4.0 VECTOR METHODS

The primary focus of linear algebra is on vector spaces and their transformations, which can be represented by matrices. We will cover vector methods in this chapter and discuss matrix methods separately. Vector space theory in its present form is a relatively recent mathematical invention. There are glimmers of the notion of vector in the works of the Greek geometers, but the core idea seems to have been dormant until it reappeared in a form that we would recognize today about the time of Isaac Newton. Although Newton worked with quantities having both magnitude and direction like force, displacement and velocity, he never took the step of regarding them as examples of the broader concept we now call "vector". Several prominent mathematicians including Gauss and Hamilton moved the concept closer to our current formulation in the first half of the nineteenth century. The German high school teacher Hermann Grassmann published an amazingly prescient theory in the 1840's touching on n-dimensional spaces and the use of matrix transformations, but his work was largely and unjustly ignored. Grassmann had no stature in the scientific community and his book was written with a cryptic notation, but he was vindicated thirty years later when the significance of his contributions was finally appreciated. The major impetus for the adoption of vector methods came from the physicists James Clerk Maxwell and Oliver Heaviside, who framed their theories of electromagnetism in terms of vectors. The American physicist J. Willard Gibbs had studied Maxwell's work and adapted it to thermodynamics. Gibbs turned out to be the prime mover in codifying and popularizing modern vector space theory. He was a professor at Yale and in the early 1880's began to circulate printed notes explaining vector analysis, first to his students and ultimately to scientists in the United States and Europe. In 1901 he published the first text on modern vector space theory entitled *Vector Analysis*. Acceptance of Gibbs' theory was wide but not unanimous. There were still holdouts who championed Hamilton's old theory of quaternions, but that view eventually faded as the vector concept proved superior.

A typical vector theory course taught today to engineering and physics students would have a practical focus probably very similar to a course Gibbs might have offered. This is understandable in view of how the subject developed. More recently, mathematicians have pursued abstractions of vector space theory that would be barely recognizable to an engineer. Our discussion will emphasize the down-to-earth problem-solving aspects of vector methods and leave the abstruse points to the theorists. We now turn to a presentation of the basic definitions.

4.1 BASIC DEFINITIONS

A **vector space** (**linear space**) is a pair of sets $\langle V, F\rangle$, where V consists of elements called **vectors** equipped with the operation of **vector addition** and F is a so-called **scalar field**, typically $\mathbb{R}$ or $\mathbb{C}$. Elements of the scalar field may be combined with vectors in accordance with **scalar multiplication**. We will write vectors in boldface and scalars unbolded. Vector addition and scalar (field) addition will be denoted by "$+$". There should be no ambiguity based on what elements are being added, but if it helps with clarity the symbol $\oplus$ can be used for vector addition. Scalar multiplication will be indicated by juxtasposition with the scalar always to the left. Operations within the field have been covered in §0.21. When the scalar field is clear from context we often abuse the notation and refer to the vector space as just V. The operations specific to a vector space behave as follows:

Table 1 - Axioms for Vector Spaces
$\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $a, b, 0, 1 \in F$

1) Associativity of addition	$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
2) Commutativity of addition	$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
3) Existence of an additive identity (the zero vector)	$\mathbf{x} + \mathbf{0} = \mathbf{x}$ (note $\mathbf{0} \neq 0$)
4) Existence of additive inverses	$\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5) Unitarity	$1\mathbf{x} = \mathbf{x}$
6) Annihilation	$0\mathbf{x} = \mathbf{0}$
7) Distributivity of scalar multiplication over vector addition	$a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$
8) Distributivity of scalar multiplication over scalar addition	$(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$
9) Associativity of scalar multiplication	$a(b\mathbf{x}) = (ab)\mathbf{x}$

Axioms 1) through 4) involve only vector addition. V with these four properties form an

abelian group.

Example 1 *Show that $\mathbb{R}^2$ is a vector space over the real field ($\mathbb{R}$ again).*

Solution: The elements of $\mathbb{R}^2$ are point pairs (x, y). Define vector addition as $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$. Define scalar multiplication by $c \in \mathbb{R}$ as $c(x, y) = (cx, cy)$. The associativity and commutativity are obvious from the properties of $\mathbb{R}$. The element $(0, 0)$ serves as the additive identity. The additive inverse of (x, y) is clearly $(-x, -y)$. Unitarity and annihilation are trivial. For distributivity of scalar multiplication over vector addition we have $c[(x_1, y_1) + (x_2, y_2)] = c(x_1, y_1) + c(x_2, y_2)$ and for distributivity of scalar multiplication over scalar addition, this is immediate from the field properties of $\mathbb{R}$. Finally, $c_1(c_2(x, y)) = (c_1 c_2)(x, y)$. The axioms are satisfied so we conclude $\langle \mathbb{R}^2, \mathbb{R} \rangle$ is a vector space.

Example 2 *Show that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$. Of course, this will supersede Example 1 and separately show that $\langle \mathbb{R}^3, \mathbb{R} \rangle$ is a vector space.*

Solution: The elements of $\mathbb{R}^n$ *are n-***tuples** $(x_1, x_2, \cdots, x_n)$. Define vector addition and scalar multiplication in complete analogy with Example 1. So $(x_1, x_2, \cdots, x_n) + (x'_1, x'_2, \cdots, x'_n) = (x_1 + x'_1, x_2 + x'_2, \cdots, x_n + x'_n)$ and $c(x_1, x_2, \cdots, x_n) = (cx_1, cx_2, \cdots, cx_n)$. Verification that elements of $\mathbb{R}^n$ conform to the axioms proceeds exactly parallel to Example 1 and we conclude that $\langle \mathbb{R}^n, \mathbb{R} \rangle$ and hence $\langle \mathbb{R}^3, \mathbb{R} \rangle$ are vector spaces.

There is a subtlety in Examples 1 and 2 that needs to be clarified. We have shown that the various n-tuples equipped with their relevant operations fulfill the nominal requirements for a vector space. $\mathbb{R}^2$ and $\mathbb{R}^3$ as cartesian products are technically just collections of points or geometric addresses. To develop a geometry on these spaces we typically impose the so-called **euclidean metric** ρ, which gives us the ability to determine the distance between points. We remark that this distance function is available for any $\mathbb{R}^n$, and occasionally you will see the notation $\mathbb{E}^n = \langle \mathbb{R}^n, \rho \rangle$ to emphasize the presence of the metric. The metric boils down to a generalization of the Pythagorean formula in n dimensions. Given $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, $\rho(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.

So far we have not seen an element that explicitly has magnitude and direction...properties that we are expecting from our naïve understanding of vector spaces. We can easily remedy that. Using $\mathbb{R}^2$ as an example, let us define the vector $\langle a, b \rangle$ to be the position vector of the point (a, b). It is a **directed line segment** connecting the origin $(0,0)$ of $\mathbb{R}^2$ as the **initial point** (**tail**) to the point (a, b) as the **terminal point** (**head**). This scheme clearly gives us a bijective map between points (a, b) and vectors $\langle a, b \rangle$. For the former we view $\mathbb{R}^2$.as just a set of locations in the xy-plane and for the latter we view $\mathbb{R}^2$ as a collection of directed line segments emanating from the origin. From context and notation it will be easy to distinguish a location from its position vector. The general vector space axioms are clearly satisfied as well for the vector interpretation, so we are now justified in calling it a vector space. The applicable metric is just the familiar Pythagorean formula in $\mathbb{R}^2$ for distance between points $P_1 = (x_1, y_1)$ and $P_2(x_2, y_2)$: $d(\overline{P_1 P_2}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. This immediately gives us the length of $\langle a, b \rangle$ as $\sqrt{a^2 + b^2}$. So now the vector $\langle a, b \rangle$ is finally shown to have both magnitude and direction, as we were anticipating.

In general, the **length** (or **norm**) of an arbitrary vector $\mathbf{v}$ is customarily written as $|\mathbf{v}|$, or, if there may be confusion (as in Table 2) with absolute value, as $\|\mathbf{v}\|$. For a vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ in $\mathbb{R}^n$ we have immediately $|\mathbf{x}| = \sqrt{\sum_{i=1}^{n} x_i^2}$. Vector spaces where such a measure of vector length is available are called **normed vector spaces**. There are abstract spaces which are not normable.

Table 2 - Axioms for Vector Space Norms

$\mathbf{x}, \mathbf{y} \in V$, $a \in F$ $\quad \|\cdot\|$ is norm, $|\cdot|$ is absolute value or modulus if $F = \mathbb{C}$

1) Faithfulness	$\|\mathbf{x}\| = 0$ implies $\mathbf{x} = \mathbf{0}$
2) Absolute (or positive) homogeneity	$\|a\mathbf{x}\| = \|a\| \cdot \|\mathbf{x}\|$
3) Subadditivity (triangle inequality)	$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

These axioms immediately imply that:

(i) $\|\mathbf{0}\| = 0$, since for any $\mathbf{x} \in V$, $\|\mathbf{0}\| = \|0\mathbf{x}\| = |0|\|\mathbf{x}\| = 0$. This seemingly obvious property is the converse of faithfulness, so now we know that the only vector with zero length is $\mathbf{0}$.

(ii) for any $\mathbf{x} \in V$, $\|\mathbf{x}\| \geq 0$, since $\|\mathbf{x} - \mathbf{x}\| \leq \|\mathbf{x}\| + \|\mathbf{x}\|$ implies $2\|\mathbf{x}\| \geq \|\mathbf{0}\| = 0$. Hence every nonzero vector has positive length.

Faithfulness deserves an explanation. It ensures that the norm function recognizes, or separates, different vectors, since if $\|\mathbf{x}-\mathbf{y}\| = 0$, then $\mathbf{x}-\mathbf{y} = \mathbf{0}$ and $\mathbf{x} = \mathbf{y}$. A norm-like function $\|\cdot\|_s$ that satisfies axioms 2) and 3), but not 1), is called a **seminorm**. A seminorm may assign zero length to a nonzero vector, so if two vectors $\mathbf{x}$ and $\mathbf{y}$ are not equal, it may still be that $\|\mathbf{x}-\mathbf{y}\|_s$ is zero. We say in that case that the seminorm does not distinguish or separate the vectors $\mathbf{x}$ and $\mathbf{y}$.

Some vector spaces come equipped with a function called a **bilinear form**. If (V,F) is a vector space, then $f : V \times V \to F$ is a bilinear form if the following axioms are satisfied.

Table 3 - Axioms for Bilinear Forms

$\mathbf{x},\mathbf{y},\mathbf{z} \in V,\ a \in F$

1) Additivity in the first argument	$f(\mathbf{x}+\mathbf{y},\mathbf{z}) = f(\mathbf{x},\mathbf{z}) + f(\mathbf{y},\mathbf{z})$
2) Additivity in the second argument	$f(\mathbf{x},\mathbf{y}+\mathbf{z}) = f(\mathbf{x},\mathbf{y}) + f(\mathbf{x},\mathbf{z})$
3) Homogeneity in either argument	$f(a\mathbf{x},\mathbf{y}) = f(\mathbf{x},a\mathbf{y}) = af(\mathbf{x},\mathbf{y})$

Mathematicians generally call a vector space equipped with a bilinear form an **inner product** space. This seems to be too general for physicists. They have a more specific concept of an inner product space that meshes with their formalism for quantum mechanics. Physicists define an inner product space as a vector space over $F = \mathbb{R}$ or $\mathbb{C}$ with a function called an inner product $\phi : V \times V \to F$ that satisfies the following axioms:

Table 4 - Axioms for Inner Product Spaces (**applied version**)

$\mathbf{x},\mathbf{y},\mathbf{z} \in V,\ a \in F$

1) Linearity in the first argument	$\phi(a\mathbf{x}+\mathbf{y},\mathbf{z}) = a\phi(\mathbf{x},\mathbf{z}) + a\phi(\mathbf{y},\mathbf{z})$
2) Conjugate symmetry	$\phi(\mathbf{x},\mathbf{y}) = \overline{\phi(\mathbf{y},\mathbf{x})}$
3) Positivity	$\phi(\mathbf{x},\mathbf{x}) \geq 0$
4) Definiteness	$\phi(\mathbf{x},\mathbf{x}) = 0$ implies $\mathbf{x} = 0$

Note axiom 2) implies $\phi(\mathbf{u},\mathbf{u}) \in \mathbb{R}$, since $\phi(\mathbf{u},\mathbf{u}) = \overline{\phi(\mathbf{u},\mathbf{u})}$. Also axiom 4) implies $\phi(\mathbf{0},\mathbf{0}) = \phi(0\mathbf{x},\mathbf{0}) = 0\phi(0\mathbf{x},\mathbf{0}) = 0$, then axiom 1) gives the converse.

These axioms immediately imply that $\phi(\mathbf{x},a\mathbf{y}+\mathbf{z}) = \overline{\phi(a\mathbf{y}+\mathbf{z},\mathbf{x})} = \bar{a}\overline{\phi(\mathbf{y},\mathbf{x})} + \bar{a}\overline{\phi(\mathbf{z},\mathbf{x})} = \bar{k}\phi(\mathbf{x},\mathbf{y}) + \bar{k}\phi(\mathbf{x},\mathbf{z})$. This is antilinearity in the second argument. If $F = \mathbb{R}$, the conjugations are irrelevant, and the inner product is both symmetric and linear in the second argument. This recovers all of the axioms for the bilinear form but also imposes symmetry, which the bilinear form does not have definitionally. A mapping which is linear in the first argument and antilinear in the second argument is often called sesquilinear (one and a half times linear) by mathematicians, although physicists often reverse the argument properties. *Bottom line*: an inner product for a real vector space is a symmetric bilinear form which is positive definite, and an inner product for a complex vector space is a conjugate symmetric sesquilinear form which is positive definite. Another notation for inner product is $\phi(\mathbf{u},\mathbf{v}) = \langle \mathbf{u},\mathbf{v}\rangle$, which steps on the toes of our vector component notation. The inner product defines a norm by the formula $\|\mathbf{v}\| = \sqrt{\phi(\mathbf{u},\mathbf{u})}$. Shortly we will introduce a version of vector multiplication called the "dot product", which is a real inner product defined for the space $\mathbb{R}^n$.

A point of slight confusion: note that it is required for any vector in $\mathbb{R}^n$ to have the origin as its initial point to conform to the axioms in Table 1. For example, a directed line segment constructed between the points $(1,1)$ and $(2,3)$ has the same length and direction as the vector $\langle 1,2\rangle$, but since it is not a position vector it is not directly expressible in the form $\langle a,b\rangle$. Hence it cannot legitimately participate in the operations defined for the vector space. Moreover, if we were to call it a vector, there would not be a unique vector corresponding to a specified length and direction. But there is a strong practical temptation to identify all directed line segments, no matter where their location, as vectors or at least vector-like. The resolution of this issue is to consider all directed line segments with the same length and direction as members of an equivalence class whose representative is always chosen as the one whose initial point is the origin, namely a bona fide vector. Then adding two arbitrary directed line segments becomes the same as adding their vector representatives.

The expressions **bound vector** and **free vector** are sometimes encountered in physics and engineering literature. Our definition of a vector in $\mathbb{R}^2$ or $\mathbb{R}^3$ corresponds to a bound vector, namely one with a specific initial point at the the origin. Arbitrary directed line segments are then free vectors which are elements of a so-called **affine space**. A not-entirely-facetious description of an affine space is "a vector space which has forgotten its origin". Every affine space has an associated vector space, so the free versus bound vector issue is perhaps a distinction without a difference. The specification of a directed line segment by its initial point (x_0,y_0,z_0) and its final point (x_1,y_1,z_1) automatically generates the corresponding bound vector $\langle x_1-x_0,y_1-y_0,z_1-z_0\rangle$, so all operations on directed line segments may be viewed as operations on bound equivalence class vector representatives. With the understanding that we are abusing our terminology, if we refer to a directed line segment as a vector we mean the bound vector that represents it.

