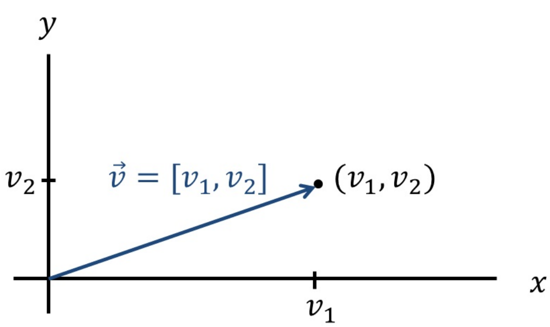
Chapter 2. Vectors and Vector Spaces

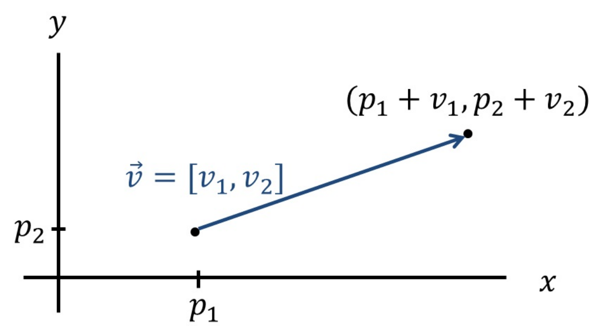
Section 2.2. Cartesian Coordinates and Geometrical Properties of Vectors

Note. There is a natural relationship between a *point* in $\mathbb{R}^n$ and a *vector* in $\mathbb{R}^n$. Both are represented by an n-tuple of real numbers, say $(x_1, x_2, \ldots, x_n)$. In sophomore linear algebra, you probably had a notational way to distinguish vectors in $\mathbb{R}^n$ from points in $\mathbb{R}^n$. For example, Fraleigh and Beauregard in *Linear Algebra*, 3rd Edition (1995), denote the point $x \in \mathbb{R}^n$ as $(x_1, x_2, \ldots, x_n)$ and the vector $\vec{x} \in \mathbb{R}^n$ as $\vec{x} = [x_1, x_2, \ldots, x_n]$. Gentle makes no such notational convention so we will need to be careful about how we deal with points and vectors in $\mathbb{R}^n$, when these topics are together. Of course, **there is a difference between points in $\mathbb{R}^n$ and vectors in $\mathbb{R}^n$** (a common question on the departmental Linear Algebra Comprehensive Exams)!!! For example, vectors in $\mathbb{R}^n$ can be added, multiplied by scalars, they have a "direction" (an informal concept based on existence of an ordered basis and the component of the vector with respect to that ordered basis). But vectors don't have any particular "position" in $\mathbb{R}^n$ and they can be translated from one position to another. Points in $\mathbb{R}^n$ do have a specific position given by the coordinates of the point. But you cannot add points, multiply them be scalars, and they have neither magnitude nor direction. So the properties which a vector in $\mathbb{R}^n$ has are not shared by a point in $\mathbb{R}^n$ and vice-versa.

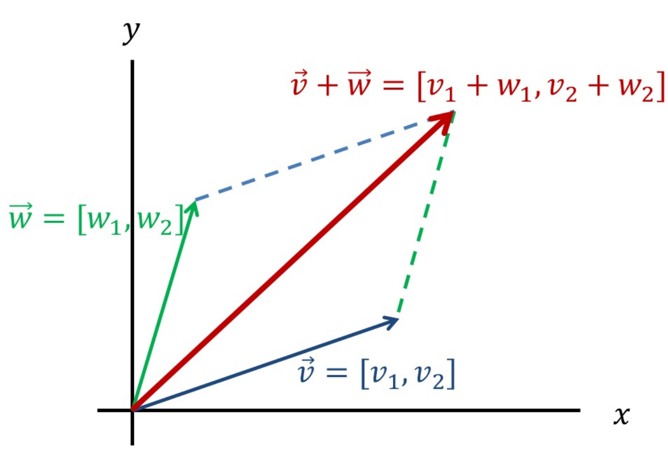
Note. We shall refer to the "natural relationship" between points in $\mathbb{R}^n$ and vectors in $\mathbb{R}^n$ as the *geometric interpretation* of vectors. A vector in $\mathbb{R}^n$ with components $x_1, x_2, \ldots, x_n$ (in order) can be geometrically interpreted as an "arrow" with its "tail" at the origin of an n-dimensional real coordinate system and its "head" at the point in $\mathbb{R}^n$ with coordinates $x_1, x_2, \ldots, x_n$ (in order). Thusly interpreted, the vector is said to be in *standard position.* When $n = 2$ this produces a nice way to illustrate vectors.

