the elements of $\mathcal{B}$. Suppose now that the basis $\{\mathcal{B}\}$ is ordered. Then the ordered set of coefficients which multiply the basis vectors completely define the vector of the linear combination, which is the content of the next result.

Theorem 3.3. *Suppose $\{\mathcal{B}\} = \{v_1, \dots, v_N\}$ is an ordered basis and $v \in V$. Then there exists a unique element $[v]_{\mathcal{B}} : (\alpha_1, \dots, \alpha_N) \in \mathbb{F}^N$ so that $v = \alpha_1 v_1 + \cdots + \alpha_N v_N$.*

Proof: Let $v \in V$. Since $\mathcal{B}$ spans V, there exists $(\alpha_1, \dots, \alpha_N) \in \mathbb{F}^N$ so that $v = \alpha_1 v_1 + \cdots + \alpha_N v_N$. We must prove the uniqueness assertion. Let $(\beta_1, \dots, \beta_N) \in \mathbb{F}^N$ also satisfy $v = \beta_1 v_1 + \cdots + \beta_N v_N$. Then

$$0 = v - v = (\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_N - \beta_N)v_N,$$

and it follows that $\alpha_n = \beta_n$ for each $n = 1, \dots, N$ since $\mathcal{B}$ is independent. Thus the representation is unique. $\square$

Th vector $[v]_{\mathcal{B}}$ of ordered coefficients is referred to as the *coordinates* or the *coordinate representation* of the vector relative to the basis $\mathcal{B}$. The previous theorem helps to clarify the concept of dimension: the dimension of a vector space is the least number of parameters that are needed to describe all its elements.

Let us consider $\mathbb{R}^2$ with ordered basis $\{e_1, e_2\}$. If $(a, b) \in \mathbb{R}^2$, then the coordinate representation of (a, b) in this basis is actually just the same vector (a, b). However, if a different basis is chosen, for instance $\mathcal{B} = \{(2, 1), (1, 0)\}$, then the coordinate representation is different. One can find $[(a, b)]_{\mathcal{B}}$ by writing the vector as a linear combination $(a,b) = \alpha_1(2,1) + \alpha_2(1,0)$ and solving the system of equations

$$\begin{aligned} \alpha_1 + \alpha_2 &= a \\ \alpha_1 &= b \end{aligned}$$

that this generates, from which one finds $\alpha_1 = b$ and $\alpha_2 = a - 2b$. The vector itself is therefore different from its coordinate representation. In fact, one should think of the original vector as just its representation in the usual basis, and be mindful that other bases could be used and are often preferred.

Here is another example in the vector space $\mathcal{P}_2$. Find the coordinate representation of the polynomial $1 + t + t^2$ in each of the two bases

$$\mathcal{B}_1 = \{1, t-1, (t-2)(t-1)\} \text{ } and \text{ } \mathcal{B}_2 = \{1, t, t^2\}$$

Writing the vector as a linear combination of the basis vectors in $\mathcal{B}_1$, one obtains the equation

$$1 + t + t^2 = \alpha_1 + \alpha_2(t-1) + \alpha_3(t-2)(t-1),$$

and expanding the right hand side gives

$$1 + t + t^2 = \alpha_1 - \alpha_2 + 2\alpha_3 + t(\alpha_2 - 3\alpha_3) + \alpha_3 t^2$$

Comparing coefficients of terms of equal power on the two sides of the equation results in the three equations

$$\alpha_3 = 1 \ , \ \alpha_2 - 3\alpha_3 = 1 \ , \ \alpha_1 - \alpha_2 + 2\alpha_3 = 1$$

which are solved for the coordinates

$$\alpha_1 = 3 \ , \ \alpha_2 = 4 \ , \ \alpha_3 = 1.$$

Therefore the coordinate representation can be written simply as $[1 + t + t^2]_{\mathcal{B}_1} = (3,4,1)$. While this basis required some algebraic calculations, the other basis, $\mathcal{B}_2$, is more convenient, and one can see immediately that $[1 + t + t^2]_{\mathcal{B}_2} = (1,1,1)$.