Example 3 *Since the vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ are models of our physical world, they are called* **spatial**. *Give an example of a vector space that is not spatial.*

Solution: Consider the set of polynomials $\mathfrak{P}_x$ of any degree in the variable x with rational coefficients. A typical element is $a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ where $a_i \in \mathbb{Q}$. Define vector addition as polynomial addition where the coefficients of monomial terms of equal degree are added. Adding two rational numbers produces another rational, so sums remain in the set. Define multiplication by $c \in \mathbb{Q}$ as
$c(a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0) = ca_nx^n + ca_{n-1}x^{n-1} + \cdots + ca_1x + ca_0$. Multiplying two rational numbers results in another rational, so scalar multiples stay in the set. The additive identity

is the zero polynomial. The additive inverse is $-a_n x^n - a_{n-1}x^{n-1} - \cdots - a_1 x - a_0$. The other axioms are satisfied since $\mathbb{Q}$ is a field. We conclude $\langle \mathfrak{P}_x, \mathbb{Q}\rangle$ is a vector space, decidedly not spatial. We could try to emulate the manner in which we wrote the elements of $\mathbb{R}^n$ in Example 2. The typical element might be written as $(a_0, a_1, a_2, \cdots a_{n-1}, a_n)$ with the understanding that the coefficient in the i^{th} position applies to x^i. A moment's reflection reveals that we may need arbitrarily many positions in the parenthetical expression because the degree of a polynomial, although finite, can be arbitrarily high. Instead of an n-tuple we really need a ω-tuple with infinitely many positions. Of course, only finitely many entries in the ω-tuple could be nonzero to ensure that it represents a proper polynomial.

Example 4 *Show that the set of all real functions on an arbitrary nonvoid set is a vector space over* $\mathbb{R}$.

Solution: This example gives us a hint as to why the theory of vector spaces is so strongly connected to functional analysis. Let $X \neq \varnothing$ be any set and $\mathfrak{F}(X) = \{f : X \to \mathbb{R}\}$. The vectors are simply real functions on X, such as $f(x)$ and $g(x)$. For all $x \in X$ we define vector addition as $(f+g)(x) = f(x) + g(x)$, and scalar multiplication by $c \in \mathbb{R}$ as $(cf)(x) = cf(x)$. The constant zero function acts as the additive identity and for each $f(x)$ its negative is clearly the additive inverse, since $f(x) + (-f(x) = 0$. Axioms 1 through 4 are therefore satisfied and the remaining axioms depending on the field properties of $\mathbb{R}$ are verified as well.

Under some circumstances a subset of a vector space $\langle V, F\rangle$ will satisfy the axioms in Table 1 and be a vector space on its own, or a **vector subspace**. So if $W \subset V$ as a set, we would like to identify the conditions under which W is a vector space. Note that we are forming a subset of the vectors V and not a subfield of the scalar field F. First of all, as a subgroup of V, it must be the case that W is closed under vector addition, that is $\mathbf{x}, \mathbf{y} \in W$ must imply $\mathbf{x} + \mathbf{y} \in W$. Likewise, W must be closed under scalar multiplication, so if $c \in F$, we must have $c\mathbf{x} \in W$ for all $\mathbf{x} \in W$. All of the operations for elements of V are inherited by W, so associativity and commutativity of vector addition are preserved in W. Zero belongs to every field F so by closure of scalar multiplication in W, the element $0\mathbf{x} \in W$. Now certainly $(0+0)\mathbf{x} = 0\mathbf{x}$. Viewing $\mathbf{x}$ as an element of V (so the axioms apply) by Axiom 7 we can write $(0+0)\mathbf{x} = 0\mathbf{x} + 0\mathbf{x} = 0\mathbf{x}$. Then by Axiom 4 we add the additive inverse of $0\mathbf{x}$ to both sides and applying Axiom 3 we obtain $0\mathbf{x} + \mathbf{0} = \mathbf{0}$. Finally another application of Axiom 3 yields $0\mathbf{x} = \mathbf{0}$. So closure of the two operations implies that $\mathbf{0} \in W$ and clearly the additive inverse of the 1 in the scalar field can be multiplied by any $\mathbf{x} \in W$ to produce the additive inverse of $\mathbf{x}$. Now we have Axioms 1 through 4 satisfied for W and again the remaining axioms are valid for W due to the properties of the scalar field. We conclude that closure of vector addition and scalar multiplication in the set W is sufficient for $\langle W, F\rangle$ to be a subspace of $\langle V, F\rangle$.

We can distill the sufficient condition for a subset of a vector space to be a subspace into one condition:

Subspace Test for Vector Spaces

Given the vector space $\langle V,F\rangle$ and $W \subset V$ as a subset,

For every $\mathbf{w}_1$ and $\mathbf{w}_2$ in W, and every $c \in F$, if $c\mathbf{w}_1 + \mathbf{w}_2 \in W$, then $\langle W,F\rangle$ is a subspace of $\langle V,F\rangle$.

This condition implies closure for vector addition and scalar multiplication in W. For closure of addition, set $c = 1$ to show $1\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{w}_1 + \mathbf{w}_2 \in W$. We know $1\mathbf{w}_1 = \mathbf{w}_1$ as an element of V. For closure of scalar multiplication, set $\mathbf{w}_1 = \mathbf{w}_2$ and let $c = -1$, then $(-1)\mathbf{w}_2 + \mathbf{w}_2 = \mathbf{0}$ as an element of V, so evidently $\mathbf{0} \in W$. Now set $\mathbf{w}_2 = \mathbf{0}$ and we see $c\mathbf{w}_1 + \mathbf{0} = c\mathbf{w}_1 \in W$. Both operations are hence closed for the subset and conformance with all the axioms is verified. We conclude the subspace test is a sufficient condition for a subset to be a subspace.

Example 5 From Example 2 and the subsequent discussion we know that $\mathbb{R}^3$ is a vector space over $\mathbb{R}$. Vectors in $\mathbb{R}^3$ are specified by ordered triples $\langle x,y,z\rangle$ where $x,y,z \in \mathbb{R}$. Let us choose a subset W of these ordered triples where the third entry is zero. So $W = \{\langle x,y,0\rangle : x,y \in \mathbb{R}$. We claim $\langle W,\mathbb{R}\rangle$ is a vector space. Take $w_1 = \langle x_1,y_1,0\rangle$ and $w_2 = \langle x_2,y_2,0\rangle$ and any $c \in \mathbb{R}$. Form $cw_1 + w_2 = c\langle x_1,y_1,0\rangle + \langle x_2,y_2,0\rangle$. This can be rewritten as $\langle cx_1 + x_2, cy_1 + y_2, 0\rangle$. But this is the form of an element in W, so by the Subspace Test we conclude $\langle W,\mathbb{R}\rangle$ is a subspace of $\mathbb{R}^3$. You may recognize W as a plane in three dimensional space, in fact, it is precisely the xy plane. If we add a nonzero constant vector $\mathbf{z} \in \mathbb{R}^3$ to every element of the subspace W, we are **translating** W.by $\mathbf{z}$. In general, if M is a subspace of the vector space V and $\mathbf{x} \in V$, the translated subspace $\mathbf{x} + M$ is an **affine subspace** (or **linear manifold**) with the same dimension as M. It has "forgotten" its origin, so it is no longer technically a subspace of V, but vector operations still work (recall free vectors above). A maximal proper affine subspace of V is called a **hyperplane**. Any hyperplane in V necessarily has dimension one less than the dimension of V, or **codimension** 1. In our example, $W + \mathbf{z}$ is a hyperplane in $\mathbb{R}^3$.

4.2 LINEAR COMBINATIONS & SPANS

A linear combination of vectors from the space $\langle V,F\rangle$ is simply a sum of vectors with various scalar coefficients applied. So $a\mathbf{x} + b\mathbf{y} + c\mathbf{z} = \mathbf{u}$ is a typical linear combination. If we choose $V = \mathbb{R}^3$, this combination would look like $a\langle x_1,x_2,x_3\rangle + b\langle y_1,y_2,y_3\rangle + c\langle z_1,z_2,z_3\rangle = \langle u_1,u_2,u_3\rangle$. Applying the scalar multiplications and then performing the vector additions we get $\langle ax_1 + bx_1 + cx_1, ax_2 + bx_2 + cx_2, ax_3 + bx_3 + cx_3\rangle = \langle u_1,u_2,u_3\rangle$. Matching the respective entries in the ordered triples we see that $u_1 = ax_1 + bx_1 + cx_1$, $u_2 = ax_2 + bx_2 + cx_2$, and

$u_3 = ax_3 + bx_3 + cx_3$.

Example 1 *In $\mathbb{R}^2$ we have the linear combination $a\langle 1,2\rangle + b\langle 1,3\rangle = \langle -2,5\rangle$. Determine a and b.*

Solution: Matching entries we obtain two equations: (i) $a + b = -2$ and (ii) $2a + 3b = 5$. Solving these simultaneously we have $a = -11$ and $b = 9$.

Example 2 *In $\langle \mathfrak{P}_x, \mathbb{Q}\rangle$ (see Example 3 in §5.1) we have the linear combination $a(x^3 - 3x^2 + 1) + b(2x^3 + 4x^2 + 3x - 5) + c(3x^3 + 2x^2 - 2) + d(x^2 - x + 1) = (11x^3 + 8x^2 + 7x - 14)$. Determine $a, b, c,$ and d.*

Solution: Matching entries we obtain four equations: (i) $a + 2b + 3c = 11$, (ii) $-3a + 4b + 2c + d = 8$, (iii) $3b - d = 7$, and (iv) $a - 5b - 2c + d = -14$. Solving simultaneously we have $a = 1$, $b = c = 2$, and $d = -1$.

If we have a set of vectors, say $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \in \langle V, F\rangle$, we might ask what other vectors could be generated as linear combinations of these. The answer is any vector that can be expressed as $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n$ where $a_i \in F$ for $1 \le i \le n$. We say the **linear span** (or just the **span**) of the set $S = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset V$ is the set $sp(S) = \{\sum_{i=1}^{n} a_i\mathbf{x}_i : a_i \in F\}$. We don't rule out the possibility that $S = \varnothing$, in which case $sp(S) = \{\mathbf{0}\}$. If S spans W, we say S is a **spanning set** of W. Note that the span of a set of vectors from V is always a subspace of V. It should be clear that if $W \subset V$ and $S \subset W$, then $sp(S) \subset W$. In other words, no vector generated as a linear combination of vectors already in W can escape from being just an element of W.

4.3 LINEAR INDEPENDENCE & BASES

We define a set of *nonzero* vectors $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ to be **linearly independent** (or simply **independent**) if every linear combination that equals the zero vector forces every coefficient in the combination to be zero. So if $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$, then $a_i = 0$ for $1 \le i \le n$. This means that there is no redundancy in the set with regard to generating their linear span. Suppose that $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ is linearly independent and $sp(\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}) = sp(\{\mathbf{x}_2, \cdots, \mathbf{x}_n\})$. This would mean that $\mathbf{x}_1$ could be written as a linear combination of the vectors in the set $\{\mathbf{x}_2, \cdots, \mathbf{x}_n\}$. So $\mathbf{x}_1 = b_2\mathbf{x}_2 + \cdots + b_n\mathbf{x}_n$. This means $\mathbf{x}_1 - b_2\mathbf{x}_2 + \cdots + b_n\mathbf{x}_n = \mathbf{0}$, and by linear independence all the coefficients (including 1) would have to be zero. This is impossible, so every member of the original set is essential for generating the span. This is obviously true for any member of the spanning set. A set of vectors that is not linearly independent is said to be **dependent**.

Given a set of vectors S in the space V, it may be the case that $sp(S) = V$ and also the vectors in S are linearly independent. In that case we call S a **basis** for V. It is beyond the scope of our discussion, but it can be shown that every vector space has a spanning set that is maximal with regard to linearly independence. A basis is not unique, but as we will show there is a property shared by all bases of a particular vector space and that is the number of vectors in the basis. This number is called the **dimension** of the vector space. A vector space V with a finite basis consisting of n **basis vectors** is said to have dimension n. This can be appreviated to $\dim V = n$. If the space does not have a finite basis we say it is infinite dimensional. Examples of finite dimensional vector spaces are the $\mathbb{R}^n$ spaces in §4.1, Example 2. An example of an infinite dimensional space is $\langle \mathfrak{P}_x, \mathbb{Q} \rangle$ in §4.1, Example 3. Sometimes we qualify a basis as being a **Hamel basis** or **algebraic basis** to distinguish it from other types of bases which are defined in terms of different properties.

Example 1 *Show that if we are given a vector space V with finite basis $B = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$, there is only one way to write an arbitrary $\mathbf{v} \in V$ in terms of the basis vectors.*

Solution: For the sake of contradiction, suppose $\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n$ and also that $\mathbf{v} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \cdots + b_n\mathbf{e}_n$. Subtracting the second from the first expression we get $(a_1 - b_1)\mathbf{e}_1 + (a_2 - b_2)\mathbf{e}_2 + \cdots + (a_n - b_n)\mathbf{e}_n = \mathbf{0}$. Since the basis vectors are linearly independent, it must be the case that $a_i - b_i = 0$ for $1 \leq i \leq n$. But this means $a_i = b_i$ and we see that the expression of any vector with respect to the given basis is unique.

Example 2 *Show that if the set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ spans the vector space V, then the set $\{\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is linearly dependent, where $\mathbf{w} \neq 0$.*

Solution: We know $\mathbf{w}$ is some linear combination of the $\mathbf{v}_i$, say $\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$. Then $\mathbf{w} - a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$. At least one $a_i \neq 0$, so the condition for linear independence is violated.

Example 3 Show that any set of $n + 1$ vectors in a vector space V of dimension n is linearly dependent.

Solution: Suppose we have an arbitrary set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n, \mathbf{v}_{n+1}\}$ in V which has a basis $B = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$. For the sake of contradiction, suppose the vectors in S are independent. This requires that the test expression $\beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_{n+1}\mathbf{v}_{n+1} = \mathbf{0}$ forces each $\beta = 0$. First express each $\mathbf{v}_i$ in terms of the basis vectors. We have $\mathbf{v}_1 = \sum_{j=1}^{n} a_{1j}\mathbf{e}_j$, and in general $\mathbf{v}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j$ for $1 \leq i \leq n + 1$. Substituting into the test expression, we get $\beta_1 \sum_{j=1}^{n} a_{1j}\mathbf{e}_j + \beta_2 \sum_{j=1}^{n} a_{2j}\mathbf{e}_j + \cdots + \beta_{n+1} \sum_{j=1}^{n} a_{(n+1)j}\mathbf{e}_j = \mathbf{0}$. Rewriting this sum by collecting coefficients of the basis vectors, we obtain $\mathbf{e}_1 \sum_{k=1}^{n+1} a_{k1}\beta_k + \mathbf{e}_2 \sum_{k=1}^{n+1} a_{k2}\beta_k + \cdots + \mathbf{e}_n \sum_{k=1}^{n+1} a_{kn}\beta_k = \mathbf{0}$. Suppose we set each coefficient sum for an $\mathbf{e}_i$ to zero. This creates a system of n equations, one for each basis vector, in $n + 1$ variables, namely the β_i. The under-determined system then has one free variable, allowing us to construct a solution with some nonzero β_i. But this contradiction establishes that the $n + 1$ vectors cannot be independent. This result clearly extends to any set of $n + k$ vectors, where $k \geq 1$.