Note. In $\mathbb{R}^2$, a vector $\vec{v}$ in $\mathbb{R}^2$ (we briefly revert to sophomore level notation) can be drawn in standard position in the Cartesian plane (left) and can be drawn translated to a point other than the origin (right):

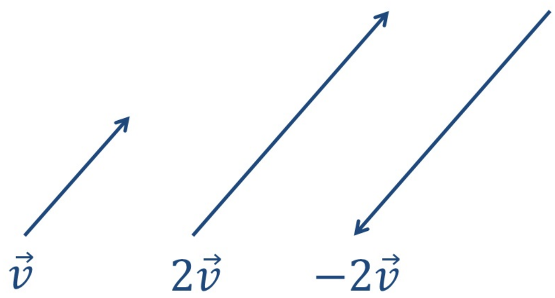




The parallelogram law of vector addition is illustrated as:



Scalar multiplication is illustrated as:



Note. In order to write a vector in terms of an orthogonal (or, preferably, an orthonormal) basis, we will make use of projections.

Definition. Let $x, y \in V$, where V is a vectors space of n-vectors. The *projection* of y onto x is $\text{proj}_x(y) = \hat{y} = \dfrac{\langle x, y\rangle}{\|x\|^2}x$.

Note. For $x, y \in V$, $\hat{y} = \text{proj}_x(y)$ is the component of y in the direction of x. This is justified when we observe that

$$\langle y - \text{proj}_x(y), x\rangle = \langle y - \hat{y}, x\rangle = \left\langle y - \frac{\langle x, y\rangle}{\|x\|^2}x, x\right\rangle = \langle y, x\rangle - \frac{\langle x, y\rangle}{\|x\|^2}\langle x, x\rangle = 0.$$

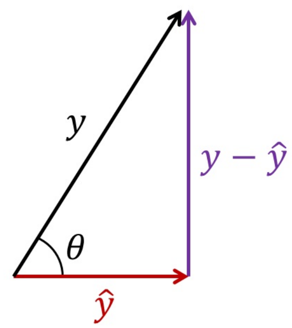
So $y - \hat{y}$ is the component of y orthogonal to x. Then $y = \hat{y} + (y - \hat{y})$ where $\hat{y}$ is parallel to x (that is, a multiple of x) and $y - \hat{y}$ is orthogonal to x. We therefore have y, $\hat{y}$, and $y - \hat{y}$ determining a right triangle. Notice that

$$\begin{aligned}
\|\hat{y}\|^2 + \|y - \hat{y}\|^2 &= \left\|\frac{\langle x, y\rangle}{\|x\|^2}x\right\|^2 + \left\|y - \frac{\langle x, y\rangle}{\|x\|^2}x\right\|^2 \\
&= \frac{\langle x, y\rangle^2}{\|x\|^2} + \left\langle y - \frac{\langle x, y\rangle}{\|x\|^2}x, y - \frac{\langle x, y\rangle}{\|x\|^2}x\right\rangle
\end{aligned}$$

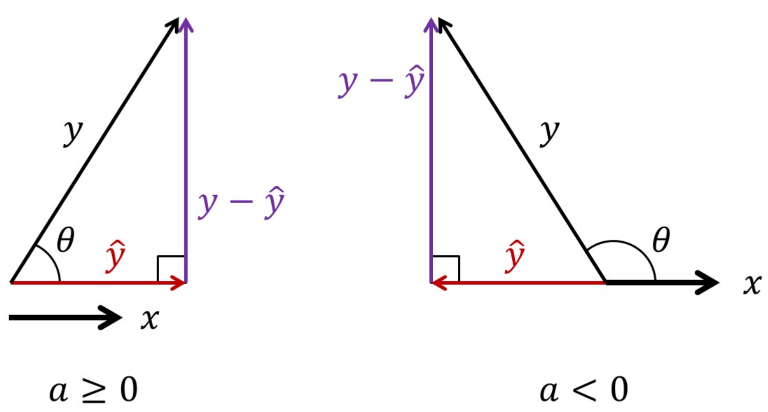
$$= \frac{\langle x,y\rangle^2}{\|x\|^2} + \langle y,y\rangle - 2\frac{\langle x,y\rangle}{\|x\|^2}\langle x,y\rangle + \frac{\langle x,y\rangle^2}{\|x\|^4}\langle x,x\rangle$$
$$= \frac{\langle x,y\rangle^2}{\|x\|^2} + \|y\|^2 - 2\frac{\langle x,y\rangle^2}{\|x\|^2} + \frac{\langle x,y\rangle^2}{\|x\|^2} = \|y\|^2,$$

and so the Pythagorean Theorem is satisfied.