4 Isomorphisms of vector spaces.

4.1 Isomorphisms of vector spaces.

Mathematics is full of objects that can be represented various different ways. For instance, the multiplicative identity of real numbers can be written as 1, $\sin^2\theta + \cos^2\theta$, $\frac{7}{7}$, e^0, etc. A considerable portion of the mathematics used in engineering is an exercise in finding the representation of an object that is usable and practical. For instance, a differential equation really contains the same information as its solution, but the differential equation is not a very convenient form to use for most purposes. In other situations one is faced with the problem of determining if two expressions are different representations of the same object. We have seen above that the same vector will have different coordinates for different bases. Another familiar example is the complex numbers $\mathbb{C} = \{x + iy : x, y \in mbR\}$, where it can naturally be thought of as $\mathbb{R}^2 = \{(x, y) : x, y \in mbR\}$. These are not the same spaces, but they exhibit exactly the same vector space properties since there is a manner to transfer all such properties from one space into the other.

Suppose V and W are vector spaces over the same field $\mathbb{F}$. A mapping $\Psi : V \to W$ is called a linear operator if $\Psi(v_1 + v_2) = \Psi(v_1) + \Psi(v_2)$ for all v_1, $v_2 \in V$ and $\Psi(\alpha v) = \alpha\Psi(v)$ for $v \in V$ and $\alpha \in \mathbb{F}$. A linear operator is thus a mapping that preserves the vector space operations. If in addition Ψ is one-to-one ($\Psi(v_1) = \Psi)v_2) \to v_1 = v_2$) and onto (for all $w \in W$, there exists $v \in V$ so that $\Psi(v) = w$), then Ψ is called an *isomorphism* from V to W. If such an isomorphism exists, then the two vector spaces V and W are said to be *isomorphic*, a property denoted by $V \simeq W$. Informally, two vector spaces are isomorphic if the only difference between them is the nomenclature used to designate the elements, where the isomorphism is the rule needed to change from one type of nomenclature to another.

It is straightforward to show that: (1) the identity map from V to V is an isomorphism; (2) if Ψ is an isomorphism from V to W, then its inverse Ψ^{-1} is well-defined and an isomorphism from W to V; and (3) if Ψ_1 is an isomorphism from V_1 to V_2, and Ψ_2 is an isomorphism from V_2 to V_3, then its inverse $\Psi_2 \circ \Psi_1$ is an isomorphism from V_1 to V_3. Thus the property of two vector spaces being isomorphic is an equivalence relation.

One of the surprising features of finite dimensional vector spaces is that for each finite dimension, there is exactly one equivalence class.

Theorem 4.1. *Two finite dimensional vectors spaces over the same field have the same dimension if and only if they are isomorphic.*

Proof: Suppose V has dimension n. From the above discussion, it suffices to show that V is isomorphic to $\mathbb{F}^n$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V, and define $\Psi : V \to \mathbb{F}^n$ by $\Psi(v) = [v]_{\mathcal{B}}$; that is, Ψ maps a vector to its coordinates. Clearly this map is well-defined and onto, and we have seen above that it is one-to-one since coordinates are unique. To show that it preserves the vector space operations, let $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = \beta_1 v_1 + \cdots + \beta_n v_n$ be elements in V. Since $v + w = (\alpha_1 + \beta_1)v_1 + \cdots + (\alpha_n + \beta_n)v_n$, we have

$$\Psi(v+w) = \big(\alpha_1+\beta_1, \ldots, \alpha_n+\beta_n\big) = \big(\alpha_1, \ldots, \alpha_n\big)+\big(\beta_1, \ldots, \beta_n\big) = \Psi(v)+\Psi(w),$$

and thus Ψ preserves addition. The proof that Ψ preserves scalar multiplication is very similar, and hence $t\Psi$ is an isomorphism. $\square$

This theorem implies that all vector calculations in an n-dimensional vector space can be carried out in the most natural of n-dimensional spaces. If the field $\mathbb{F} = \mathbb{R}$, then this space is $\mathbb{R}^n$. Thus, there is no need to investigate the properties of unusual and uncommon finite dimensional vector spaces because their vectorial properties can all be determined by investigating $\mathbb{R}^n$. This is the reason for focusing on $\mathbb{R}^n$ in the following sections.

It should be remembered, however, that only the algebraic operations of addition and scalar multiplication are preserved by an isomorphism. Other natural operations and properties of a particular space do not generally carry over in any obvious way. For instance, the complex numbers as a real vector space is isomorphic to $\mathbb{R}^2$. There is a natural way to multiply two complex numbers that is not so natural in $\mathbb{R}^2$. Another instance is to consider the space of polynomials $\mathcal{P}_n$, in which one could consider the roots of its members. The space $\mathcal{P}_n$ is isomorphic to the space $\mathbb{F}^{n+1}$, but on the face of it, it makes no sense to speak of a root of an element of $\mathbb{F}^{n+1}$.