Example 4 Show that the dimension of a finite dimensional vector space V is well defined.

Solution: Consider all linearly independent spanning sets for $V \neq \varnothing$. There is at least one. By well-ordering there is a minimum cardinality n among these sets. Since such a set would constitute a basis for V, and any set with $n+1$ or more vectors could not be a basis by Example 3, we conclude that $\dim V = n$.

4.4 VECTORS IN COMPONENT FORM

One of the many advantages of identifying a basis for a vector space is that arbitrary vectors can be expressed in a uniform way. For example, we can show that a basis for the vector space $\mathbb{R}^2$ is $\{\langle 1,0\rangle,\langle 0,1\rangle\}$. It is traditional to denote the basis vectors $\langle 1,0\rangle$ by $\hat{\mathbf{i}}$ and $\langle 0,1\rangle$ by $\hat{\mathbf{j}}$. An arbitrary vector $\langle x,y\rangle$ can then be written $x\hat{\mathbf{i}} + y\hat{\mathbf{j}}$. This is yet another way to distinguish points in $\mathbb{R}^2$ from vectors in $\mathbb{R}^2$. §4.3 Example 1 assures us that writing a vector in this or any other manner relative to a fixed basis is unique. We call $\hat{\mathbf{i}}$ the **unit vector** in the x-direction and $\hat{\mathbf{j}}$ the corresponding unit vector in the y-direction. The vector space $\mathbb{R}^3$ is an easy extension of this. We add $\hat{\mathbf{k}}$, the unit vector in the z-direction, and $\langle x,y,z\rangle$ becomes $x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\,\hat{\mathbf{k}}$. The set $\{\hat{\mathbf{i}},\hat{\mathbf{j}},\hat{\mathbf{k}}\}$ is called the **standard basis** for $\mathbb{R}^3$ to distinguish it from other possible bases. We assume the standard basis is in use for $\mathbb{R}^2$ and $\mathbb{R}^3$ unless otherwise noted. The **components** of a vector relative to a given basis are the coefficients of the linear combination of basis vectors expressing the vector. So the x-component of $\langle 3,-5,6\rangle$ is 3, and so forth. The space $\langle \mathfrak{P}_x,\mathbb{Q}\rangle$ has basis $B = \{x^n : n \in \mathbb{N}_0\}$. This has to be an infinite spanning set in order to express all possible rational polynomials of arbitrary but finite degree. The condition of linear independence is applied just to finite subsets of B. Relative to B the polynomial $3x^{10} - 2x^8 + 5x^7 - 4x^6 + 7x^4 + 11x^3 - x^2 + x + 6$ would be written $\langle 6,1,-1,11,7,0,-4,5,-2,0,2,0,\cdots,0,\cdots\rangle$. Some advanced books on algebra start with this expression to define a polynomial.

Vector addition, subtraction, and scalar multiplication can all be done component-wise. So in $\mathbb{R}^3$ we would have $\langle 8,-2,7\rangle + \langle 4,5,11\rangle = \langle 12,3,18\rangle$ and $3\langle 7,-5,9\rangle = \langle 21,-15,27\rangle$. Extension to other vector spaces is immediately obvious.

Example 1 *A cruise ship leaves port due east and travels for ten hours at fifteen knots, then turns south and steams for five hours at twenty knots, then turns west and goes for 3 hours at ten knots. How far from port is it and which direction should it take to return?*

Solution: This is a case where if we drew the port and waypoints to scale on a map, some of the displacements would be directed line segments (free vectors). The task is to find the vectors in $\mathbb{R}^2$ that represent displacements along each leg of the journey. The answer to the question will be the vector that is the additive inverse of the sum of the three legs. We don't have displacements yet...the problem data is in terms of speed and time. Orient the xy-plane so the east aligns with positive values of x. Measuring distance in (nautical) miles the eastward leg is $10 \cdot 15 = 150$ miles, so the first leg displacement vector is $\langle 150, 0\rangle$. The second is $5 \cdot 20 = 100$ miles, so that vector is $\langle 0, -100\rangle$. The length of the third vector is $3 \cdot 10 = 30$ miles, so that vector is $\langle -30, 0\rangle$. The sum of these three vectors represents where the ship is. So the vector describing its position is $\langle x, y\rangle = \langle 150, 0\rangle + \langle 0, -100\rangle + \langle -30, 0\rangle = \langle 120, -100\rangle$. The vector that leads from that position to port is then $\langle -120, 100\rangle$. The norm of this vector is $\sqrt{(-120)^2 + 100^2} = 156.2$ miles. The direction in which to proceed back to port would be $\theta = 129.8^o$ for an engineer or $N50.2^oW$ for an actual sailor. Recall θ is the angle in the xy-plane measured counterclockwise from the positive x-axis.

Example 2 *A $50\ N$ force is acting in the direction of the line segment running from $(1,2,3)$ to $(4,7,10)$ Express this force as a vector.*

Solution: The force vector must have a magnitude of 50 newtons. The direction of the force will be the same as the direction of the line segment starting at $(1,2,3)$ and terminating at $(4,7,10)$. This line segment extends $4 - 1 = 3$ units in the x-direction, $7 - 2 = 5$ units in the y-direction, and $10 - 3 = 7$ units in the z-direction. So the corresponding vector is $\langle 3,5,7\rangle$. The norm of $\langle 3,5,7\rangle$ is $\sqrt{3^2 + 5^2 + 7^2} = \sqrt{83} = 9.11$, so it captures the direction but not the correct magnitude of the force. It would be very convenient to have a vector of length 1 in the direction of $\langle 3,5,7\rangle$. The operation of **unitizing** a nonzero vector is always possible and easy to do. We simply divide any vector by its norm to produce a unit vector in the same direction. So $\frac{1}{9.11}\langle 3,5,7\rangle$ saves the direction of $\langle 3,5,7\rangle$ but has length 1. The force vector we want has this direction but norm 50, so our solution is $\mathbf{F} = \frac{50}{9.11}\langle 3,5,7\rangle$ $N = \langle 16.47, 27.44, 38.42\rangle$. Checking, $\sqrt{(16.47)^2 + (27.44)^2 + (38.42)^2} \approx 50\ N$.

Unitizing a Vector

Given vector $\mathbf{v} \neq \mathbf{0}$, a parallel unit vector $\mathbf{u_v}$ is $\frac{\mathbf{v}}{|\mathbf{v}|}$

Given $\mathbf{v} = \langle a,b,c\rangle$, $\mathbf{u_v} = \frac{\langle a,b,c\rangle}{\sqrt{a^2 + b^2 + c^2}}$

Similarly for $\mathbf{v} \in \mathbb{R}^n$

Example 3 *Three forces in a common plane are acting at the same point on a structural connection. Force #1 is ten tons pulling due north. Force #2 is five tons pulling due west. Force #3 is fifteen tons pulling southeast. What additional force and in what direction will balance the existing forces?*

Solution: The vectors corresponding to these forces are respectively $\langle 0,10\rangle$, $\langle -5,0\rangle$, and $\frac{\sqrt{2}}{2}\langle 15,-15\rangle$. Adding component-wise we get the net force $\mathbf{F} = \langle 5.61, -0.61\rangle$ tons. So the balancing force is $-\mathbf{F} = \langle -5.61, 0.61\rangle$ tons.

4.5 DOT PRODUCT

As we mentioned above in connection with general inner product spaces, there is a type of multiplication between vectors in $\mathbb{R}^n$ called the **dot product**. The dot product combines two vectors in $\mathbb{R}^n$ and returns us to $\mathbb{R}$. Viewed as the mapping "$\cdot$" we have "$\cdot$": $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, where "$\cdot$"$(\langle x_i \rangle, \langle y_j \rangle) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \delta_{ij}$. The bookkeeping symbol δ_{ij} is known as the **Kronecker delta** and is defined to be 1 if $i = j$ and 0 otherwise. Practically speaking, all we are doing is multiplying the corresponding components of two n-dimensional vectors and adding up all the individual products. So in $\mathbb{R}^3$, for example, $\langle 2,5,8 \rangle \cdot \langle -1,5,7 \rangle = (2)(-1) + (5)(5) + (8)(7) = -2 + 25 + 56 = 79$. All of our applications will be in $\mathbb{R}^2$ or $\mathbb{R}^3$, but it is worth knowing that the construction is valid for arbitrary n. The dot product is a well-behaved operation. The following computational rules are simply an adaptation of the inner product axioms given in Table 4.

Table 5 - Dot Product Rules for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$

1) Commutativity	$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
2) Distributivity over vector addition	$\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$
3) Associativity of scalar multiplication	$c(\mathbf{x} \cdot \mathbf{y}) = (c\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (c\mathbf{y})$
4) Annihilation	$\mathbf{0} \cdot \mathbf{x} = 0$
5) Norm	$\sqrt{\mathbf{x} \cdot \mathbf{x}} = \lvert\mathbf{x}\rvert$

There are some particularly convenient formulas involving the dot product. They help reduce some sticky problems in 2 and 3 dimensional geometry to almost routine calculations.

Example 1 *Show that the angle θ between two vectors $v, w \in \mathbb{R}^3$ is given by* $\theta = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{\lvert\mathbf{v}\rvert\lvert\mathbf{w}\rvert}\right)$.

Solution: The terminal points of the two vectors along with their common initial point determine a plane, so this problem immediately devolves to $\mathbb{R}^2$. The Law of Cosines is a generalization of the Pythagorean formula. It supplies a correction term that applies when the "right" angle fails to be right. Suppose a scalene triangle has sides a, b and c where we want to calculate the length of c in terms of a and b but the angle θ opposite side c fails to be 90^o. The Law of Cosines says $c^2 = a^2 + b^2 - 2ab\cos\theta$. Now let us use this in the context of vectors.

Suppose $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ and $\mathbf{w} = \langle w_1, w_2, w_3 \rangle$. The directed line segment that runs from the terminal point of $\mathbf{v}$ to the terminal point of $\mathbf{w}$ completes a triangle with $\mathbf{v}$ and $\mathbf{w}$ as the other sides. The vector $\mathbf{w} - \mathbf{v}$ has the same length and orientation as this line segment. That length is $|\mathbf{w}-\mathbf{v}| = \sqrt{(w_1-v_1)^2+(w_2-v_2)^2+(w_3-v_3)^2}$. Now identify this expression as c in the Law of Cosines and likewise treat a as $|\mathbf{v}|$ and $|\mathbf{w}|$ as b. With these substitutions we have $(w_1-v_1)^2+(w_2-v_2)^2+(w_3-v_3)^2 = (v_1^2+v_2^2+v_3^2)+(w_1^2+w_2^2+w_3^2)-2(|\mathbf{v}||\mathbf{w}|)\cos\theta$. Bit of a mess at this point, but wait...expanding the left hand side we get $(w_1^2-2w_1v_1+v_1^2)+(w_2^2-2w_2v_2+v_2^2)+(w_3^2-2w_3v_3+v_3^2) = (v_1^2+v_2^2+v_3^2)+(w_1^2+w_2^2+w_3^2)-2(w$ Now harvest the cancellations on both sides to get $-2(w_1v_1+w_2v_2+w_3v_3) = -2(|\mathbf{v}||\mathbf{w}|)\cos\theta$. Dividing by -2 and noticing that $w_1v_1+w_2v_2+w_3v_3 = \mathbf{v}\cdot\mathbf{w}$, we finally have $\mathbf{v}\cdot\mathbf{w} = |\mathbf{v}||\mathbf{w}|\cos\theta$ or $\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{v}||\mathbf{w}|}$. The desired formula follows immediately.

Example 2 *A one hundred foot tall building with a square fifty foot by fifty foot base is oriented so that one wall faces due east. A drain pipe is to be run from the northeast corner of the roof to the southwest corner of the base. The pipe will descend along the east wall fifty feet to the southeast corner of the building, then turn and descend another fifty feet along the south wall to its destination. An angled fitting must be manufactured to accomodate the turn at the southeast corner. What should be the angle?*

Solution: A 90^o fitting won't work because the drain is turning through 90^o at the same time as it is descending. We will impose a coordinate system on the building and locate the origin at the point of the fitting. Then the two runs of pipe will be represented by vectors, and our formula will give us the necessary angle. Letting the z-axis coincide with the southeast corner of the building extending towards the roof, we then locate the y-axis in the south face of the building. Now we can find the components of the two vectors. The pipe running from the origin to the roof has no extension in the y-direction, moves -50 feet in the x-direction, and 50 feet in the z-direction. So its vector is $\langle -50, 0, 50\rangle$. The pipe descending across the south face has no extension in the x-direction, moves -50 feet in the y-direction, and -50 feet in the z-direction. Then its vector is $\langle 0, -50, -50\rangle$. We compute the fitting angle

$$\theta = \arccos\left(\frac{\langle -50,0,50\rangle\cdot\langle 0,-50,-50\rangle}{|\langle -50,0,50\rangle||\langle 0,-50,-50\rangle|}\right) = \arccos\left(\frac{-2500}{\left(\sqrt{5000}\right)^2}\right) = \arccos\left(-\frac{1}{2}\right) = 120^o.$$

Example 3 *Show that* $\mathbf{v}\cdot\mathbf{w} = |\mathbf{v}||\mathbf{w}|\cos\theta$ *and therefore the dot product detects perpendicularity (**orthogonality**).*

Solution: This expression for the dot product is often used as the fundamental definition, but it has the disadvantage that the angle between vectors needs to be known. It is an obvious consequence of the calculation in Example 1 that if the dot product of two nonzero vectors is zero, then the cosine of the angle between them must be zero. It follows that the acute angle must be $\frac{\pi}{2}$ or 90^o.

Example 4 *Show that thescalar **projection** of the vector* $\mathbf{v}$ *in the direction of vector* $\mathbf{w}$ *is* $\frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|}$.

Solution: The scalar projection of a vector is the length of its "shadow" in some direction. Specifically it is the amount of the vector's length that aligns with a given direction.

If that direction is given in the form of a unit vector $\mathbf{u}$, which makes an angle θ with $\mathbf{v}$, then the projection of $\mathbf{v}$ in that direction is $|\mathbf{v}|\cos\theta = \mathbf{v}\cdot\mathbf{u}$. Note that $\frac{\mathbf{w}}{|\mathbf{w}|}$ is a unit vector in the direction of $\mathbf{w}$, so $\mathbf{v}\cdot\frac{\mathbf{w}}{|\mathbf{w}|}$ is precisely the projection of $\mathbf{v}$ in the direction of vector $\mathbf{w}$.