Note. With θ an angle between vectors x and y, from the previous note we expect a geometric interpretation as follows:



So $\cos\theta = \dfrac{\|\hat{y}\|}{\|y\|} = \dfrac{\left\|\frac{\langle x,y\rangle}{\|x\|^2}x\right\|}{\|y\|} = \dfrac{|\langle x,y\rangle|}{\|x\|\|y\|}$. Now $\hat{y}$ is a scalar multiple of x, say $\hat{y} = ax$, then $\langle x,y\rangle = \langle x, \hat{y} + (y-\hat{y})\rangle = \langle x,\hat{y}\rangle + \langle x, y-\hat{y}\rangle = \langle x, ax\rangle + 0 = a\langle x,x\rangle = a\|x\|^2$. If $a \geq 0$ then $\langle x,y\rangle \geq 0$ and $\cos\theta = \langle x,y\rangle/(\|x\|\|y\|)$. If $a < 0$ then $\langle x,y\rangle < 0$ and $\cos\theta = \langle x,y\rangle/(\|x\|\|y\|) < 0$. The geometric interpretation of these two cases are:



This inspires the following definition.

Definition. The *angle* θ between vectors x and y is $\theta = \cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right)$. With e_i as the ith unit vector $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$, the *ith direction cosine* is $\theta_i = \cos^{-1}\left(\frac{\langle x, e_i\rangle}{\|x\|}\right)$.

Note. For $x = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$, we can easily express x as a linear combination of $e_1, e_2, \ldots, e_n$ as $x = \sum_{i=1}^n \langle x, e_i\rangle e_i$. Of course, $\{e_1, e_2, \ldots, e_n\}$ is a basis (actually, an orthonormal basis) for $\mathbb{R}^n$, called the *standard basis* for $\mathbb{R}^n$. With $\cos\theta_i = \frac{\langle x, e_i\rangle}{\|x\|}$ we have

$$\begin{aligned}
\cos^2\theta_1 + \cos^2\theta_2 + \cdots + \cos^2\theta_n &= \left(\frac{\langle x, e_1\rangle}{\|x\|}\right)^2 + \left(\frac{\langle x, e_2\rangle}{\|x\|}\right)^2 + \cdots + \left(\frac{\langle x, e_n\rangle}{\|x\|}\right)^2 \\
&= \frac{1}{\|x\|^2}\left(\langle x, e_1\rangle^2 + \langle x, e_2\rangle^2 + \cdots + \langle x, e_n\rangle^2\right) \\
&= \frac{1}{\|x\|^2}\langle\langle x, e_1\rangle e_1 + \langle x, e_2\rangle e_2 + \cdots + \langle x, e_n\rangle e_n, \\
&\quad \langle x, e_1\rangle e_1 + \langle x, e_2\rangle e_2 + \cdots + \langle x, e_n\rangle e_n\rangle \\
&\quad \text{since } \{e_1, e_2, \ldots, e_n\} \text{ is an orthonormal set} \\
&= \frac{1}{\|x\|^2}\langle x, x\rangle = \frac{\|x\|^2}{\|x\|^2} = 1.
\end{aligned}$$

Note. The representation above of $x = [x_1, x_2, \ldots, x_n]$ in terms of the standard basis $\{e_1, e_2, \ldots, e_n\}$ as $x = \sum_{i=1}^n \langle x, e_i\rangle e_i$ is suggestive of the ease of such representations when using an orthonormal basis. The following theorem allows us to represent x in terms of an orthogonal basis.