Another caveat is in order for interpreting the above theorem. If two vector spaces have the same dimension, why not just say they are equal and dispense with the notion of isomorphism altogether? The reason is that different vector spaces may have coordinates representing very different concepts, and referring to them as equal would cause confusion by suggesting that coordinates and vectors from different spaces could be interchanged. For instance, in problems in mechanics, vectors have coordinates that represent positions, velocities, angular momentums, forces, and torques. The position coordinates are very different from the velocities, and both of these are different from the forces, etc.

4.2 Subspaces and sums.

Given a vector space V, there are often smaller spaces that naturally sit inside V but that are vector spaces in the own right. These are called subspaces. For example, the physical world we inhabit and move around is three-dimensional, but often we work on, say, a piece of paper, which is two-dimensional. The concept of a subspace is intuitively natural if one thinks of a line or plane through the origin sitting inside as the Euclidean space $\mathbb{R}^3$. This idea is formalized in the following definition.

Definition: *Let W be a subset of the vector space V. The W is a subspace of V if W is also a vector space in its own right with the operations of addition and scalar multiplication inherited from V.* $\square$

Since the operations of V already are assumed to satisfy the the axioms of a vector space, it is immediate that W is a subspace if and only if the addition of two vectors from W again belongs to W, and that scalar multiplication of a vector from W again belongs to W. These two criteria can be combined in one expression by noting that W is a subspace of V if and only if

$$w_1,\ w_2 \in W,\ \alpha \in \mathbb{F} \quad \text{implies} \quad \alpha w_1 + w_2 \in W.$$

Let us first note a few facts: (1) in every vector space V, there are two trivial subspaces, namely $\{0\}$ and the whole space V, (2) if W_1 and W_2 are subspaces, then $W_1 \cap W_2$ is also a subspace (3) if $\{mathcalB\}$ is any set of vectors in V, then $W = span\mathcal{B}$ is a subspace; (4) every vector space has subspaces of each dimension less than equal to the dimension of V.

We next give specific examples.

Consider the 2-dimensional Euclidean plane $\mathbb{R}^2$, for which we write the elements as (x, y). Since $\mathbb{R}^2$ has dimension 2, the only subspaces that are not $\{0\}$ and $\mathbb{R}^2$ itself must have dimension 1. Now every subspace of dimension one has one basis element, so every nontrivial subspace has the form $\{\alpha(x_0, y_0) : \alpha \in \mathbb{R}\}$ where $(x_0, y_0) \in \mathbb{R}^2$ is a nonzero vector. Two obvious such subspaces are those corresponding to the usual coordinate directions, i.e. the subspaces consisting of the x-axis where $y_0 = 0$, and the y-axis where $x_0 = 0$. In general, the subspaces are straight lines through the origin with a slope of y_0/x_0.

In 3-dimensional Euclidean space $\mathbb{R}^3$, the nontrivial subspaces are either one (lines) or two (planes) dimensional. Since every subspace must contain the origin, the one dimensional subspaces have the form $\{\alpha(x_0, y_0, z_0) : \alpha \in \mathbb{R}\}$ for some nonzero $(x_0, y_0, z_0) \in \mathbb{R}^3$, and the two dimensional ones can be written as $\{(x, y, z) : n_1 x + n_2 y + n_3 z = 0\}$ for some nonzero $(n_1, n_2, n_3) \in \mathbb{R}^3$. More specifically, the vectors $(0, 7, 1)$ and $(1, 2, 3)$ span a subspace of $\mathbb{R}^3$ of dimension two. The parametric representation of this plane is $(\alpha_2, 7\alpha_1 + 2\alpha_2, \alpha_1 + 3\alpha_2)$, where $\alpha_1, \alpha_2 \in \mathbb{R}$. This set of points can equivalently be written as all those points (x, y, z) satisfying $19x + y - 7z = 0$.

In n-dimensional Euclidean space $\mathbb{R}^n$, there are two natural ways to describe subspaces. The parametric description of an m dimensional subspace W consists of finding vectors that span W. In fact one can find a minimal spanning set $\{v_1, \ldots, v_m\}$ (in fact these vectors form a basis for W and so there must be m of them). Now consider a matrix A whose i^{th} column is v_i. Then W is precisely the range of A thought of as a map from $\mathbb{R}^m$ to $\mathbb{R}^n$. That is, $W = \{Aw : w \in \mathbb{R}^m\}$.