4.6 CROSS PRODUCT

In the vector space $\mathbb{R}^3$ there is another type of multiplication between vectors called the **cross product**. Viewed as a mapping "$\times$" we have "$\times$": $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, so we are combining two vectors and the result is another vector. It may seem odd, but this version of vector multiplication only works in $\mathbb{R}^3$ for reasons that are beyond the scope of our discussion. However this won't prevent us from using it effectively for practical problems. The definition of the cross product may seem unduly complicated at first, but once we appreciate its utility that feeling should give way to respect. To economize on notation we are going to give the standard unit vectors in $\mathbb{R}^3$ temporary new names: $\widehat{\mathbf{i}} = \widehat{\mathbf{e}}_1$, $\widehat{\mathbf{j}} = \widehat{\mathbf{e}}_2$, and $\widehat{\mathbf{k}} = \widehat{\mathbf{e}}_3$. Then we define $\mathbf{x}\times\mathbf{y} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\varepsilon_{ijk}\widehat{\mathbf{e}}_i x_j y_k$ The bookkeeping symbol ε_{ijk} is known as the **Levi-Civita** symbol and is defined to be 1 if ijk is an even permutation of 123, -1 if ijk is an odd permutation of 123, and 0 for any other case. The even permutations of 123 are 123, 312, and 231. These retain the original order with only "wraparounds" permitted. The odd permutations are 132, 213, and 321. So apparently only six of the twenty-seven possible terms of the triple sum survive. We are thankful for that. More explicitly, if $\mathbf{x} = \langle x_1,x_2,x_3\rangle$ and $\mathbf{y} = \langle y_1,y_2,y_3\rangle$ then $\mathbf{x}\times\mathbf{y} = \widehat{\mathbf{i}}(x_2y_3 - x_3y_2) + \widehat{\mathbf{j}}(x_3y_1 - x_1y_3) + \widehat{\mathbf{k}}(x_1y_2 - x_2y_1)$. After we discuss matrices and determinants, we will have a scheme using formal determinants that organizes this calculation for us. Conveniently enough, it is possible to extend this formula to vectors in $\mathbb{R}^2$ by artificially giving each vector a zero component in the $\widehat{\mathbf{k}}$ direction. Of course, the result will turn out to be a vector in $\mathbb{R}^3$ with only a $\widehat{\mathbf{k}}$ component.

The cross product is less well-behaved than the dot product.

Table 6 - Cross Product Rules for $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathbb{R}^3$

1) Anti-commutativity	$\mathbf{x}\times\mathbf{y} = -\mathbf{y}\times\mathbf{x}$
2) Left distributivity over vector addition	$\mathbf{x}\times(\mathbf{y}+\mathbf{z}) = \mathbf{x}\times\mathbf{y}+\mathbf{x}\times\mathbf{z}$
3) Right distributivity over vector addition	$(\mathbf{y}+\mathbf{z})\times\mathbf{x} = \mathbf{y}\times\mathbf{x}+\mathbf{z}\times\mathbf{x}$
4) Associativity of scalar multiplication	$c(\mathbf{x}\times\mathbf{y}) = (c\mathbf{x})\times\mathbf{y} = \mathbf{x}\times(c\mathbf{y})$
5) Annihilation	$\mathbf{0}\times\mathbf{x} = \mathbf{0}$

Example 1 *An alternative definition of the cross product is* $\mathbf{x}\times\mathbf{y} = |\mathbf{x}||\mathbf{y}|\sin\theta\,\widehat{\mathbf{e}}_\perp$, *where* θ *is the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *and* $\widehat{\mathbf{e}}_\perp$ *is a unit vector perpendicular to the common*

plane of $\mathbf{x}$ *and* $\mathbf{y}$ *and points in the direction determined by the* ***right-hand-rule****. Show that this is equivalent to the definition above.*

Solution: Consider the following eight equations which develop the puzzle pieces in an orderly way. Much cancellation will be possible.

(i)

$(\mathbf{x} \times \mathbf{y}) \cdot (\mathbf{x} \times \mathbf{y}) = \left(\hat{\mathbf{i}}(x_2y_3 - x_3y_2) + \hat{\mathbf{j}}(x_3y_1 - x_1y_3) + \hat{\mathbf{k}}(x_1y_2 - x_2y_1)\right) \cdot \left(\hat{\mathbf{i}}(x_2y_3 - x_3y_2) + \hat{\mathbf{j}}(x_3y_1 -\right.$

(ii) Expanding the squares...

$(x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 + (x_1y_2 - x_2y_1)^2 = (x_2y_3)^2 - 2(x_2y_2x_3y_3) + (x_3y_2)^2 + (x_3y_1)^2 - 2(x_3$

then regrouping...

(iii)

$(x_2y_3)^2 - 2(x_2y_2x_3y_3) + (x_3y_2)^2 + (x_3y_1)^2 - 2(x_3y_1x_1y_3) + (x_1y_3)^2 + (x_1y_2)^2 - 2(x_1y_2x_2y_1) + (x_2y_1)$

(iv) Generating similar terms...$|\mathbf{x}|^2|\mathbf{y}|^2 = (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) = (x_1^2y_1^2 + x_1^2y_2^2 + x_1^2y_3^2) + (x_2^2y_1^2 + x_2^2y_2^2 + x_2^2y_3^2) + (x_3^2y_1^2$

(v) More similar terms...$(\mathbf{x} \cdot \mathbf{y})^2 = (x_1y_1 + x_2y_2 + x_3y_3)^2 = (x_1y_1)^2 + (x_2y_2)^2 + (x_3y_3)^2 + x_1y_1x_2y_2 + x_1y_1x_3y_3 + x_2y_2$

(vi) Regrouping (v)...$(\mathbf{x} \cdot \mathbf{y})^2 = (x_1y_1)^2 + (x_2y_2)^2 + (x_3y_3)^2 + 2[x_1y_1x_2y_2 + x_1y_1x_3y_3 + x_2y_2x_3y_3]$

(vii) Setting up cancellations...$|\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2 = [(x_1^2y_1^2 + x_1^2y_2^2 + x_1^2y_3^2) + (x_2^2y_1^2 + x_2^2y_2^2 + x_2^2y_3^2) + (x_3^2y_1^2 + x_3^2y_2^2 + x_3^2y$

(viii) Harvesting cancellations...

$|\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2 = [(x_2^2y_1^2 + x_2^2y_2^2 + x_2^2y_3^2) + (x_3^2y_1^2 + x_3^2y_2^2 + x_3^2y_3^2)] - 2[x_2y_2x_3y_3 + x_3y_1x_1y_3 + x_1y_2x_2y_1$

Compare equation (iii) with equation (viii) and assemble the puzzle pieces to get $(\mathbf{x} \times \mathbf{y}) \cdot (\mathbf{x} \times \mathbf{y}) = |\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2$. Since $(\mathbf{x} \cdot \mathbf{y})^2 = |\mathbf{x}|^2|\mathbf{y}|^2 \cos^2\theta$, evidently $|\mathbf{x} \times \mathbf{y}|^2 = |\mathbf{x}|^2|\mathbf{y}|^2(1 - \cos^2\theta) = |\mathbf{x}|^2|\mathbf{y}|^2 \sin^2\theta$. Upon taking square roots we finally get $|\mathbf{x} \times \mathbf{y}| = |\mathbf{x}||\mathbf{y}| \sin\theta$.

Now for the matter of $\hat{\mathbf{e}}_\perp$. Notice that $\mathbf{x} \cdot (\mathbf{x} \times \mathbf{y}) = x_1(x_2y_3 - x_3y_2) + x_2(x_3y_1 - x_1y_3) + x_3(x_1y_2 - x_2y_1) = 0$. Likewise $\mathbf{y} \cdot (\mathbf{x} \times \mathbf{y}) = 0$. We know that a zero dot product is an indicator of orthogonality, so we can conclude that $\mathbf{x} \times \mathbf{y}$ is perpendicular to the common plane of $\mathbf{x}$ and $\mathbf{y}$. Using the convention of a right-handed coordinate system, we will choose the direction of $\hat{\mathbf{e}}_\perp$ to be given by the right-hand rule. If the fingers of the right hand are extended in the direction of $\mathbf{x}$ and then curled in the direction of $\mathbf{y}$, the thumb will point in the direction of $\hat{\mathbf{e}}_\perp$. This completes the verification of the equivalence of the two definitions. As with the dot product, the trade-off for using the simpler definition is having to know the angle between vectors in advance.

Example 2 *If two vectors form adjacent sides of a parallelogram, find the area of the figure.*

Solution: Let $\mathbf{x}$ and $\mathbf{y}$ be the two vectors. The length of an altitude of the parallelogram drawn from the terminal point of $\mathbf{x}$ and intersecting $\mathbf{y}$, or if necessary a parallel extension of $\mathbf{y}$, would be $|\mathbf{x}| \sin\theta$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. The base of the figure is $|\mathbf{y}|$. So the area of the parallelogram is base times altitude or $|y||x| \sin\theta$. This happens to be $|\mathbf{x}\times\mathbf{y}|$. For a triangle take half, so the area would be $\frac{1}{2}|\mathbf{x}\times\mathbf{y}|$.

Example 3 *Find the area of a triangle in* $\mathbb{R}^3$ *with vertices at* $(1,2,3)$, $(3,5,7)$*, and* $(-1.-1,4)$.

Solution: *Label the points* P_1, P_2, and P_3 respectively. The line segment connecting P_1 and P_2 will be written $\overline{P_1P_2}$, and so forth. Pick any point...P_1 will do...and find the vectors that represent the line segments $\overline{P_1P_2}$ and $\overline{P_1P_3}$. For $\overline{P_1P_2}$ we have $\langle 3-1, 5-2, 7-3\rangle$ or $\mathbf{x} = \langle 2,3,4\rangle$. For $\overline{P_1P_3}$ we have $\langle -1-1, -1-2, 4-3\rangle$ or $\mathbf{y} = \langle -2,-3,1\rangle$. We don't know the angle between these vectors, so we resort to the component definition of cross product $\mathbf{x}\times\mathbf{y} = \hat{\mathbf{i}}(x_2y_3 - x_3y_2) + \hat{\mathbf{j}}(x_3y_1 - x_1y_3) + \hat{\mathbf{k}}(x_1y_2 - x_2y_1)$. Then in this case, $\mathbf{x}\times\mathbf{y} = \hat{\mathbf{i}}(3\cdot 1 - 4(-3)) + \hat{\mathbf{j}}(4(-2) - 2\cdot 1) + \hat{\mathbf{k}}(2(-3) - 3(-1))$. Simplified, this is the vector $\langle 15,-10,3\rangle$. The area is $\frac{1}{2}|\langle 15,-10,3\rangle| = \frac{1}{2}\sqrt{225+100+9} = 9.14$ units. This method is easier to use in practice than a purely analytical geometry approach.

Example 4 *Show that the cross product detects parallelism and antiparallelism.*

Solution: For nonzero vectors, if $\mathbf{x}\times\mathbf{y} = |\mathbf{x}||\mathbf{y}|\sin\theta\,\hat{\mathbf{e}}_\perp = \mathbf{0}$, it must be the case that $\sin\theta = 0$, or $\theta = 0^o$ for parallelism or 180^o for antiparallelism.

Example 5 $(\int)$ With just the triangle area formula and some very basic differential calculus we can prove Kepler's Second Law, namely that the planets travel in their orbits in such a way that a line connecting the sun to a planet sweeps out equal areas in equal times. This was cutting-edge physics in the 1590's. The practical effect is that when a planet is near the sun (near perihelion) it travels faster than when it is far from the sun (near aphelion). Kepler's First Law states that the planets travel around the sun in elliptical orbits with the sun at one focus. He deduced these laws from extensive naked-eye astrolabe observations prior to 1600 (the telescope was invented in 1608 in Holland) without any prior valid quantitative theory of planetary motion. We present this calculation as a testament to the power of vector methods.

Suppose the vector connecting the sun to a planet is $\mathbf{r}(t)$, where t is time. After an increment Δt of time passes, the vector changes to $\mathbf{r}(t+\Delta t)$ and the planet travels $\Delta\mathbf{r}(t) = \mathbf{r}(t+\Delta t) - \mathbf{r}(t)$. Since Δt is assumed small, we can regard the area swept out by the radius vector during Δt to be a (very thin) triangle. Now $\mathbf{r}(t)$ and $\Delta\mathbf{r}(t)$ are two sides of this triangle, so its area $A(t) = \frac{1}{2}|\mathbf{r}(t)\times\Delta\mathbf{r}(t)|$. We would like to develop an expression for $\frac{dA}{dt}$, the rate of sweeping out area. Specifically, we want to show it is a constant. To do that we will need to look at $\frac{d^2A}{dt^2}$ and show it is zero. So $\frac{dA}{dt} = \lim_{\Delta t\to 0}\frac{\Delta A}{\Delta t} = \lim_{\Delta t\to 0}\frac{1}{2}\left|\mathbf{r}(t)\times\frac{\Delta\mathbf{r}(t)}{\Delta t}\right|$. Evidently $\frac{dA}{dt} = \frac{1}{2}\left|\mathbf{r}(t)\times\lim_{\Delta t\to 0}\frac{\Delta\mathbf{r}(t)}{\Delta t}\right| = \frac{1}{2}\left|\mathbf{r}(t)\times\frac{d\mathbf{r}(t)}{dt}\right|$. Now we use the product rule to differentiate a cross product (see Table 7 below). It works the same way as for a regular scalar product as long as the factors are kept in the same order (recall the cross product does not commute). The result is $\frac{d^2A}{dt^2} = \frac{1}{2}\frac{d}{dt}\left|\mathbf{r}(t) \times \frac{d\mathbf{r}(t)}{dt}\right| = \frac{1}{2}\left|\frac{d\mathbf{r}(t)}{dt} \times \frac{d\mathbf{r}(t)}{dt} + \mathbf{r}(t) \times \frac{d^2\mathbf{r}(t)}{dt^2}\right|$. The first cross product vanishes because a vector is parallel to itself (see Example 4). In the second cross product, the term $\frac{d^2\mathbf{r}(t)}{dt^2}$ is the acceleration of the planet along the radius. The force of gravity causes this acceleration and it is given by Newton's Gravity Law $\mathbf{F}(r) = -\frac{GM_s m_p}{r^2}\widehat{e}_r$. In this formula the force of gravity acting on the planet at distance $|\mathbf{r}| = r$ from the sun is $\mathbf{F}(r)$, G is the gravitational constant, and M_s and m_p are the masses of the sun and planet, respectively. The unit vector $\widehat{e}_r = \frac{\mathbf{r}(t)}{|\mathbf{r}(t)|}$ points from the sun towards the planet. Since $\mathbf{F}(r) = m_p\frac{d^2\mathbf{r}(t)}{dt^2}$ (Newton's Second Law of Motion) we see that $\widehat{e}_r$ and $\frac{d^2\mathbf{r}(t)}{dt^2}$ are parallel. Hence the second cross product also vanishes and we conclude $\frac{d^2A}{dt^2} = 0$. Since this is the derivative of $\frac{dA}{dt}$, it must be that $\frac{dA}{dt}$ is a constant. But this is precisely Kepler's Second Law. So our little formula leads to the validation of a major historical advance in astronomy.

Example 6 There is a three dimensional analogue of the cross product formula for parallelogram area. A parallelepiped is a six-sided figure in three dimensions bounded by opposite pairs of congruent parallelograms. Imagine a rectangular solid like a shoebox that has been subjected to shear in two directions. Any corner of the figure has three vectors emanating from it. Two define a parallelogram base and the third extends out of that plane to give the figure volume. Let the vectors forming the base be $\mathbf{B}$ and $\mathbf{C}$. The area of the base is then $|\mathbf{B} \times \mathbf{C}|$. But before we take the norm, note that the vector $\mathbf{B} \times \mathbf{C}$ is perpendicular to the parallelogram base and that is exactly the direction in which the altitude of the figure is measured. In fact, the altitude is the projection of the out-of-plane third vector $\mathbf{A}$ onto $\mathbf{B} \times \mathbf{C}$. The scalar projection of $\mathbf{A}$ onto $\mathbf{B} \times \mathbf{C}$ is precisely $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C}$. Since the volume V of the parallelepiped is the area of its base times its altitude, we have $V = |\mathbf{A} \cdot \mathbf{B} \times \mathbf{C}|$. The vertical bars here signify absolute value.