Theorem 2.2.1. Let $\{v_1, v_2, \ldots, v_k\}$ be a basis for vector space V of n-vectors where the basis vectors are mutually orthogonal. Then for $x \in V$ we have

$$x = \frac{\langle x, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \frac{\langle x, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 + \cdots + \frac{\langle x, v_k \rangle}{\langle v_k, v_k \rangle} v_k.$$

Corollary 2.2.2. Let $\{v_1, v_2, \ldots, v_k\}$ be an orthonormal basis for vector space V of n-vectors. Then for $x \in V$ we have

$$x = \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 + \cdots + \langle x, v_k \rangle v_k.$$

Definition. If $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal basis for vector space V of n-vectors, then for $x \in V$ the formula $x = \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 + \cdots + \langle x, v_k \rangle v_k$ is the *Fourier expansion* of x and the $\langle x, v_i \rangle$ are the Fourier coefficients (with respect to the given basis).

Note. We now give a technique by which any basis for a vector space V can be transformed into an orthonormal basis.

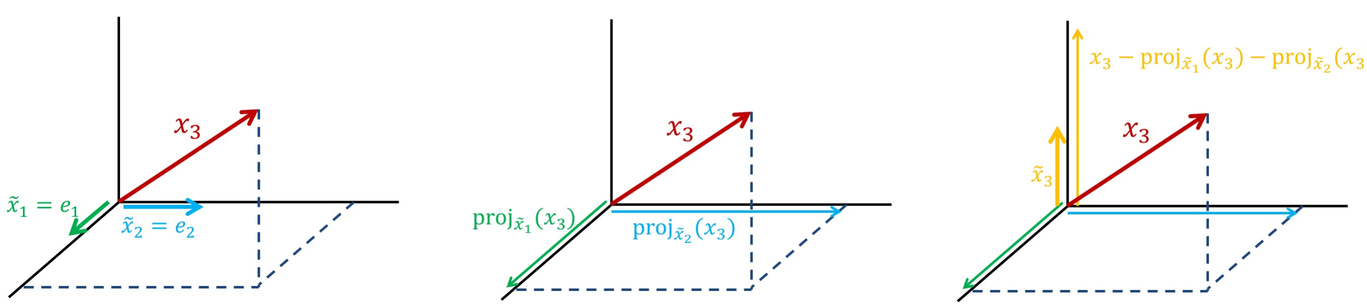
Definition. Let $\{x_1, x_2, \ldots, x_m\}$ be a set of linearly independent n-vectors. Define $\tilde{x}_1 = x_1 / \|x_1\|$. For $k = 2, 3, \ldots, m$ define

$$\tilde{x}_k = \left(x_k - \sum_{i=1}^{k-1} \langle \tilde{x}_i, x_k \rangle \tilde{x}_i \right) \Bigg/ \left\| x_k - \sum_{i=1}^{k-1} \langle \tilde{x}_i, x_k \rangle \tilde{x}_i \right\|.$$

This transformation of set $\{x_1, x_2, \ldots, x_m\}$ into $\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m\}$ is called the *Gram-Schmidt Process.*

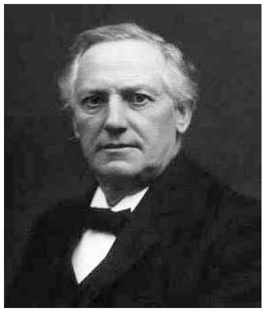
Note. In the Gram-Schmidt Process, for each $1 \le k \le m$ we have that $\tilde{x}_k$ is a linear combination of $x_1, x_2, \ldots, x_k$ (due to the recursive definition). Also, by the definition of $\tilde{x}_k$, we see that x_k is a linear combination of $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k$. So $\text{span}\{x_1, x_2, \ldots, x_k\} = \text{span}\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k\}$. By the Note above which shows that $(y - \text{proj}_x(y)) \perp x$, we have that $\tilde{x}_2 \perp \tilde{x}_1$; $\tilde{x}_3 \perp \tilde{x}_1$ and $\tilde{x}_3 \perp \tilde{x}_2$; $\tilde{x}_4 \perp \tilde{x}_3$, $\tilde{x}_4 \perp \tilde{x}_2$, $\tilde{x}_4 \perp \tilde{x}_1$; etc. Of course each $\tilde{x}_k$ is normalized, so $\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m\}$ is an orthonormal basis for $\text{span}\{x_1, x_2, \ldots, x_m\}$.

Note. Since each $\tilde{x}_k$ is normalized, we can think of the term $\langle \tilde{x}_i, x_k\rangle \tilde{x}_i$ as $\text{proj}_{\tilde{x}_i}(x_k)$. This gives the Gram-Schmidt Process a very geometric flavor! For example, in $\mathbb{R}^3$ suppose that we already have $\tilde{x}_1 = e_1$ and $\tilde{x}_2 = e_2$. With x_3 as illustrated below, we produce $\tilde{x}_3$ as shown.