A second way to describe a subspace W of $\mathbb{R}^n$ is as the null space of a certain matrix $B \in \mathcal{M}nm$. If $B \in \mathcal{M}nm$ then it can be easily checked that $W = \{v : Bv = 0\}$ is a subspace, which is called the null space of B. The converse is also true, which means that given an m dimensional subspace W, then there exists $B \in \mathcal{M}nm$ so that W is of this form. We shall see how to define B later.

Examples of function subspaces include the following. On the real line, let $\mathcal{P}_n$ $(n \geq 0)$ be the set of polynomials of degree less than or equal to n, $\mathcal{P}$ the set of polynomials of arbitrary degree, and C^k $(k \geq 0)$ the k-times continuously differentiable functions, where C^0 is just the set of continuous functions. It is clear that for each nonnegative integers $n_1 < n_2$ and $k_1 < k_2$, we have the subset inclusions

$$\mathcal{P}_{n_1} \subset \mathcal{P}_{n_2} \subset \mathcal{P} \subset C^{k_2} \subset C^{k_1}.$$

However, since all of these are vector spaces, it is also the case that each smaller space is a subspace of the larger one.

Evidently, all subspaces must contain the zero element so any two subspaces must have at least this element in common. If the zero element is the only element that two subspaces have in common, then the two subspaces are said to be *disjoint*. This terminology is justified by the following consideration. Suppose $\mathcal{B}$ is a linear independent set in V, $\mathcal{B}_1$ and $\mathcal{B}_2$ are subsets of $\mathcal{B}$, and $W_1 = span\mathcal{B}_1$ $W_2 = span\mathcal{B}_2$. Then the subspaces W_1 and W_2 are disjoint subspaces if and only if $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint subsets.

4.3 Vector products.

As we emphasized earlier, there is no generic multiplication of vectors defined on a vector space. Nonetheless, there are several types of products between vectors with different characteristics and notation that can be defined. We explore some of those in this section.

4.3.1 Inner product.

The most widely used product is called the scalar or dot product, and is defined from $\mathbb{C}^n \times \mathbb{C}^n$ into$\mathbb{C}$ by

$$v \bullet w = \sum_{i=1}^{n} v_i \bar{w}_i \tag{1}$$

where the overbar indicates complex conjugation (if $\alpha = a + ib \in \mathbb{C}$, then its conjugate is $\bar{\alpha} = a - ib$). One has a similar product defined from $\mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}$, where the conjugation in the above definition becomes superfluous. As we shall soon see, the scalar product leads to additional geometric concepts like length and orthogonality.

A generalization of the scalar product is an inner product, which is defined axiomatically.

Definition: *An inner product is a binary operation which maps two elements of a complex (resp. real) vector space V into $\mathbb{C}$ (resp. $\mathbb{R}$ that satisfies the following properties, where u, v, and w are arbitrary vectors and α is a scalar.*

(i) (Additivity) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$;

(ii) (Homogeneity) $\langle \alpha v, w\rangle = \alpha\langle v, w\rangle$;

(iii) (Conjugate Symmetry) $\langle v, w\rangle = \langle\overline{w, v}\rangle$;

(iv) (Positive definiteness) $\langle v, v\rangle > 0$, *if* $v \neq 0$.

Properties (i) and (ii) together are usually referred to as the linearity properties of the inner product. For real vector spaces, the conjugate in (iii) is again superfluous. From (i)-(iii), it follows that

$$\langle u, v + \alpha w\rangle = \langle u, v\rangle + \bar{\alpha}\langle u, w\rangle.$$

Note also that the symmetry property (iv) says that that the inner product of a vector with itself is real even if the associated field is complex. The requirement of positive definiteness is therefore meaningful even when the vector contains complex elements. A vector space on which an inner product is defined is called an *inner product space*. It is easy to check that the scalar product is an example of an inner product.

Now suppose V is a vector space with an inner product $\langle \cdot, \cdot \rangle$. The length of a vector v is defined as

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

and the angle θ between two vectors v and w is defined to satisfy

$$\langle v, w \rangle = |v||w| \cos \theta$$

A vector v is a *unit* of *normal* vector if it has length 1. Any non-zero vector v can be *normalized* into a unit vector by dividing it by its length. Note that if v is a unit vector, then $\langle w, v \rangle$ is the length of the *orthogonal projection* in the direction of v, and the vector $\langle w, v \rangle v$ is the *component* of w in the direction v. Finally, two vectors, different from zero, are said to be *orthogonal* if their inner product is zero. It is important to note that the geometrical concepts all depend on the particular given inner product - different inner products will yield different results.