$\mathbf{A} \cdot \mathbf{B} \times \mathbf{C}$ is called the **scalar triple product** of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. It is an unambiguous expression as the cross product must be executed first. Computationally, if $\mathbf{A} = \langle a_1, a_2, a_3\rangle$, $\mathbf{B} = \langle b_1, b_2, b_3\rangle$, and $\mathbf{C} = \langle c_1, c_2, c_3\rangle$, we have $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1)$. When we discuss determinants, we will present a scheme that self-organizes this calculation. Interestingly, any even permutation of the vectors while holding the operation symbols in place does not change the value of the scalar triple product. So $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = \mathbf{B} \cdot \mathbf{C} \times \mathbf{A} = \mathbf{C} \cdot \mathbf{A} \times \mathbf{B}$. An odd permutation changes the sign.

Example 7 *A three-dimensional lattice is described by a repeating unit in the* *shape of a parallelepiped with corner vectors* $\langle 0,1,0\rangle$*,* $\langle -2,2,0\rangle$*, and* $\langle -1,2,3\rangle$*. Units are in millimeters. How many lattice units would be in a cubic centimeter?*

Solution: $V = |\langle 0,1,0\rangle \cdot \langle -2,2,0\rangle \times \langle -1,2,3\rangle| = 6\times 10^{-9}\ m^3$, so $\dfrac{1\times 10^{-6}}{6\times 10^{-9}} = 1.67\times 10^2$ units.

Example 8 It makes sense to form $\mathbf{A}\times\mathbf{B}\times\mathbf{C}$, but as written the expression is ambiguous because the cross product is not associative. In other words, $\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) \neq (\mathbf{A}\times\mathbf{B})\times\mathbf{C}$. We can develop a convenient expression for $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ two different ways. The first involves grinding out all the components by brute calculation and then seeking simplifications. This will work but it is not as pleasant a derivation as one based on our bookkeeping symbols δ_{ij} and ε_{ijk}. We need two things...a fact and a notational convention.

(i) The bookkeeping symbols satisfy the following identity: $\varepsilon_{kij}\varepsilon_{kpq} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$, as may be established by checking the non-vanishing terms (almost all do).

(ii) Let us use the so-called Einstein summation convention, where if two indices are repeated in an expression there is an implied summation. The range of the sum is clear from context. For example, we would interpret the expression $a_{ij}b_{jk}$ as $\sum_{j=1}^{3} a_{ij}b_{jk}$ where $1 \le j \le 3$. Another example is the dot product definition in $\mathbb{R}^3$: $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j \delta_{ij}$ reduces to just $x_i y_j \delta_{ij}$.

So, with the vector components as in Example 6,
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \varepsilon_{ijk}\hat{\mathbf{e}}_i a_j(\mathbf{B}\times\mathbf{C})_k = \varepsilon_{ijk}\hat{\mathbf{e}}_i a_j(\varepsilon_{kpq}b_p c_q) = \varepsilon_{ijk}\varepsilon_{kpq}\hat{\mathbf{e}}_i a_j b_p c_q$. Now $\varepsilon_{ijk} = \varepsilon_{kij}$, and $\varepsilon_{kij}\varepsilon_{kpq} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$, so
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})\hat{\mathbf{e}}_i a_j b_p c_q = \hat{\mathbf{e}}_i a_j b_i c_j - \hat{\mathbf{e}}_i a_j b_j c_i = \mathbf{B}(a_j c_j) - \mathbf{C}(a_j b_j) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}$
That's it. The beauty of the bookkeeping symbols is that they act like scalars and can be shifted arouund in expressions without worrying about commutativity. In summary, $\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})$. This is universally called the **BAC-CAB** identity for the **vector triple product**. It is useful in characterizing certain electromagnetic phenomena, but the physics gets pretty deep, so we will just illustrate an identity due to Jacobi that comes up in non-associative (like the cross product) algebras.

Example 9 *With* $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$ *as above, show that*
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) + \mathbf{B}\times(\mathbf{C}\times\mathbf{A}) + \mathbf{C}\times(\mathbf{A}\times\mathbf{B}) = \mathbf{0}.$

Solution:
$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})$$
$$\mathbf{B}\times(\mathbf{C}\times\mathbf{A}) = \mathbf{C}(\mathbf{B}\cdot\mathbf{A}) - \mathbf{A}(\mathbf{B}\cdot\mathbf{C})$$
$$\mathbf{C}\times(\mathbf{A}\times\mathbf{B}) = \mathbf{A}(\mathbf{C}\cdot\mathbf{B}) - \mathbf{B}(\mathbf{C}\cdot\mathbf{A})$$

Adding the six terms on the right and cancelling in pairs (recall the dot product is commutative), we obtain the zero vector.

4.7 VECTOR GEOMETRY

We can harness the tools developed thus far to solve additional two- and three-dimensional geometric problems involving space curves.

4.71 STRAIGHT LINES

Restricting our discussion to $\mathbb{R}^2$ or the xy-plane for the moment, we ask what is the bare minimum information required to specify a straight line graph. We need either two points (according to Euclid) or a point along with a direction. Given two points (x_1,y_1) and (x_2,y_2) we can determine the relative inclination (change in y divided by change in x) of the line running through them by calculating $\frac{y_2-y_1}{x_2-x_1} = m$. Reasoning that this should be the same as the inclination using an arbitrary point (x,y) on the line, we have $\frac{y-y_1}{x-x_1} = m$ or $y = mx + (y_1 - mx_1)$. This is the **point-slope** form of the straight line equation. Calling m the **slope** and $b = y_1 - mx_1$ the **y-intercept**, we have recovered the formula $y = mx + b$, which is the **slope-intercept** form of the straight line equation. Given a generic point (x,y) on a line and a direction specified as the slope m, we can easily calculate $b = y - mx$ and reconstruct $y = mx + b$ for that case. In the event that a line is vertical ($m = \infty$) and cuts the x-axis at $x = c$, that expression will serve as the equation of the line. We now have a complete picture of graphing straight lines in $\mathbb{R}^2$. What about $\mathbb{R}^3$? Suddenly, the planar experience in $\mathbb{R}^2$ seems inadequate.

We will develop an analogue of the slope-intercept formula for three dimensions. Once again, we start off with two points (a_1,a_2,a_3) and (b_1,b_2,b_3). How shall we define the slope of the line through these points? There are three reference planes to choose from, the xy-plane, xz-plane, and yz-plane. A better idea is to let the **direction vector** $\mathbf{v} = \langle b_1-a_1, b_2-a_2, b_3-a_3\rangle$ express the orientation of the line in space. This is effectively the "slope" in $\mathbb{R}^3$. Now we can write the **equation of a line** as $\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}t$, where $\mathbf{r}(t)$ is the position vector of a point on the line, $\mathbf{r}_0$ is the position vector of a fixed point on the line, $\mathbf{v}$ is the direction vector, and t is a **parameter**, typically ($-\infty < t < \infty$). Thinking of t as time does no harm and will make practical sense when we view the line as the path of a moving object and we need to compute velocity and acceleration. You can imagine starting at the origin $(0,0,0)$ and moving along the vector $\mathbf{r}_0 = \langle a_1,a_2,a_3\rangle$ to get to the line (the bus stop), then moving along the line (riding the bus) by taking various scalar multiples of $\mathbf{v}$. In effect you are adding the constant vector $\mathbf{r}_0$ to a multiple t of the direction vector $\mathbf{v}$ to generate the vector $\mathbf{r}(t)$, which is the position vector of a generic point on the line.

Example 1 *Write the vector equation for the straight line through the point* $(1,2,3)$ *in the direction of the vector* $\langle -1,1,4\rangle$*. Does the line go through* $(-2,5,15)$ *and if so what is t?*

Solution: $\mathbf{r}_0 = \langle 1,2,3\rangle$ and $\mathbf{v} = \langle -1,1,4\rangle$, so $\mathbf{r}(t) = \langle 1,2,3\rangle + t\langle -1,1,4\rangle$. Note $\langle 1,2,3\rangle + 3\langle -1,1,4\rangle = \langle -2,5,15\rangle$, so "yes" when $t = 3$.

Example 2 *Find the equation of the straight line that goes through* $(2,1,4)$ *and* $(7,-7,3)$.

Solution: $\mathbf{v} = \langle 7-2, -7-1, 3-4\rangle = \langle 5,-8,-1\rangle$. Use either given point for $\mathbf{r}_0$, say $\mathbf{r}_0 = \langle 2,1,4\rangle$, then $\mathbf{r}(t) = \langle 2,1,4\rangle + t\langle 5,-8,-1\rangle$. Note that the order of subtraction in calculating $\mathbf{v}$ from the two given points could be reversed. Then reversing the sign of t generates the same line in the same direction.

Example 3 *Express the vector equation in Example 2 in component form.*
Solution: Separating the components of $\langle x(t), y(t), z(t)\rangle$ we get $x(t) = 2 + 5t$, $y(t) = 1 - 8t$, and $z(t) = 4 - t$.

Example 4 *Find the equation of a line parallel to the line in Example 1 but passing through* $(1,1,1)$.
Solution: We use the same direction vector as in Example 1 but we adjust $\mathbf{r}_0$ to $\langle 1,1,1\rangle$. So $\mathbf{r}(t) = \langle 1,1,1\rangle + t\langle -1,1,4\rangle$.

Example 5 *Do the lines* $r_1(t) = \langle 1,3,5\rangle + t_1\langle 4,-5,3\rangle$ *and* $r_2(t) = \langle 6,2,2\rangle + t_2\langle 4,-5,3\rangle$ *intersect?*
Solution: The direction vector for both lines is the same, so they are parallel. If they intersect they would be the same line, so if $t_1 = 0$ there must be some t_2 such that $\langle 6,2,2\rangle + t_2\langle 4,-5,3\rangle = \langle 1,3,5\rangle$. Checking the first two component equations, (i) $6 + 4t_2 = 1$ and (ii) $2 - 5t_2 = 3$. Equation (i) gives $t_2 = -\frac{5}{4}$ and (ii) gives $t_2 = -\frac{1}{5}$. Evidently there is no consistent parameter value that would allow $\mathbf{r}_2(t)$ to equal $\langle 1,3,5\rangle$, so the lines can't be the same, hence they do not intersect.

Example 6 *Find the equation of a line through* $(1,2,3)$ *perpendicular to* $r(t) = \langle 1,2,3\rangle + t\langle 2,4,3\rangle$.
Solution: Suppose the direction vector of the perpendicular line is $\langle x,y,z\rangle$, then the dot product of the two direction vectors should be zero, so $\langle 2,4,3\rangle \cdot \langle x,y,z\rangle = 0$. All we need is a solution to $2x + 4y + 3z = 0$. We have two degrees of freedom to work with, so we can easily find $z = 0$, $x = 2$, and then $y = -1$ as a solution. Then $\mathbf{r}_1(t) = \langle 1,2,3\rangle + t\langle 2,-1,0\rangle$ satisfies the criteria. Recall that in $\mathbb{R}^2$ a line that is perpendicular to that given by $y = mx + b$ is $y = m'x + b'$, where $mm' = -1$. Restating these formulas in our vector notation, $y = mx + b$ becomes $\mathbf{r}(t) = \langle 0,b\rangle + x\langle 1,m\rangle$ and $\mathbf{r}'(t) = \langle 0,b'\rangle + x\langle 1,m'\rangle$. Then the dot product condition yields $\langle 1,m\rangle \cdot \langle 1,m'\rangle = mm' + 1 = 0$ which returns the perpendicular slope formula $mm' = -1$.

A **point-to-line** distance is easily calculated using our vector methods. Suppose we are given the line $\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}t$ and the point $Q = (a,b,c)$. We can pick a completely arbitrary point P on the line and construct the vector $\mathbf{w}$ from P to Q. Note that a unit vector parallel to the line is $\frac{\mathbf{v}}{|\mathbf{v}|}$. Then the scalar projection of $\mathbf{w}$ in the direction perpendicular to the line is $|\mathbf{w}|\sin\theta$, where θ is the angle between $\mathbf{w}$ and $\mathbf{v}$. Using our cross product formula we see that the perpendicular point-to-line distance $d = |\mathbf{w}|\sin\theta = \frac{|\mathbf{w} \times \mathbf{v}|}{|\mathbf{v}|}$. Some examples will make this clear.

Example 7 What is the distance between the point $(4,3,7)$ and the line $\mathbf{r}(t) = \langle 1,2,3\rangle + \langle 2,-3,1\rangle t$?

Solution: We first pick the point P on the line. The simplest case usually is to let $t = 0$, so $P = (1,2,3)$. The vector $\mathbf{w}$ from P to $Q = (4,3,7)$ is then $\langle 3,1,4\rangle$. Finding $\frac{\mathbf{v}}{|\mathbf{v}|}$ we have $\frac{\langle 2,-3,1\rangle}{\sqrt{14}}$. Finally, using our cross product formula, we get $\langle 3,1,4\rangle \times \frac{\langle 2,-3,1\rangle}{\sqrt{14}} = \frac{\langle 13,5,-11\rangle}{\sqrt{14}}$, so the distance d is $\frac{1}{\sqrt{14}}|\langle 13,5,-11\rangle| = 4.74$.

Example 8 What is the distance from the point $(-1,5,9)$ to the line $\mathbf{r}(t) = \langle -2,-2,11\rangle + \langle 1,-5,8\rangle t$?

Solution: With $t = 0$, $P = (-2,-2,11)$ and $Q = (-1,5,9)$. Then $\mathbf{w} = \langle 1,7,-2\rangle$ and $\frac{\mathbf{v}}{|\mathbf{v}|} = \frac{\langle 1,-5,8\rangle}{3\sqrt{10}}$. The cross product is $\langle 1,7,-2\rangle \times \frac{\langle 1,-5,8\rangle}{3\sqrt{10}} = \frac{|\langle 46,-10,-12\rangle|}{3\sqrt{10}}$. Then $d = \left|\frac{48.58}{9.49}\right| = 5.12$.

There are three cases to consider in connection with **line-to-line** distances. Two lines can intersect, they can be parallel, or they can be neither. Lines that are not parallel and do not intersect are called **skew** lines. Lines that intersect obviously have zero distance between them. For lines that are parallel, the distance between them can be found by picking an arbitrary point on one line and finding its distance to the other line as we did above. Skew lines offer more of a challenge. The distance between skew lines, of course, means the distance perpendicular to both. Our plan is to create a vector $\mathbf{w}$ by connecting two arbitrary points, one on each line. Then,.by the magic of vector methods,.find a unit vector $\mathbf{v}$ that is simultaneously perpendicular to both lines. The scalar projection of $\mathbf{w}$ in the direction of $\mathbf{v}$ will be the required perpendicular distance d. So given $\mathbf{r}(t_1) = \mathbf{r}_0 + \boldsymbol{\alpha} t_1$ and $\mathbf{s}(t_2) = \mathbf{s}_0 + \boldsymbol{\beta} t_2$ we form the unit vector $\mathbf{v} = \boldsymbol{\alpha} \times \boldsymbol{\beta}$ and then unitize it. This unit vector is necessarily perpendicular to both lines and will serve as the "screen" upon which we project the vector $\mathbf{w}$. Now suppose the points on the two lines are $P = (a,b,c)$ and $Q = (d,e,f)$, so $\mathbf{w} = \langle a-d, b-e, c-f\rangle$. The distance between skew lines $d = \left|\frac{\mathbf{w} \cdot \boldsymbol{\alpha} \times \boldsymbol{\beta}}{\|\boldsymbol{\alpha} \times \boldsymbol{\beta}\|}\right|$. Here we have used the convention that if norms and absolute values appear in the same expression, norms are denoted by $\|\cdot\|$.

Example 9 What is the distance between the lines given by $\mathbf{r}(t_1) = \langle 1,2,3\rangle + \langle 1,0,3\rangle t_1$ and $\mathbf{s}(t_2) = \langle -1,5,7\rangle + \langle 4,4,6\rangle t_2$?

