Chapter 2. Vectors and Vector Spaces

Section 2.1. Operations on Vectors

Note. In this section, we define several arithmetic operations on vectors (especially, vector addition and scalar multiplication). We reintroduce much of the terminology associated with vectors from sophomore level Linear Algebra (MATH 2010).

Definition. A (real) *vector space* is a set V of *vectors* along with an operation of addition $+$ of vectors and multiplication of a vector by a scalar (real number), which satisfies the following. For all vectors $x, y, z \in V$ and for all scalars $a, b \in \mathbb{R}$:
(A1) $(x + y) + z = x + (y + z)$ (Associativity of Vector Addition)
(A2) $x + y = y + x$ (Commutivity of Vector Addition)
(A3) There exists $0 \in V$ such that $0 + x = x$ (Additive Identity)
(A4) $x + (-x) = 0$ (Additive Inverses)
(S1) $a(x+y) = ax+ay$ (Distribution of Scalar Multiplication over Vector Addition)
(S2) $(a+b)x = ax+bx$ (Distribution of Scalar Addition over Scalar Multiplication)
(S3) $a(bx) = (ab)x$ (Associativity)
(S4) $1x = x$ (Preservation of Scale).

Note. More generally, we can consider a vector space over a *field*. For details, see my online notes on "Vector Spaces" for Introduction to Modern Algebra 2 (MATH 4137/5137). However, unless otherwise noted, we only consider finite dimensional vector spaces with real scalars.

Definition. Two vectors in $\mathbb{R}^n$, $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_n]$ are *equal* if $x_i = y_i$ for $i = 1, 2, \ldots, n$. The *zero vector* in $\mathbb{R}^n$ is $0_n = 0 = [0, 0, \ldots, 0]$. The *one vector* in $\mathbb{R}^n$ (also called the "summing vector") is $1_n = 1 = [1, 1, \ldots, 1]$. (Later when using 1_n to generate sums, we will treat it as a column vector.)

Definition. Let $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_n]$ be vectors in $\mathbb{R}^n$. Let $a \in \mathbb{R}$ be a scalar. Define

(1) $x + y = [x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n]$, and

(2) $ax = [ax_1, ax_2, \ldots, ax_n]$.

Theorem 2.1.1. Properties of Vector Algebra in $\mathbb{R}^n$.
Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$ be scalars. Then:

A1. $(x + y) + z = x + (y + z)$ (Associativity of Vector Addition)

A2. $x + y = y + x$ (Commutivity of Vector Addition)

A3. $0 + x = x$ (Additive Identity)

A4. $x + (-x) = 0$ (Additive Inverses)

S1. $a(x+y) = ax+ay$ (Distribution of Scalar Multiplication over Vector Addition)

S2. $(a+b)x = ax+bx$ (Distribution of Scalar Addition over Scalar Multiplication)

S3. $a(bx) = (ab)x$ (Associativity)

S4. $1x = x$ (Preservation of Scale)

Note. The statement of Theorem 2.1.1 is as it is stated in Fraleigh and Beauregard's *Linear Algebra*, 3rd Edition, Addison-Wesley Publishing Company (1995). The proof is very elementary and is based on the definition of vector addition, scalar multiplication, and the field properties of $\mathbb{R}$.

Definition. With vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ and scalars $a_1, a_2, \ldots, a_k \in \mathbb{R}$, we have the *linear combination* $a_1v_1 + a_2v_2 + \cdots + a_kv_k \in \mathbb{R}^n$. The set of vectors in $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_k\}$ is *linearly independent* if the equation $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$ implies that $a_1 = a_2 = \cdots = a_k = 0$. A set of vectors in $\mathbb{R}^n$ is *linearly dependent* if it is not linearly independent.

Note. If the set of vectors in $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_k\}$, is linearly dependent then there are some scalars $a_1, a_2, \ldots, a_k \in \mathbb{R}$ not all zero such that $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$.

Note. In Exercise 2.1, the following is established:

Exercise 2.1. The maximum number of vectors in $\mathbb{R}^n$ that forms a linearly independent set is n.

Definition. A set of n-vectors, $V \subset \mathbb{R}^n$, is a *vector space* if it is closed under linear combinations. That is, if for any $x, y \in V$ and any scalars $a, b \in \mathbb{R}$ we have $ax + by \in V$.

Note. The above definition may seem unusually simple to you. Technically, we are basing all the properties in the traditional definition of "vector space" on Theorem 2.1.1 and the previous definition just justifies the fact that V is a *subspace* of $\mathbb{R}^n$.