To see how inner products may arise generally, suppose V is a vector space with a given basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Recall that $[v]_{\mathcal{B}} \in \mathcal{F}^n$ denotes the coordinates of a vector $v \in V$. The inner product associated to $\mathcal{B}$ can be defined as the scalar product of the coordinates of the vectors with repect to $\mathcal{B}$:

$$\langle v, w \rangle_{\mathcal{B}} = [v]_{\mathcal{B}} \bullet [w]_{\mathcal{B}}$$

One can verify that the axioms (i)-(iv) hold. Note that with this inner product, each basis vector is a unit vector and is orthogonal to every other basis vector. If $V = \mathbb{C}^n$ or $\mathbb{R}^n$, and $\mathcal{B}$ is the canonical basis $\{e_1, \ldots, e_n\}$, then this inner product coincides with the scalar product.

A basis that has pairwise orthogonal elements is called an *orthogonal basis*, and is called an *orthonormal basis* if in addition each basis element is a normal vector. We have just seen that any basis is an orthonormal basis with respect to some inner product, but given an inner product, there are many such orthonormal bases associated to it. We shall discuss how to construct some of these later, but next we illustrate the utility and advantage of working with an orthonormal basis to represent and manipulate vectors.

Suppose $\{v_1, \ldots, v_n\}$ is an orthonormal basis and $v \in V$. The coordinate representation $[v]_{\mathcal{B}} = (\alpha_1, \ldots, \alpha_n$ of v means that

$$v = \sum_{i=1}^{n} \alpha_i v_i.$$

Now fix an index $1 \leq j \leq n$, and take the inner product of both sides of the expression with an arbitrary basis vector v_j. The additive and homogeneous properties of the inner product, and the orthonormal assumption on the basis yields that

$$\langle v, v_j \rangle = \left\langle \sum_{i=1}^{n} \alpha_i v_i, v \right\rangle = \sum_{i=1}^{n} \alpha_i \langle v_i, v_j \rangle = \alpha_j.$$

That is, the coordinates of an arbitrary vector $v \in V$ with respect to an orthonormal basis can be found easily be taking the inner of v with the corresponding basis vector.

4.3.2 Cross product

The cross product $v \times w$ is defined only between vectors v and w in $\mathbb{R}^3$, and is another vector in $\mathbb{R}^3$. It can be written several ways,

$$v \times w = |v||w| \sin(\theta_{vw}) n_{vw}$$

where θ_{vw} is the angle between the two vectors and n_{vw} is the unit normal to the v, w plane, pointing in the direction indicated by the right hand rule (in a right handed coordinate system) applied to v, w or by the left hand rule (in a left handed coordinate system). Geometrically, $v \times w$ is a vector of magnitude equal to the area of the parallelogram spanned by v and w, orthogonal to the plane of this parallelogram, and pointing in the direction indicated by the right hand rule. An equivalent definition in terms of the usual coordinate representation $v = (v_1, v_2, v_3)$ and $(w = (w_1, w_2, w_3)$ of the cross product is

$$v \times w = (v_2 w_3 - v_3 w_2, v_3 w_1 - v_1 w_3, v_1 w_2 - v_2 w_1).$$

A third form can be given as the determinate (which will be defined below) of the 3×3 matrix

$$v \times w = \begin{vmatrix} e_1 & e_2 & e_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

One can check that $v \times w = -w \times v$.

4.3.3 Triple Scalar Product

The *triple scalar product* is defined for three vectors of $\mathbb{R}^3$ by

$$[v, w, u] = v \bullet (w \times u),$$

and it equals the volume of the parallelepiped formed by the three vectors. In terms of coordinates

$$[v, w, u] = v_1(w_2u_3 - w_3u_2) + v_2(w_3u_1 - w_1u_3) + v_3(w_1u_2 - w_2u_1) = \begin{vmatrix} v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \end{vmatrix}$$

4.3.4 Dyad or outer product.

The dyad or outer product $\rangle v, w\langle$ of two vectors v and w in $\mathbb{F}^n$ is defined as the $n \times n$ matrix whose entry in the i^{th} row and j^{th} column is v_iw_j. The outer product is not commutative. For example in $\mathbb{R}^3$ with $v = (1, 2, 3)$ and $w = (6, -2, 4)$, one has

$$\rangle v, w\langle = \begin{pmatrix} 6 & -2 & 4 \\ 12 & -4 & 8 \\ 18 & -6 & 12 \end{pmatrix} \qquad and \qquad \rangle w, v\langle = \begin{pmatrix} 6 & 12 & 18 \\ -2 & -4 & -6 \\ 4 & 8 & 12 \end{pmatrix}.$$