Solution: First, $\mathbf{v} = \langle 1,0,3\rangle \times \langle 4,4,6\rangle = \langle -12,6,4\rangle$. Next, set $t_1 = t_2 = 0$ so that $P = (1,2,3)$ and $Q = (-1,5,7)$. Then $\mathbf{w} = \langle 2,-3,-4\rangle$. Finally,

$d = \left| \frac{\langle 2,-3,-4\rangle \cdot \langle 1,0,3\rangle}{\|\langle 1,0,3\rangle\|} \right| = \left| \frac{-10}{\sqrt{10}} \right| = 3.162$. We wanted to be precise in representing the two line equations so we distinguished the two parameters. This isn't strictly necessary since any value in either line equation still yields a point on the line (provided $-\infty < t < \infty$)

4.72 GENERAL CURVES (∫)

We can investigate more general space curves than just straight lines by considering the **curve** or **path** function $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$, where t is a parameter and $x(t)$, $y(t)$, and $z(t)$ are component functions. It is customary to think of $t \in [t_1, t_2]$ as time and then $\mathbf{r}(t)$ as the position vector of a point in space which we imagine has some velocity as it traces out the curve. There are several restrictions we impose on $\mathbf{r}(t)$ to ensure it is mathematically well-behaved. First, we insist that $\mathbf{r}(t)$ must be a continuous function so the curve is never broken. Second, we require $\mathbf{r}(t)$ to be an injective function on (t_1, t_2), so the curve never crosses itself. We allow $\mathbf{r}(t_1) = \mathbf{r}(t_2)$ in the case of a **closed curve**. Third, we impose the condition that $\mathbf{r}'(t) \neq 0$, which ensures that the tracing process never pauses. It could never reverse itself either, since this would violate injectivity. Fourth and finally, we want $\mathbf{r}(t)$ to be a **smooth** function, which means $\mathbf{r}'(t)$ should be continuous. This is another guarantee that tracing could not reverse itself, since there would be a time t_0 when $\mathbf{r}'(t_0) = 0$. Many of the space curves encountered in real world settings also have continuous higher order path functions (orbits, trajectories, and so forth) and this allows us to characterize them more completely.

Example 1 *Find the three dimensional path function corresponding to a circle of radius* 2 *centered at* $(0,0,1)$ *in the* z*-plane*

Solution: $\mathbf{r}(t) = \langle 2\cos t, 2\sin t, 1\rangle$ where $0 \le t \le 2\pi$.

Example 2 *Find the path function in the* xy*-plane corresponding to an ellipse with equation* $\frac{x^2}{9} + \frac{y^2}{25} = 1$.

Solution: Let $x = a\cos t$ and $y = b\sin t$. Then $\frac{a^2\cos^2 t}{9} + \frac{b^2\sin^2 t}{25} = 1$. Regrouping, $\left(\frac{a}{3}\right)^2\cos^2 t + \left(\frac{b}{5}\right)^2\sin^2 t = 1$. We see $\left(\frac{a}{3}\right)^2 = \left(\frac{b}{5}\right)^2 = 1$, so $a = \pm 3$ and $b = \pm 5$. Choosing the positive values, we can finally write $\mathbf{r}(t) = \langle 3\cos t, 5\sin t\rangle$.

Example 3 *Find the path function for a circular helix of radius* 3 *and pitch* 1 *with axis coincident with the positive* z*-axis and extending from* $z = 0$ to $z = 10$.

Solution: Looking down at the helix along the z-axis from a height above $z = 10$ we would see a circle of radius 3. Following Example 1 the path function would be $\mathbf{r}(t) = \langle 3\cos t, 3\sin t, z\rangle$, where we have initialized the curve at the point $(3,0,0)$ and the dependence in the z direction still needs to be determined. Since the pitch is one unit, whenever the parameter t increases by 2π units, the position vector has the same x and y coordinates, but the z coordinate has increased by exactly one. We can make this happen by setting $z = \frac{t}{2\pi}$. Now the complete path function is $\mathbf{r}(t) = \left\langle 3\cos t, 3\sin t, \frac{t}{2\pi} \right\rangle$ and we allow t to range over the interval $[0, 20\pi]$ to ensure that the helix completes ten revolutions and terminates at $(3, 0, 10)$.

We can view a space curve as a completed entity where the tracing process implicit in the definition of the path function is considered to be a fait accompli as in the prior three examples. However, if the path function describes some dynamic process like an airplane flying or a satellite orbiting, we might be more interested in the local rather than global properties of the curve. Given $\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle$ we can easily get the **velocity** vector along the curve by calculating $\mathbf{r}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}$, which is the usual definition of derivative for a function of one variable. Applying this to $\mathbf{r}(t)$ written in terms of components, we can easily confirm that $\mathbf{r}'(t) = \langle x'(t), y'(t), z'(t) \rangle$. Note that $|\mathbf{r}'(t)|$ is the scalar **speed**. This vector is never zero ($\mathbf{r}'(t) \neq 0$ was assumed) and represents the direction vector of a straight line tangent to the curve at a given point, much like the derivative $y'(x)$ represents the slope of a tangent to the graph of $y(x)$. If we want only the direction of the tangent, we can unitize $\mathbf{r}'(t)$ and define $\frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \mathbf{T}(t)$ as the **unit tangent vector**. As the curve wends its way through space, the unit tangent vector will generally be changing its direction, unless we are on a piece of "straight track", so to speak. Assuming we can differentiate $\mathbf{r}(t)$ one more time, the derivative $\mathbf{T}'(t)$ measures this rate of change. There is no reason to suppose that the raw derivative $\mathbf{T}'(t)$ is a unit vector, so we unitize that and define $\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}$. $\mathbf{N}(t)$ is called the **unit** (**principal**) **normal vector** for the curve $\mathbf{r}(t)$. It is perpendicular (normal) to $\mathbf{T}(t)$ and points in the direction that the unit tangent vector is turning towards. It will be convenient to have a table showing the interplay of differentiation and vector operations.

Table 7 - Differentiation of Vectors for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$

1) Constant multiple rule	$(c\mathbf{x})' = c(\mathbf{x}')$
2) Sum rule	$(\mathbf{x} + \mathbf{y})' = \mathbf{x}' + \mathbf{y}'$
3) Difference rule	$(\mathbf{x} - \mathbf{y})' = \mathbf{x}' - \mathbf{y}'$
4) Dot product rule	$(\mathbf{x} \cdot \mathbf{y})' = \mathbf{x}' \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{y}'$
5) Cross product rule ($k = 3$)	$(\mathbf{x} \times \mathbf{y})' = \mathbf{x}' \times \mathbf{y} + \mathbf{x} \times \mathbf{y}'$ (maintain order!)

Example 4 *Show why $\mathbf{T}(t)$ and $\mathbf{N}(t)$ are perpendicular.*

Solution: This is a consequence of the more general fact that if a vector $\mathbf{v}(t)$ has constant length, then $\mathbf{v}'(t)$ is perpendicular to $\mathbf{v}(t)$. If $\mathbf{v}(t)$ has constant length, then $(\mathbf{v}(t) \cdot \mathbf{v}(t))' = \left[|\mathbf{v}(t)|^2\right]' = 0$. Then by the dot product rule, $\mathbf{v}'(t) \cdot \mathbf{v}(t) + \mathbf{v}(t) \cdot \mathbf{v}(t)' = 2(\mathbf{v}(t) \cdot \mathbf{v}(t)') = 0$. Evidently $\mathbf{v}'(t) \perp \mathbf{v}(t)$. Now $\mathbf{T}(t)$ is a unit vector so it has constant length one. Then $\mathbf{T}(t) \cdot \mathbf{T}'(t) = 0$ by the preceding and therefore $\mathbf{T}(t) \perp \mathbf{T}'(t)$. A constant scalar multiple of one of these vectors does not influence perpendicularity, so $\mathbf{T}(t) \perp \dfrac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}$ or $\mathbf{T}(t) \perp \mathbf{N}(t)$.

Now that we have $\mathbf{T}(t)$ and $\mathbf{N}(t)$, we can define $\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t)$. Note $\mathbf{T}(t) \perp \mathbf{N}(t)$ means $|\mathbf{B}(t)| = |\mathbf{T}(t)||\mathbf{N}(t)|\sin\frac{\pi}{2} = 1$, so we are justified in calling $\mathbf{B}(t)$ the **unit binormal vector**. $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ are mutually perpendicular and are attached to each point of a space curve as a miniature mobile version of the basis vectors for $\mathbb{R}^3$. Together they are called the "moving trihedron" and provide a localized cartesian coordinate system to investigate motion along the path. In pairs they define three orthogonal planes: $\mathbf{T}$ and $\mathbf{N}$ determine the **osculating plane**, $\mathbf{T}$ and $\mathbf{B}$ determine the **rectifying plane**, and $\mathbf{N}$ and $\mathbf{B}$ together determine the **normal plane**. A good intuitive picture of what $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ represent is this: imagine an airplane flying along the space curve with these three axes emanating from the center of the plane. A climb or dive would correspond to rotation around $\mathbf{N}$. A roll to either side would corespond to rotation around $\mathbf{T}$. And a yaw (turn without banking) would correspond to rotation around $\mathbf{B}$.

Two scalar quantities are also useful in characterizing the local behavior of space curves. The first is **curvature**, denoted by $\kappa(t)$. We discuss the second below. Curvature is, roughly speaking, the ratio of the tendency of the curve to change direction to the velocity along the curve. By definition, $\kappa(t) = \dfrac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|}$. Intuitively, if the change in direction is small compared to velocity, the curvature of the path is small. Conversely, if the direction change is significant for a small velocity, the curvature is large. The **radius of curvature** $\rho(t) = \dfrac{1}{\kappa(t)}$. This is the imputed radius of a circle having the same curvature as $\mathbf{r}(t)$. This circle lies in the osculating plane and is called the **osculating circle**. It represents the circle that most nearly conforms to the curve at some momentary t. If $\mathbf{r}(t)$ is not changing direction at a point, then $\kappa(t) = 0$ and $\rho(t) = \infty$. Curvature can be expressed directly in terms of $\mathbf{r}(t)$ as $\kappa = \dfrac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}$.

Example 5 *Find $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ and κ for the circle in the xy-plane given by $\mathbf{r}(t) = \langle a\cos t, a\sin t\rangle$ for $0 \le t \le 2\pi$.*

Solution: First we need $\mathbf{r}'(t) = \langle -a\sin t, a\cos t\rangle$, so

$\mathbf{T}(t) = \dfrac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \dfrac{\langle -a\sin t, a\cos t\rangle}{\sqrt{a^2\cos^2 t + a^2\sin^2 t}} = \langle -\sin t, \cos t\rangle$. Then

$\mathbf{N}(t) = \dfrac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|} = \dfrac{\langle -\cos t, -\sin t\rangle}{\sqrt{\cos^2 t + \sin^2 t}} = \langle -\cos t, -\sin t\rangle$. Recall from §4.6 that the cross product is a three dimensional construction, but we can apply it to two dimensional vectors by embedding them in $\mathbb{R}^3$ each with a zero component in the z direction. The cross product formula then reduces to $\mathbf{x} \times \mathbf{y} = \widehat{\mathbf{k}}(x_1y_2 - x_2y_1)$. It follows that
$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t) = \langle -\sin t, \cos t, 0\rangle \times \langle -\cos t, -\sin t, 0\rangle = \widehat{\mathbf{k}}[(-\sin t)(-\sin t) - (-\cos t)(\cos t)] = \widehat{\mathbf{k}}$.
The alternate formula $\mathbf{x} \times \mathbf{y} = |\mathbf{x}||\mathbf{y}|\sin\theta\,\widehat{\mathbf{e}}_\perp$ (§4.6 Example 1) yields the same result with $\widehat{\mathbf{e}}_\perp$ given by the right-hand rule as $\widehat{\mathbf{k}}$. Now for the curvature: $\kappa(t) = \dfrac{|\langle -\cos t, -\sin t\rangle|}{|\langle -a\sin t, a\cos t\rangle|} = \dfrac{1}{a}$. Since $\rho(t) = \dfrac{1}{\kappa(t)}$, the radius of curvature is the radius of the original circle a. In fact, the osculating circle is the given circle itself and the osculating plane is the xy-plane. Although $\mathbf{T}(t)$ and $\mathbf{N}(t)$ vary with t, $\mathbf{B}(t) = \widehat{\mathbf{k}}$ is constant.

Example 6 *Find* $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ *and* κ *for the circular helix in Example 3 above.*

We have $\mathbf{r}(t) = \left\langle 3\cos t, 3\sin t, \dfrac{t}{2\pi}\right\rangle$ for the helix, where $0 \le t \le 20\pi$. Then
$\mathbf{r}'(t) = \left\langle -3\sin t, 3\cos t, \dfrac{1}{2\pi}\right\rangle$ and $|\mathbf{r}'(t)| = \sqrt{9 + \dfrac{1}{4\pi^2}} = \dfrac{\sqrt{36\pi^2+1}}{2\pi}$. So
$\mathbf{T}(t) = \dfrac{2\pi}{\sqrt{36\pi^2+1}}\left\langle -3\sin t, 3\cos t, \dfrac{1}{2\pi}\right\rangle$. Let us set $\alpha = \dfrac{2\pi}{\sqrt{36\pi^2+1}}$ to simplify notation. Next,
$\mathbf{N}(t) = \dfrac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|} = \dfrac{\alpha\langle -3\cos t, -3\sin t, 0\rangle}{\alpha\sqrt{9\cos^2 t + 9\sin^2 t}} = \langle -\cos t, -\sin t, 0\rangle$. So $\mathbf{N}(t)$ is antiparallel to the projection of $\mathbf{r}(t)$ onto the xy-plane. In other words, as the helix winds up the z axis, $\mathbf{N}(t)$ constantly points to that axis. Getting the unit binormal won't be as easy as in Example 5.
$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t) = \alpha\left\langle -3\sin t, 3\cos t, \dfrac{1}{2\pi}\right\rangle \times \langle -\cos t, -\sin t, 0\rangle$. This boils down to
$\mathbf{B}(t) = \alpha\left\langle \dfrac{\sin t}{2\pi}, \dfrac{-\cos t}{2\pi}, 3\right\rangle$. As a check we compute
$|\mathbf{B}(t)| = \dfrac{2\pi}{\sqrt{36\pi^2+1}}\left|\left\langle \dfrac{\sin t}{2\pi}, \dfrac{-\cos t}{2\pi}, 3\right\rangle\right| = \dfrac{2\pi}{\sqrt{36\pi^2+1}} \cdot \dfrac{\sqrt{36\pi^2+1}}{2\pi} = 1$. Finally,
$\kappa(t) = \dfrac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|} = \dfrac{3\alpha}{1/\alpha} = 3\alpha^2 = 3 \cdot \dfrac{4\pi^2}{36\pi^2+1} = \dfrac{12\pi^2}{36\pi^2+1}$. Note that $\dfrac{12\pi^2}{36\pi^2+1} < \dfrac{12\pi^2}{36\pi^2} = \dfrac{1}{3}$.
By Example 5 a circle of radius 3 would have curvature $\dfrac{1}{3}$. The helix goes around the z axis at a constant distance of 3 units, so in that respect it behaves like the circle, but it also climbs 1 unit every time it goes around. The extra distance that the helical path function traverses compared to the circular path function with the same nominal diameter "opens up" the helix and causes its curvature to be less than that for the circle. Also note that the **TNB** frame tracking the helix has a constant backwards tilt in order to follow $\mathbf{T}$ along the curve.