Note. We might think of the Gram Schmidt Process as producing a "nice" basis from a given basis. One thing nice about an orthonormal basis is the ease with which we can express a vector in terms of the basis elements, as shown in Corollary 2.2.2.

Note. The Gram-Schmidt Process is named for Jorgen Gram (1850–1916) and Erhard Schmidt (1876–1959).



Jorgen Gram (1850–1916)

Erhard Schmidt (1876–1959)

Jorgen Gram was a Danish (i.e., from Denmark) mathematician who worked for an insurance company and published papers in forestry and on probability and numerical methods. In 1883 he published "On Series Determined by the Methods of Least Squares" in *Journal für Mathematik*. This paper contained the process co-named after him (though the process was used by Laplace and Cauchy before Gram). It also is a fundamental paper in the development of integral equations. He did work in number theory, including work on the Riemann zeta function. Unlike most prominent mathematicians, Gram never taught. He died at the age of 65 when he was hit by a bicycle. Erhard Schmidt was born in Estonia and did his doctoral work at the University of Göttingen under David Hilbert's supervision. He spent his professional career at the University of Berlin. Like Gram, he studied integral equations (in fact, for both Gram and Schmidt, the vectors they considered were in fact continuous functions). In 1907 he published a paper on integral equations which contained a proof of the Gram-Schmidt Process (in which he mentions Gram's work). Most of his work was in the area of Hilbert spaces. These historical comments are based on the "MacTutor History of Mathematics archive" and historical comments from Fraliegh and Bauregard's *Linear Algebra*, 3rd edition (1995). The Gram-Schmidt Process also holds in a Hilbert space with a countable basis; for details, see my online notes on Section 5.4, Projections and Hilbert Space Isomorphisms, from *Real Analysis with an Introduction to Wavelets and Applications*, D. Hong, J. Wang, and R. Gardner, Elsevier Press (2005).

Note. With $\{u_1, u_2, \ldots, u_n\}$ an orthonormal basis for a vector space, for any $x = c_1u_1 + c_2u_2 + \cdots + c_nu_n$ we have $\|x\|^2 = \sum_{i=1}^n c_i^2$. This is *Parseval's identity.*

Note. We now consider a collection of linearly independent n-vectors $c_1, c_2, \ldots, c_m$ (column vectors) and an n-vector x (a column vector of variables). For constants $b_1, b_2, \ldots, b_m$ (or scalars; technically, these are 1×1 matrices) we can consider the m equations (since the c_i are linearly independent n-vectors; notice that $m \leq n$ by Exercise 2.1):

$$\begin{aligned} c_1^T x &= b_1 \\ c_2^T x &= b_2 \\ &\vdots \\ c_m^T x &= b_m \end{aligned}$$

This may be more familiar to you as the matrix equation $C\vec{x} = \vec{b}$ where C is a matrix with ith row as the row vector c_i^T (so C is an $m \times n$ matrix), $\vec{x}$ as an $n \times 1$ column vector, and $\vec{b}$ as an $m \times 1$ column vector. So we have m linear equations in n unknowns. From Linear Algebra (MATH 2010), since the rows of C are linearly independent then each row of the row echelon form of C contains a pivot and so there exists a solution to the system of equations. If $m = n$, this means that C is invertible and there is a unique solution. Otherwise the set of solutions has $n - m$ free variables.

Definition. In the system of m equations in n unknowns given above, the set of vectors x which are solutions form a *flat* (or *affine space*) in $\mathbb{R}^n$. If $b_1 = b_2 = \cdots = b_m = 0$ then the system of equations is *homogeneous*. If $m = 1$, the set of vectors x which are solutions form a *hyperplane* in $\mathbb{R}^n$.

Theorem 2.2.2. The solutions to a homogeneous system of equations form a subspace of $\mathbb{R}^n$.

Note. A nonhomogeneous system of equations cannot have a solution set which contains the zero vector. So the solution set, in this case, is not a subspace of $\mathbb{R}^n$. So a flat (or affine space) or hyperplane may not be a subspace. You are probably used to describing flats as "translates of subspaces" and writing them as a translation vector t plus a subspace $\text{span}\{v_1, v_2, \ldots, v_k\}$: $t + \text{span}\{v_1, v_2, \ldots, v_k\}$.