4.4 Gram-Schmidt Orthogonalization.

The definition of a basis is purely algebraic and may have nothing to do with any inherent geometry in the space. We have seen that any basis is orthonormal with respect to some inner product; however, it often turns out that a specific inner product is desired and one wants to find an orthonormal basis with respect to that product. The main example of course is $\mathbb{F}^n$ with the scalar product, but the procedure introduced in this section, known as Gram-Schmidt orthogonalization, is valid for any given inner product. The rationale for the procedure is that orthogonal bases are more convenient and easier to use in practice than non-orthogonal bases. In fact, many engineers have probably never worked problems with non-orthogonal bases and do not appreciate the difficulties that could arise if the basis vectors are not orthogonal. The Gram-Schmidt orthogonalization procedure is an algorithm that converts a given basis into an orthonormal one with respect to a given inner product.

Let $\{v_k\}_{k=1,n}$ be a given, not necessarily orthogonal basis. We shall construct an orthogonal basis $\{w_k\}_{k=1,n}$ in such a way that $span\{v_k : k = 1, \ldots, n\} = span\{v_k : k = 1, \ldots, n\}$ for each $n = 1, \ldots, n$. It is clear that the first new basis vector should be $w_1 = v_1$. Observe that $\{v_2, w_1\}$ is indpendent, which is equivalent to saying $v_2 - cw_1 \neq 0$ for every c. The idea is to choose c so that the resulting vector w_2 is orthogonal to w_1. The value of c can be easily determined by setting $0 = \langle v_2 - cw_1, w_1\rangle = \langle v_2, w_1\rangle - c\langle w_1, w_1\rangle$ (by the linearity properties of the inner product), and thus $c = \frac{\langle v_2, w_1\rangle}{|w_1|^2}$. The second new basis vector is thus

$$w_2 = v_2 - \frac{\langle v_2, w_1\rangle}{|w_1|^2} w_1.$$

Clearly $span\{v_1, v_2\} = span\{w_1, w_2\}$ and $\{w_1, w_2\}$ is orthogonal. We continue this reasoning. Since $\{v_3, w_2, w_1\}$ is independent, then $v_3 - c_1 w_1 - c_2 w_2 \neq 0$ for all constants c_1 and c_2 and we want to choose these constants so the resulting vector is orthogonal to both w_1 and w_2. Being orthogonal to w_1 means that $0 = \langle v_3 - c_1 w_1 - c_2 w_2, w_1\rangle = \langle v_3, w_1\rangle - c_1 |w_1|^2$, and so $c_1 = \frac{\langle v_3, w_1\rangle}{|w_1|^2}$. Similarly, being orthogonal to w_2 forces $c_2 = \frac{\langle v_3, w_2\rangle}{|w_2|^2}$, and w_3 is defined as

$$w_3 = v_3 - \frac{\langle v_3, w_1\rangle}{|w_1|^2} w_1 - \frac{\langle v_3, w_2\rangle}{|w_2|^2} w_2.$$

In general, w_n is found in the same way and equals

$$w_n = v_n - \sum_{k=1}^{n-1} \frac{\langle v_n, w_k\rangle}{|w_k|^2} w_k.$$

It is sometimes more convenient to normalize the new basis vector at each stage by replacing w_n as defined above by $w_n/|w_n|$. Then an orthonormal basis $\{w_1, \ldots, w_n\}$ is produced by the recursive formula

$$w_n = \frac{v_n - \sum_{k=1}^{n-1} \langle v_n, w_k\rangle w_k}{|v_n - \sum_{k=1}^{n-1} \langle v_n, w_k\rangle w_k|}$$

The Gram-Schmidt procedure has an attractive geometric interpretation: at each stage, the new vector w_n consists of the subtraction from v_n of all the components of v_n that lie in the directions $w_1, \ldots, w_{n-1}$, and then is normalized.