Example 7 The curve $\mathbf{r}(t) = \langle t, t^2, t^3\rangle$, known as the "twisted cubic", is often used as an example illustrating the features of space curve theory. Find $\mathbf{T}, \mathbf{N},$ and $\mathbf{B}$ when $t = 1$.

For the unit vectors: $\mathbf{T}(t) = \dfrac{\langle 1, 2t, 3t^2\rangle}{\sqrt{1 + 4t^2 + 9t^4}}$ in general, and $\mathbf{T}(1) = \dfrac{\langle 1, 2, 3\rangle}{\sqrt{14}}$.

Now for $\mathbf{N}$:

$\frac{d}{dt}\mathbf{T}(t) = \frac{d}{dt}\left(\frac{1}{\sqrt{1+4t^2+9t^4}}\right)\hat{\mathbf{i}} + \frac{d}{dt}\left(\frac{2t}{\sqrt{1+4t^2+9t^4}}\right)\hat{\mathbf{j}} + \frac{d}{dt}\left(\frac{3t^2}{\sqrt{1+4t^2+9t^4}}\right)\hat{\mathbf{k}}$ where

$\frac{d}{dt}\left(\frac{1}{\sqrt{1+4t^2+9t^4}}\right) = -\frac{1}{2}\left(\frac{36t^3+8t}{(1+4t^2+9t^4)^{3/2}}\right)$,

$\frac{d}{dt}\left(\frac{2t}{\sqrt{1+4t^2+9t^4}}\right) = \frac{2}{\sqrt{1+4t^2+9t^4}} - \frac{t(36t^3+8t)}{(1+4t^2+9t^4)^{3/2}}$, and

$\frac{d}{dt}\left(\frac{3t^2}{\sqrt{1+4t^2+9t^4}}\right) = \frac{6t}{\sqrt{1+4t^2+9t^4}} - \frac{3}{2}\left(\frac{t^2(36t^3+8t)}{(1+4t^2+9t^4)^{3/2}}\right)$. With these substitutions, this is $\mathbf{T}'(t)$. Since $\mathbf{N}(1) = \frac{\mathbf{T}'(1)}{|\mathbf{T}'(1)|}$, we have $\mathbf{N}(1) = \frac{\frac{\sqrt{14}}{98}\langle 11,8,-9\rangle}{\frac{\sqrt{14}}{98}\sqrt{121+64+81}} = \frac{1}{\sqrt{266}}\langle 11,8,-9\rangle$, which is the unit principal normal at $t = 1$.

Then $\mathbf{B}(1) = \mathbf{T}(1) \times \mathbf{N}(1) = \frac{1}{61.02}\langle -42,42,-14\rangle = 0.2294\langle -3,3,-1\rangle$., which is the unit binormal at $t = 1$. Lots of calculation for a seemingly simple curve.

The second scalar quantity that gives insight into the local behavior of space curves is **torsion**, denoted by $\tau(t)$. With non-planar curves, the osculating plane can rock back and forth as the curve is traced out. This means the direction of the unit normal (and the other two unit vectors, of course) may be changing. The tendency of the unit normal to change direction along the curve, measured by its projection onto the unit binormal, is called the torsion of the curve. We have thus far developed all of our space curve formulas in terms of $\mathbf{r}(t)$ and its derivatives. The corresponding expression for torsion is unfortunately rather complicated, but for the sake of completeness we present it here: $\tau(t) = \frac{[\mathbf{r}'(t) \times \mathbf{r}''(t)] \cdot \mathbf{r}'''(t)}{|\mathbf{r}'(t) \times \mathbf{r}''(t)|^2}$. We are now going to explore a different but equivalent way of characterizing space curves that has a more "natural" feel to it.

Up to this point we have expressed space curves in terms of their intuitively appealing but somewhat artificial dynamic behavior with respect to the parameter time. Many applications in physics and engineering involve objects large and small following paths in space over time, so this viewpoint is certainly convenient and it makes sense to develop the theory around this parameter. In the mid nineteenth century mathematicians wished to develop a more intrinsic way of parametrizing curves that did not depend (explicitly at least) on the notion of time. The French mathematician J.-F. Frenet proposed using the distance along a curve from some fixed starting point as a more natural parameter (the Frenet coordinate). This allowed space curves to be expressed in terms of a parameter that was already contained, so to speak, in the curve itself, viewed as a static and completed geometric object. The first order of business in this approach is to find an expression that yields the length of a curve.

Let $s(t)$ be the length along the curve given by $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ where we start measuring length at t_0 and stop at t_1. The goal is to get an expression for $\frac{ds}{dt}$ in terms of what we know and then form the integral $s(t_1) - s(t_0) = \int_{t_0}^{t_1} ds = \int_{t_0}^{t_1} \frac{ds}{dt} dt$. Differential distance ds in $\mathbb{R}^3$ obeys the Pythagorean theorem: $ds^2 = dx^2 + dy^2 + dz^2$, so we can write $\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$. Note that the velocity vector along the curve is $\mathbf{r}'(t) = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle$. Evidently $\mathbf{r}'(t) \cdot \mathbf{r}'(t) = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2$, so $\sqrt{\mathbf{r}'(t) \cdot \mathbf{r}'(t)} = |\mathbf{r}'(t)| = \frac{ds}{dt}$. Plugging this into our integral we have $s(t_1) - s(t_0) = \int_{t_0}^{t_1} |\mathbf{r}'(t)| dt$. Setting $t_0 = 0$ and $s(t_0) = 0$, we can re-express the integral for arc length as a function of time as $s(t) = \int_0^t |\mathbf{r}'(\tau)| d\tau$. We have used τ as the dummy integration variable time.

Example 8 *Find the total length L along the helix of Example 3 above. The curve is given by* $\mathbf{r}(t) = \left\langle 3\cos t, 3\sin t, \frac{t}{2\pi} \right\rangle$ *and t ranges from 0 to 20π.*

Solution: $L = s(20\pi) = \int_0^{20\pi} |\mathbf{r}'(t)| dt$. We know $\mathbf{r}'(t) = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle = \left\langle -3\sin t, 3\cos t, \frac{1}{2\pi} \right\rangle$, so $|\mathbf{r}'(t)| = \sqrt{9\sin^2 + 9\cos^2 + \frac{1}{4\pi^2}} = \sqrt{9 + \frac{1}{4\pi^2}} \approx 3.0042$. Then $L = (3.0042)(20\pi) = 188.76$ units.

Example 9 *Find the length L of the "twisted cubic"* $\mathbf{r}(t) = \langle t, t^2, t^3 \rangle$ *in Example 7 from $t = 0$ to $t = 1$.*

Solution: $\mathbf{r}'(t) = \langle 1, 2t, 3t^2 \rangle$, so $|\mathbf{r}'(t)| = \sqrt{1 + 4t^2 + 9t^4}$. Then $L = s(1) = \int_0^1 \sqrt{1 + 4t^2 + 9t^4}\, dt \approx 1.863$ (by numerical integration)

With our formula for the arc length $s(t)$ we are in a position to restate the formulas for $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ and κ relative to the Frenet coordinate s. Applying the Fundamental Theorem of Calculus to the arc length formula $s(t) = \int_0^t |\mathbf{r}'(\tau)| d\tau$ we see that $s(t)$ is a strictly increasing function, since the norm $|\mathbf{r}'(t)| > 0$ (curve tracing never stops). It follows that $s(t)$ is injective and therefore invertible (at least in principle) as $t(s)$. We can also compose $\mathbf{r}(t)$ with $t(s)$ to form $\mathbf{r}(t(s)) = \mathbf{r}(s)$. The explicit appearance of the time parameter is now eliminated. Moreover, since $\frac{ds}{dt} = |\mathbf{r}'(t)|$ and $|\mathbf{r}'(t)|$ is never zero, we have $\frac{dt}{ds} = \frac{1}{|\mathbf{r}'(t)|}$. We will use this fact immediately below.

Consider $\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}$ by definition. Note that $\frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \frac{d\mathbf{r}}{dt}\frac{dt}{ds} = \frac{d\mathbf{r}}{ds}$, so $\frac{d\mathbf{r}}{ds}$ is already the unit tangent vector. We denote $\frac{d\mathbf{r}}{ds}$ by $\mathbf{T}(s)$. Next we define the curvature κ in terms of the arc length parameter s. $\kappa = \left|\frac{d\mathbf{T(s)}}{ds}\right|$. Simpler than using t. Now for the unit normal: $\mathbf{N}(s) = \frac{1}{\kappa}\frac{d\mathbf{T(s)}}{ds}$. The curvature was exactly the factor needed to normalize $\frac{d\mathbf{T(s)}}{ds}$. Finally the binormal: $\mathbf{B}(s) = \mathbf{T}(s) \times \mathbf{N}(s)$. Our program of expressing the vectors of the $\mathbf{TNB}$ frame along with curvature in terms of arc length is now complete. One loose end remains, namely how we can express torsion τ in terms of s. Translating into symbols our qualitative description of torsion as the projection of the rate of change of the unit normal onto the unit binormal, we have $\tau = \frac{d\mathbf{N(s)}}{ds} \cdot \mathbf{B}(s)$. This is mercifully simpler than the formula for torsion purely in terms of $\mathbf{r}(t)$ and its derivatives.

One might reasonably ask why we bother to develop the space curve formalism in terms of time when everything seems simpler in terms of arc length. The disappointing lesson here is that the formulas expressed in terms of arc length appear to be relatively simpler than those for the time parameter, but that may be an illusion, since finding $s(t)$ in order to eliminate t can easily become a nightmare...see Example 9 above for a fairly innocent but frustrating case. We note that $\mathbf{r}(t)$ does have its charm for practical problems, especially those involving motion along space curves.

4.73 TRAJECTORIES (∫)

Ballistic trajectories are those space curves followed by non-self-propelled objects launched from the surface of the Earth and subject to the returning force of gravity. These curves are assumed to happen in a plane, so we can frame our analysis in $\mathbb{R}^2$. It is customary to neglect air resistance or any other force other than the initial launch. Mass of the object does not enter into the calculation of the flight path since downward acceleration is a constant $g = 32.16$ feet per second^2 (9.81 meters per second^2) for any object under these assumptions.

Suppose the projectile is launched along a straight line elevated at an angle θ with respect to the ground or other datum plane with initial velocity v_0. You can imagine a cannon with the barrel inclined at angle θ firing a cannonball with muzzle velocity v_0. The horizontal component of velocity $v_h = v_0\cos\theta$ and the vertical component of velocity is $v_v = v_0\sin\theta$. Gravity causes the cannonball to have a net upward velocity at time t of $v_0\sin\theta - gt$. Expressing the trajectory as the vector function $\mathbf{r}(t)$, what we have at this point is $\mathbf{r}'(t) = \langle v_0\cos\theta, v_0\sin\theta - gt\rangle$. To recover $\mathbf{r}(t)$ we integrate each component with respect to time, so $\mathbf{r}(t) = \int_0^t \mathbf{r}'(\xi)d\xi = \left\langle v_0 t\cos\theta, v_0 t\sin\theta - \frac{1}{2}gt^2\right\rangle$. Several important facts about the trajectory can be immediately deduced from this expression.

The **flight time** t_f for the projectile is the non-zero solution to the quadratic equation $v_0 t\sin\theta - \frac{1}{2}gt^2 = 0$. This gives $t_f = \frac{2v_0\sin\theta}{g}$. Since we are assuming no air resistance v_h is undiminished, and the total horizontal distance traveled by the projectile until it comes back to earth is $v_0 t_f\cos\theta = v_0\cos\theta \cdot \frac{2v_0\sin\theta}{g} = \frac{v_0^2\sin 2\theta}{g}$ by the trigonometric identity $\sin 2\theta = 2\sin\theta\cos\theta$. This distance is called the **range** of the projectile. Also under this assumption the path $\mathbf{r}(t)$ is a parabola whose vertex is the highest point reached by the projectile. By symmetry, this height is reached at time $\frac{1}{2}t_f = \frac{v_0 \sin\theta}{g}$ after launch. At this moment the vertical velocity of the projectile is momentarily zero before it starts to accelerate back to earth. Substituting $\frac{1}{2}t_f = \frac{v_0 \sin\theta}{g}$ into the $\hat{\mathbf{j}}$ component of $\mathbf{r}(t)$ we get the **maximum height** reached is $h_{\max} = (v_0 \sin\theta)\left(\frac{v_0 \sin\theta}{g}\right) - \frac{1}{2}g\left(\frac{v_0 \sin\theta}{g}\right)^2 = \frac{(v_0 \sin\theta)^2}{2g}$.

In the event that the projectile is launched at the point (a,b) instead of the origin, the trajectory would be revised to $\mathbf{r}(t) = \left\langle a + v_0 t\cos\theta, b + v_0 t\sin\theta - \frac{1}{2}gt^2\right\rangle$. The formulas for flight time, range, and maximum height would have to be adjusted accordingly. The presence of a headwind or tailwind in the plane of flight would result in the adjustment of the horizontal velocity component but not the vertical.

Table 8 - **Key Trajectory Formulas**
Launch at origin with velocity v_0, inclination θ over level ground

1) Flight time	$\frac{2v_0 \sin\theta}{g}$
2) Range	$\frac{v_0^2 \sin 2\theta}{g}$
3) Maximum height	$\frac{(v_0 \sin\theta)^2}{2g}$

In the examples below, unless otherwise noted, assume the projectile is launched at the origin and travels over level ground.

Example 1 *A rifle shoots a bullet with velocity* 1000 *feet per second (fps) with the barrel elevated* 30^o *above horizontal. Find the flight time, range, and maximum height attained.*

Solution: $v_0 = 1000\ [fps]$, $\theta = 30^o$, $g = 32.16\ [fps^2]$. Flight time is $\frac{2(1000)\sin 30^o}{32.16} = 31.09\ [s]$. Range is $\frac{(1000)^2 \sin 60^o}{32.16} = 26928.7\ [ft]$ or about $5.1\ [mi]$. Maximum height is $\frac{(1000 \sin 30^o)^2}{2(32.16)} = 3886.8\ [ft]$.