Definition. For a set V of n-vectors (not necessarily a vector space nor a cone) the *dual cone* of V, denoted V^*, is

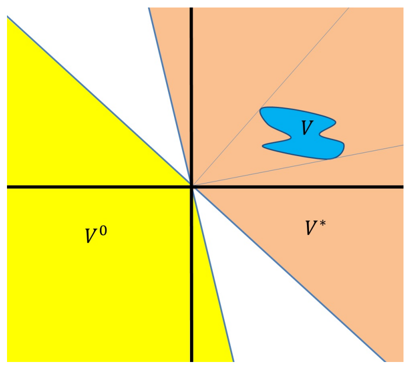
$$V^* = \{y^* \in \mathbb{R}^n \mid (y^*)^T y \geq 0 \text{ for all } y \in V\}.$$

(Notice that all vectors here are treated as column vectors, so the products are defined.) The *polar cone* of V, denoted V^0, is

$$V^0 = \{y^0 \in \mathbb{R}^n \mid (y^0)^T y \leq 0 \text{ for all } y \in V\}.$$

Note. If $y^* \in V^*$ then $(y^*)^T y \geq 0$ for all $y \in V$ and so $-(y^*)^Y y = (-y^*)^T y \leq 0$ for all $y \in V$; that is, $-y^* \in V^0$. Similarly, if $y^0 \in V^0$ then $-y^0 \in V^*$. So $V^0 = -V^*$.

Note. Since $\langle y^*, y\rangle = (y^*)^T y$ (though technically, the left hand side is a scalar and the right hand side is a 1×1 matrix), then V^* includes all vectors in $\mathbb{R}^n$ that make an angle with all elements of set V of less than or equal to $\pi/2$ (recall, the angle θ between two vectors x and y is $\theta = \cos^{-1}(\langle x, y\rangle/(\|x\|\|y\|))$. So we can illustrate a dual cone and polar cone in $\mathbb{R}^2$ as follows (where all vectors are geometrically interpreted to be in standard position):



Example. If $V \subset \mathbb{R}^n$ satisfies $v \in V$ implies $v \geq 0$ (that is, V is a subset of the "nonnegative orthant" and all entries of vector $v \in V$ are nonnegative) then $V^* = \{x \in \mathbb{R}^n \mid x \geq 0\}$ (that is, V^* is the nonnegative orthant) and $V^0 = \{x \in \mathbb{R}^n \mid x \leq 0\}$ (that is, V^0 is the nonpositive orthant). The nonnegative orthant is a convex cone and is its own dual. Our modified Exercise 2.12 shows that for any set V^* and V^0 are cones, that $V^* \cup V^0$ is closed under scalar multiplication, but that $V^* \cup V^0$ is not a vector space (even when V itself is a convex cone).

Definition. For $x = [x_1, x_2, x_3], y = [y_1, y_2, y_3] \in \mathbb{R}^3$, define the *vector cross product* $x \times y$ as

$$x \times y = [x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1].$$

Theorem 2.2.3. Properties of Cross Product.

Let $x, y, z \in \mathbb{R}^3$ and $a \in \mathbb{R}$. Then:

1. $x \times x = 0$ (Self-nilpotency)

2. $x \times y = -y \times x$ (Anti-commutivity)

3. $(ax) \times y = a(x \times y) = x \times (ay)$ (Factoring of Scalar Multiplication)

4. $(x+y) \times z = (x \times z) + (y \times z)$ (Relation of Vector Addition to Addition of Cross Products)

5. $\langle x, x \times y\rangle = \langle y, x \times y\rangle = 0$ (Perpendicular Property)

6. $x \times (y \times z) = \langle x, z\rangle y - \langle x, y\rangle z$.

Note. We can also show that the cross product is not in general associative. For example,

$$[1,0,0] \times ([1,1,0] \times [1,1,1]) = [1,0,0] \times [1,-1,0] = [0,0,-1]$$

$$\neq ([1,0,0] \times [1,1,0]) \times [1,1,1] = [0,0,1] \times [1,1,1] = [-1,1,0].$$

Note. The cross product can be used to find a normal vector to a plane in $\mathbb{R}^3$ which, in turn, gives the formula for a plane in $\mathbb{R}^3$. See my online Calculus 3 notes on "Lines and Planes in Space."

Revised: 6/10/2020