We illustrate the Gram-Schmidt orthogonalization procedure by considering the basis $\mathcal{B}$ of $\mathbb{R}^4$ given by

$$\mathcal{B} = \{v_1, v_2, v_3, v_4\} = \left\{ \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix} \right\}$$

The first vector in the orthogonal basis and needed information is

$$w_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}; \quad |w_1| = \sqrt{14}.$$

Since $\langle v_2, w_1 \rangle = 4$, the second orthogonal basis vector is calculated as

$$w_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{7} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 7 \\ 5 \\ -4 \\ 1 \end{pmatrix}; \quad |w_2| = \frac{\sqrt{91}}{7}.$$

We see that $\langle v_3, w_1 \rangle = 10$ and $\langle v_3, w_2 \rangle = \frac{1}{7}$, and the third new basis vector w_3 is

$$w_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix} - \frac{5}{7} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} - \left(\frac{1}{13}\right) \left(\frac{1}{7}\right) \begin{pmatrix} 7 \\ 5 \\ -4 \\ 1 \end{pmatrix} = \frac{1}{13} \begin{pmatrix} 12 \\ -10 \\ 8 \\ -2 \end{pmatrix}; \quad |w_3| = \frac{2\sqrt{78}}{13}$$

Finally, since $\langle v_4, w_1 \rangle = 8$, $\langle v_4, w_2 \rangle = \frac{12}{7}$, and $\langle v_4, w_3 \rangle = \frac{28}{13}$, we have

$$w_4 = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix} - \frac{4}{7} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} - \frac{12}{91} \begin{pmatrix} 7 \\ 5 \\ -4 \\ 1 \end{pmatrix} - \frac{7}{78} \begin{pmatrix} 12 \\ -10 \\ 8 \\ -2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 0 \\ 2 \\ 2 \\ -2 \end{pmatrix}.$$

5 Matrices.

Matrices are ordered sets of numbers of the form $\{a_{n,m}\}_{n=1,N;m=1,M}$. The first subscript of an element is said to be the row number of that element, the second subscript the column number. The matrix $\{a_{n,m}\}_{n=1,N;m=1,M}$ is said to have N rows and M columns. Matrices, if they are not too large, are usually written as tables of numbers

$$A_{N\times M} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,M} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,M} \end{pmatrix}$$

The *main diagonal* of a matrix are the elements with the same row and column index, $a_{n,n}$. When the number of rows equals the number of columns a matrix is said to be *square*. A square matrix is called a *diagonal matrix* if the elements outside the main diagonal are zero. A square matrix is said to be *upper triangular* if all elements below the diagonal are zero. Similarly a *lower triangular* matrix is a square matrix with zeros in the elements above the diagonal.

The *transpose* of a matrix A is written A^T and is the matrix one obtains by interchanging the rows and columns of A, i.e. $\{a_{n,m}\}^T = \{a_{m,n}\}$ or

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,M} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \cdots & a_{N,M} \end{pmatrix}^T = \begin{pmatrix} a_{1,1} & \cdots & a_{N,1} \\ \vdots & \ddots & \vdots \\ a_{1,M} & \cdots & a_{N,M} \end{pmatrix}$$

A matrix is said to be *symmetric* if it is equal to its own transpose. Only square matrices can be symmetric. A matrix is *antisymmetric* or *skew-symmetric* if it equals minus its transpose. Only square matrices with zeros in the diagonal elements can be antisymmetric. The *identity matrix* is a square matrix with 1's in the main diagonal and zero's in the off diagonal elements.

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

If the dimension of the identity matrix is relevant, then we write I_N for the square matrix of size N.

5.1 Matrix Algebra.

As mentioned earlier, matrices with the same number of N rows and M columns can be added element by element

$$\{a_{n,m}\}_{n=1,N;m=1,M} + \{b_{n,m}\}_{n=1,N;m=1,M} = \{a_{n,m} + b_{n,m}\}_{n=1,N;m=1,M}.$$

Moreover, multiplication of a matrix by a scalar is defined by $\alpha\{a_{n,m}\}_{n=1,N;m=1,M} = \{\alpha a_{n,m}\}_{n=1,N;m=1,M}$. These operations satisfy the axioms of a vector space, and in fact one can easily produce an isomrphism between $\mathcal{M}_{N\times M}$ and $\mathbb{F}^{NM}$, thus $\mathcal{M}_{N\times M}$ has dimension NM.

In addition to the vector space structure of $\mathcal{M}_{mn}$, matrices can be multiplied if the first matrix has the same number of columns as the second has rows. Suppose $A = \{a_{mn}\} \in \mathcal{M}_{mn}$ and $B = \{b_{nk}\} \in \mathcal{M}_{nk}$ are two matrices. Their product $C = \{c_{mk}\} = AB$ belongs to $\mathcal{M}_{mk}$ and is given by

$$c_{mk} = \sum_{n=1}^{n} a_{mn} b_{nk}.$$

Matrix algebra operations satisfy many of the familiar properties of arithmetic, with the notable exception of commutativity of multiplication. One should note that even if $m = n = k$, the product AB will generally be different from BA, which means that multiplication is not commutative. However, the following properties hold, where in each case we assume the dimensions of the matrix are such that the operations are defined.