Example 2 *The "Green Monster" at Fenway Park in Boston is the left field fence at the baseball park. The fence is 310 feet from home plate down the left field line. Although this distance is smaller than that in most major league parks, the fact that the fence is 37 feet high makes it difficult to hit a home run down the left field line. If a baseball is hit down the left field line starting at* 3 *feet off the ground and traveling at* 110 *mph initial velocity* *(which is constant since we are neglecting air resistance), what is the minimum angle of inclination from horizontal for which the ball would just clear the fence and what is the maximum angle that would still result in a home run?*

Solution: *The trajectory equation is* $\mathbf{r}(t) = (v_0 \cos\theta)t\hat{\mathbf{i}} + ((v_0 \sin\theta)t - 16t^2)\hat{\mathbf{j}}$. *We do the problem in feet per second and note that* 110 *miles per hour is* 161.333 $[fps]$. *Each component in the trajectory equation gives a relation between* t *and* θ, *but although we have two equations in two unknowns, it is a non-linear situation. We solve by eliminating* t *and obtain a trigonometric relation for* θ. *So...,since the fence is* 310 *ft from the point of launch, the time to reach the wall is given by* $t_{wall} = \dfrac{310}{161.333\cos\theta} = \dfrac{1.921}{\cos\theta}$. *Plugging this into the equation for height and setting it equal to* 34 *ft (we start* 3 *ft off the ground) we get* $(161.333\sin\theta)\left(\dfrac{1.921}{\cos\theta}\right) - 16\left(\dfrac{1.921}{\cos\theta}\right)^2 = 34$. *This simplifies to* $310\dfrac{\sin\theta}{\cos\theta} - \dfrac{59.044}{\cos^2\theta} = 34$, *which we write as* $310\tan\theta - 59.044\sec^2\theta = 34$, since $\dfrac{1}{\cos\theta} = \sec\theta$. *Using the trigonometric formula* $\sec^2\theta = 1 + \tan^2\theta$ *we recover a quadratic equation in* $\tan\theta$:
$310\tan\theta - 59.044(1+\tan^2\theta) = 34$, or $59.044\tan^2\theta - 310\tan\theta + 93.044 = 0$. Solving this (*quadratic formula) we get the roots* $\tan\theta = 0.3196$.and 4.931. These yield the angles 17.72^o *and* 78.54^o. *The smaller angle corresponds to a flat trajectory wshere the ball just makes it over the fence and the larger angle corresponds to a high parabolic arc that, if it were any higher, would come down and hit the fence or the field before* 310 *feet. Any launch angle in between these angles would result in the ball clearing the fence.*

Example 3 *For a cannon with given muzzle velocity* v_0, *find the launch angle yielding maximum range. Assume both launch and impact at ground level.*

Solution: *Let* $R(\theta)$ *be the range as a function of launch angle for a given* v_0. *We know* $R(\theta) = \dfrac{v_0^2 \sin 2\theta}{g}$, so the answer to the question involves finding the angle θ that maximizes $\sin 2\theta$. The sine function attains its maximum at 90^o, so $2\theta = 90^o$ gives $\theta = 45^o$ as the angle yielding maximum range for any muzzle velocity.

Example 4 *An anti-aircraft gun is aimed due east and fires projectiles traveling at* 1200 $[fps]$. *An enemy aircraft is flying due west towards the location of the gun at a constant altitude of* $5{,}000$ $[ft]$. *If the gun is inclined* 60^o *from horizontal, is it capable of shooting down the plane? If so, how many miles east of the gun location could this occur?*

Solution:The trajectory equation is $r(t) = (1200\cos 60^o)t\hat{\mathbf{i}} + ((1200\sin 60^o)t - 16t^2)\hat{\mathbf{j}}$. The anti-aircraft shell reaches the plane altitude whenever $(1200\sin 60^o)t - 16t^2 = 5000$. Rewriting this as a quadratic, $16t^2 - 1039.2t + 5000 = 0$. This has solutions 5.233 and 59.717 $[s]$. Obviously, the shorter time corresponds to hitting the enemy plane as the shell rises rather than falls. After 5.233 seconds, the shell is $(1200)(0.500)(5.233) = 3139.8$ feet (0.595 $[mi]$) east of the gun at the point of interception.

4.74 PLANES

A **plane** in $\mathbb{R}^3$ is uniquely determined by three non-collinear points, say $P_1 = (a_1,b_1,c_1)$, $P_2 = (a_2,b_2,c_2)$ and $P_3 = (a_3,b_3,c_3)$. We would like to derive the equation of the plane they determine in the form $ax + by + cz = D$. This is the **canonical** or **standard** form of the equation of a plane in $\mathbb{R}^3$. Here is a simple plan to achieve this. Form the vectors $\mathbf{v}_1 = \overrightarrow{P_1P_2}$ and $\mathbf{v}_2 = \overrightarrow{P_1P_3}$ and take the cross product $\mathbf{v}_1 \times \mathbf{v}_2$. By the properties of the cross product we know $\mathbf{v}_1 \times \mathbf{v}_2$ will be perpendicular to both $\mathbf{v}_1$ and $\mathbf{v}_2$, hence perpendicular to the common plane of P_1, P_2, and P_3. We will use this fact to construct a generic vector in that common plane. Let $\mathbf{v} = \langle x - a_1, y - b_1, z - c_1\rangle$. Now set $\mathbf{v} \cdot \mathbf{v}_1 \times \mathbf{v}_2 = 0$. By the properties of the dot product the generic vector $\mathbf{v}$ is now perpendicular to the cross product which by design was already perpendicular to the original plane. In other words, the generic vector is back in the original plane. Let us carry out this construction in finer detail and actually derive the equation of the plane in canonical form.

We have $\mathbf{v}_1 = \langle a_2 - a_1, b_2 - b_1, c_2 - c_1\rangle$ and $\mathbf{v}_2 = \langle a_3 - a_1, b_3 - b_1, c_3 - c_1\rangle$. Before we apply our formula for the cross product, let us compress our notation. Set $\langle a_2 - a_1, b_2 - b_1, c_2 - c_1\rangle = \langle A_2, B_2, C_2\rangle$ and $\langle a_3 - a_1, b_3 - b_1, c_3 - c_1\rangle = \langle A_3, B_3, C_3\rangle$. Now $\mathbf{v}_1 \times \mathbf{v}_2 = \langle B_2C_3 - B_3C_2, -A_2C_3 + A_3C_2, A_2B_3 - A_3B_2\rangle$. Dotting this with $\mathbf{v} = \langle x - a_1, y - b_1, z - c_1\rangle$ we finally can write $(x - a_1)(B_2C_3 - B_3C_2) + (y - b_1)(-A_2C_3 + A_3C_2) + (z - c_1)(A_2B_3 - A_3B_2) = 0$. Putting this in canonical form, $(B_2C_3 - B_3C_2)x + (-A_2C_3 + A_3C_2)y + (A_2B_3 - A_3B_2)z = a_1(B_2C_3 - B_3C_2) + b_1(-A_2C_3 + A_3C_2) + c$ Some examples are in order. The first is very simple.

Example 1 *Find the equation of the plane in standard form that contains the points* $(0,0,0)$*,* $(1,2,3)$*, and* $(3,5,7)$*.*

Solution: Let the origin be the common initial point for the two vectors. Then $\mathbf{v}_1 = \langle 1,2,3\rangle$ and $\mathbf{v}_2 = \langle 3,5,7\rangle$. We have $\mathbf{v}_1 \times \mathbf{v}_2 = \langle -1,2,-1\rangle$. Now let $\mathbf{v} = \langle x - 0, y - 0, z - 0\rangle$. Finally $\mathbf{v} \cdot \mathbf{v}_1 \times \mathbf{v}_2 = 0$ yields $-x + 2y - z = 0$. Done. We can easily check that the three points belong to the plane.

Example 2 *Find the equation of the plane in standard form that contains the points* $(1,2,2)$*,* $(3,-1,4)$*, and* $(6,5,8)$*.*

Solution: Select $(1,2,2)$ as the common initial point. Then $\mathbf{v}_1 = \langle 2,-3,2\rangle$ and $\mathbf{v}_2 = \langle 5,3,6\rangle$. We have $\mathbf{v}_1 \times \mathbf{v}_2 = \langle -24,-2,21\rangle$. Also $\mathbf{v} = \langle x - 1, y - 2, z - 2\rangle$, so $\mathbf{v} \cdot \mathbf{v}_1 \times \mathbf{v}_2 = 0$ yields $-24(x - 1) - 2(y - 2) + 21(z - 2) = 0$. Putting this in the required form we have $-24x - 2y + 21z = 14$.

A **direction vector** for a plane is any vector perpendicular to the plane. We can use the second part of the preceding construction to recover the equation of the plane from a direction vector if we also know one point on the plane. Note that a vector in the plane is not sufficient to define the plane.

Example 3 *A plane containing the point* $(1,1,1)$ *has direction vector* $\mathbf{u} = \langle 2,-1,3\rangle$*. Find the equation of the plane in standard form.*

Solution: A generic vector in the plane is $\mathbf{v} = \langle x-1, y-1, z-1\rangle$. Since $\mathbf{u}$ is perpendicular to $\mathbf{v}$ we have $\mathbf{u}\cdot\mathbf{v} = 0$. So $2(x-1)-(y-1)+3(z-1) = 0$. Putting in standard form we get $2x-y+3z = 4$. Notice that the components of the direction vector always reappear as the respective coefficients in the plane equation. This is a very convenient fact that we will exploit regularly. If the direction vector had been reversed (antiparallel), we would still have $(-\mathbf{u})\cdot\mathbf{v} = -(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} = 0$ with no change to the plane equation.

Example 4 *A plane containing the point* $(4,7,-13)$ *has direction vector* $\langle 11,-23,8\rangle$*. Find its equation in standard form.*

Solution: From the remark in Example 3 we can immediately write $11(x-4)-23(y-7)+8(z+13) = 0$. Moving the constants to the right hand side we have $11x-23y+8z = 44-161-104 = -221$.

Planes can either be parallel (possibly identical) or intersect in a line. The angle between intersecting planes is called a **dihedral angle**. The dihedral angle between planes is the same as the angle between their direction vectors. Recall (§4.5, Example 1) that this angle is given by the formula $\theta = \arccos\left(\frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{v}||\mathbf{w}|}\right)$. If two planes are parallel, the respective coefficients in the standard equation for one plane would be the same multiple as the coefficients in the other one. For example, the plane given by $6x-15y+3z = 13$ is parallel to the plane given by $2x-5y+z = 11$, since the respective coefficients of the former are three times the coefficients of the latter. In the event that the constant terms are also in the same ratio, the two equations would describe the identical plane. So $6x-15y+3z = 33$ clearly gives the same plane as the one given by $2x-5y+z = 11$. Note that a direction vector for the line of intersection between two planes is simply the cross product of the direction vectors for the two planes.

Example 5 *Determine if the planes given by* $2x-5y+z = 11$ *and* $3x+7y-4z = 9$ *are parallel. If not, find the dihedral angle in which they intersect.*

Solution: Checking the respective coefficients, we confirm they are not related by a fixed multiple, so they are not parallel. To get the angle of intersection we construct direction vectors based on the respective coefficients. Let $\mathbf{v} = \langle 2,-5,1\rangle$ and $\mathbf{w} = \langle 3,7,-4\rangle$. Then $\theta = \arccos\left(\frac{\langle 2,-5,1\rangle\cdot\langle 3,7,-4\rangle}{|\langle 2,-5,1\rangle||\langle 3,7,-4\rangle|}\right) = \arccos\left(\frac{-33}{\sqrt{30}\,\sqrt{74}}\right) = \arccos(-0.701) = 134.5^o$. The corresponding acute angle is 45.5^o.

Example 6 *Find the equation of the line of intersection of the two non-parallel planes given by* $x - y + z = 6$ *and* $2x + y - z = 3$.

Solution: We want the points (x,y,z) that belong to both planes, so we solve $x - y + z = 6$ and $2x + y - z = 3$ simultaneously. Since we have two equations in three unknowns we will have one variable free as a parameter. The solution with z as the parameter is $x = 3$, $y = z - 3$, and $z = z$. Expressing this in the manner of §4.71 we find the line of intersection is $\mathbf{r}(t) = \langle 3,-3,0\rangle + t\langle 0,1,1\rangle$.

Example 7 *Continuing Example 6, find a vector that lies in the plane given by* $x - y + z = 6$ *but has its initial point on the line* $\mathbf{r}(t)$ *and is perpendicular to that line.*

Solution: We are free to pick the initial point on the line, so to keep it simple we select $(3,-3,0)$. The proposed vector will be $\mathbf{v} = \langle x-3, y+3, z\rangle$. It must satisfy the condition of perpendicularity with respect to the line $\mathbf{r}(t)$ and also have its components satisfy $x - y + z = 6$. For perpendicularity we have $\mathbf{v} \cdot \langle 0,1,1\rangle = 0$ or $(y+3) + z = 0$. Solving this simultaneously with $x - y + z = 6$ we find $x = 3 - 2z$, $y = -3 - z$, and $z = z$ give a set of solutions with parameter z. We may choose z arbitrarily, so for simplicity take $z = 1$. Now $(x,y,z) = (1,-4,1)$. Plugging these values back into $\mathbf{v} = \langle x-3, y+3, z\rangle$, we find $\mathbf{v} = \langle -2,-1,1\rangle$. This vector answers the question. Checking that it is perpendicular to the direction vector of the line we get $\langle -2,-1,1\rangle \cdot \langle 0,1,1\rangle = 0$. Likewise with respect to the direction vector of the plane we have $\langle -2,-1,1\rangle \cdot \langle 1,-1,1\rangle = 0$, so it lies in the plane.

It will be useful to have a formula that gives a **point-to-plane distance**. Our derivation is based on connecting a given point to any arbitrary point in the given plane and calculating the length of that line segment. Of course, it would be very unlikely that the line segment chosen so randomly would be perpendicular to the plane, which would be required to get the true distance of the point. The idea of projecting the line segment onto the direction vector of the plane comes to our rescue. The projected length becomes the perpendicular and therefore true distance point-to-plane. Suppose we are given the point (x_1,y_1,z_1) and the plane with equation $ax + by + cz = D$. Choose an arbitrary point (x_2,y_2,z_2) in the plane and construct the vector $\mathbf{v} = \langle x_1 - x_2, y_1 - y_2, z_1 - z_2\rangle$. A unit vector parallel to the direction vector of the plane is $\mathbf{u} = \dfrac{\langle a,b,c\rangle}{\sqrt{a^2+b^2+c^2}}$. The scalar projection of $\mathbf{v}$ in the direction of $\mathbf{u}$ will be the (signed) perpendicular distance of (x_1,y_1,z_1) from the plane. From §4.5, Example 4 this projection is

$$\frac{\langle a,b,c\rangle}{\sqrt{a^2+b^2+c^2}} \cdot \langle x_1 - x_2, y_1 - y_2, z_1 - z_2\rangle = \frac{a(x_1-x_2) + b(y_1-y_2) + c(z_1-z_2)}{\sqrt{a^2+b^2+c^2}}.$$

So the point-to-plane distance $d = \left| \dfrac{a(x_1-x_2) + b(y_1-y_2) + c(z_1-z_2)}{\sqrt{a^2+b^2+c^2}} \right|$.

Example 8 *Given the plane with equation* $2x + 3y - 5z = 10$*, find the distance from the origin to the plane.*

Solution: To get an arbitrary point in the plane, the simplest approach is to set $x = y = 0$ so that $z = -2$. Then the given point is $(0,0,0)$ and the arbitrary point in the plane is $(0,0,-2)$. The vector $\mathbf{v}$ defined above is $\langle 0,0,2\rangle$, and the direction vector of the plane is $\langle 2,3,-5\rangle$. Using our formula $d = \left| \dfrac{2(0) + 3(0) - 5(2)}{\sqrt{2^2 + 3^2 + (-5)^2}} \right| = \dfrac{10}{\sqrt{38}} \approx 1.62$.

Example 9 *Given the plane with equation* $x - 2y + 4z = 7$*, find the distance from the point* $(2,1,5)$ *to the plane.*

Solution: Keeping things as simple as possible, let the point in the plane be $(7,0,0)$. Then $\left| \dfrac{a(x_1 - x_2) + b(y_1 - y_2) + c(z_1 - z_2)}{\sqrt{a^2 + b^2 + c^2}} \right| = \left| \dfrac{(2-7) - 2(1-0) + 4(5-0)}{\sqrt{1^2 + (-2)^2 + 4^2}} \right| = \dfrac{13}{\sqrt{33}} \approx 2.26$.

Regarding **plane-to-plane distance**, recall that two planes can have one of three relationships: (i) they can intersect, (ii) they can be parallel and separated by a positive distance, or (iii) they can be parallel and be identical. In cases (i) and (iii), the distance between planes is zero. To handle case (ii) select a point in either plane and find its distance to the other plane. The distance from a point to a solid figure bounded by planes can be determined by finding the equations of the planes bounding the solid and carrying out the calculation in Examples 8 or 9 with respect to each of the bounding planes. Then select the shortest distance.