(i)	$C(A+B) = CA + CB$	(left distributive law)
(ii)	$(A+B)C = AC + BC$	(right distributive law)
(iii)	$A(BC) = (AB)C$	(multiplication is associative)
(iv)	If $A \in \mathcal{M}_{MN}$ then $I_M A = A = A I_N$	(identity)
(v)	$(AB)^T = B^T A^T$	(transpose of product)

Suppose $A \in \mathcal{M}_{MN}$. Then A has a left inverse if there exists a matrix $A_l^{-1} \in \mathcal{M}_{MM}$ satisfying $A_l^{-1} A = I_M$, and has a right inverse if there exists $A_r^{-1} \in \mathcal{M}_{NN}$ satisfying $AA_r^{-1} = I_N$.

Let us note that the columns of an $M \times N$ matrix are independent vectors in $\mathbb{F}^M$ if and only if whenever $Av = 0$ for $v \in \mathbb{F}^N$, then necessarily $v = 0$; this is just another way of writing the definition of linear independence in terms of matrix multiplication. If A has a left inverse A_l^{-1}, then $Av = 0$ implies $v = I_N v = A_l^{-1} Av = 0$, and hence the columns of A are independent. In particular, the existence of a left inverse implies $N \le M$.

We can apply the same reasoning to the transpose. The transpose of product rule implies that if A has a right inverse A_r^{-1}, then A^T has $(A_r^{-1})^T$ as a left inverse. Hence in this case, the columns of A^T are independent, which is equivalent to saying the rows of A are independent, and in particular $M \leq N$.

If a matrix has both a left and a right inverse, then it follows from above that $M = N$. In fact the left and right inverses must be equal, as can be seen from the following short calculation.

$$A_l^{-1} = A_l^{-1}(AA_r^{-1}) = (A_l^{-1}A)A_r^{-1} = A_r^{-1}.$$

A square matrix A is called invertible if it has a left and right inverse, and since these must be the same matrix, we write it simply as A^{-1}. We shall see a later that for a square matrix, the existence of a right inverse implies the existence of a left one, and vice versa.

We emphasize that not every nonzero matrix has a multiplicative inverse, which is unlike ordinary numbers. For example, the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

has no left inverse because the second row of the product AB will be all zero for any matrix B.

If the inverse of a square matrix does exist, it is a useful tool for solving matrix problems. For instance, a set of coupled, linear equations with the same number of unknowns as equations can be written as $Ax = b$, so the solution is simply obtained by premultiplying each side by the inverse of A: $x = A^{-1}b$. Even something more general is true: x solves $Ax = b$ if and only if x solves the system $SAx = Sb$ for any $M \times M$ invertible square matrix S. The goal, then, is to find an invertible matrix S so that the solutions of $SAx = Sb$ are very simple to find.

5.2 Gauss Elimination.

We now explore systematically how to solve a system $Ax = b$. Suppose A is an $M \times N$ matrix with entries (a_{mn}) and $b \in \mathbb{R}^M$ with components b_m. We seek to find $x \in mbF^N$ that will satisfy the system of equations $Ax = b$, which written out in full looks like

$$\begin{array}{ccccccccc}
a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots & + & a_{1,N}x_N & = & b_1 \\
a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots & + & a_{2,N}x_N & = & b_2 \\
\vdots & & & & \ddots & & & \vdots & \\
a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots & + & a_{m,N}x_N & = & b_m \\
\vdots & & & & \ddots & & & \vdots & \\
a_{M,1}x_1 & + & a_{M,2}x_2 & + & \cdots & + & a_{M,N}x_N & = & b_M
\end{array}$$

There may be no solution at all; for example, if all $a_{mn} = 0$ and one of the $b_m \neq 0$, there obviously cannot be a solution. Or there may be infinitely many solutions; for example, with $M = 1$, $N = 2$, $a_{11} = a_{12} = 1$, and $b_1 = 0$, then any $x \in \mathbb{R}^2$ with $x_1 = -x_2$ solves the system. *Gauss elimination* is a procedure that routinely converts a system of equations into an equivalent system (that is, it has exactly the same set of solutions) which is amenable to determining if solutions exist, and if so, to finding them.

There are two steps to Gauss elimination. The first step is called forward substitution and obtains an matrix with zeros below the main diagonal. The second step, called back substitution, aims to obtain zeros above the main diagonal.