Note. In Exercise 2.1.B you are asked to prove:

> "For vector space V (as defined above), if W_1 and W_2 are finite subsets of V which are linearly independent and of maximal size then W_1 and W_2 are the same size (or "cardinality"); that is, $|W_1| = |W_2|$."

Notice that by Exercise 2.1, for linearly independent sets W_1 and W_2 we have $|W_1| \leq n$ and $|W_2| \leq n$. This justifies the following definition.

Definition. The maximum number of linearly independent n-vectors in a vector space V is the *dimension* of the vector space, denoted $\dim(V)$. When V consists of n-vectors, n is the *order* of vector space V.

Definition. Two vector spaces V_1 and V_2 both of order n are *essentially disjoint* if $V_1 \cap V_2 = \{0\}$.

Definition. The *ith unit vector* in $\mathbb{R}^n$, denoted e_i, has a 1 in the ith position and 0's in all other positions: $e_i = [0, 0, \ldots, 0, 1, 0, \ldots, 0]$. For $x = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$, the *sign vector* is defined as $\text{sign}(x) = [\text{sign}(x_1), \text{sign}(x_2), \ldots, \text{sign}(x_n)] \in \mathbb{R}^n$ where

$$\text{sign}(x_i) = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i = 0 \\ -1 & \text{if } x_i < 0. \end{cases}$$

Note. An ordering $\preceq$ on a set is a reflexive, antisymmetric, transitive relation. A total ordering on set A is an ordering on A such that for any $a, b \in A$ either $a \preceq b$ or $b \preceq a$. An ordering on set A which is not a total ordering is a partial ordering. Applications of partial orderings often involve the partial ordering of subset inclusion. For example, $\subseteq$ is a partial ordering on $A = \{\{a\}, \{b\}, \{a, b\}\}$ since $\{a\} \subseteq \{a, b\}$, $\{b\} \subseteq \{a, b\}$, but neither $\{a\} \subset \{b\}$ nor $\{b\} \subseteq \{a\}$ holds.

Definition. We define a partial ordering on $\mathbb{R}^n$ as follows. We say x is *greater than* y denoted $x > y$ if $x_i > y_i$ for $i = 1, 2, \ldots, n$. Similarly, we say x is *greater than or equal to* y, denoted $x \geq y$, if $x_i \geq y_i$ for $i = 1, 2, \ldots, n$.

Definition. A subset of a vector space V that is itself a vector space is a *subspace* of V.

Theorem 2.1.2. Let V_1 and V_2 be vector spaces of n-vectors. Then $V_1 \cap V_2$ is a vector space.

Note. A union of two vector spaces of n-vectors need not be a vector space, as you will show by example in Exercise 2.2.

Note. The text refers to a set of vectors of the same order as a "space of vectors." A better term would be a "set of n-vectors."

Definition. If V_1 and V_2 are sets of n-vectors, then the *sum* of V_1 and V_2 is $V_1 + V_2 = \{v_1 + v_2 \mid v_1 \in V_1, v_2 \in V_2\}$.

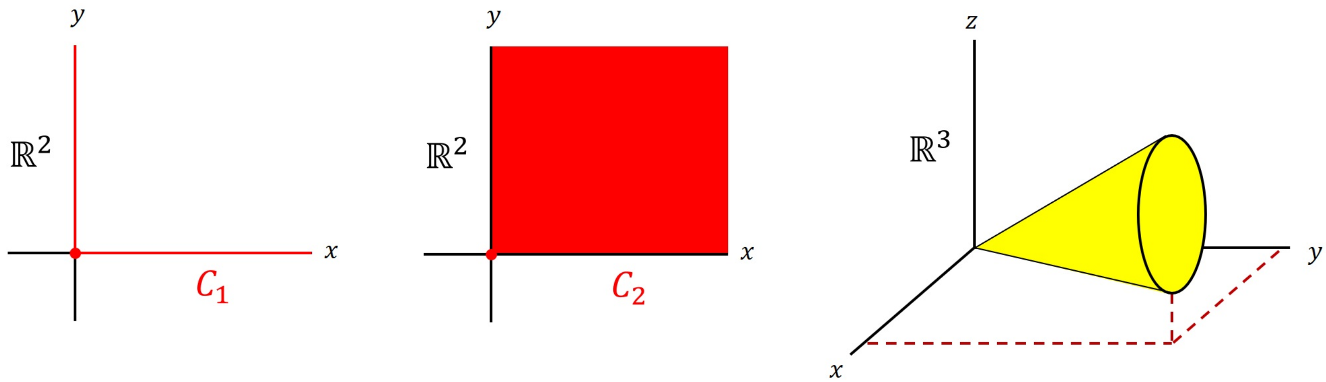
Theorem 2.1.3. If V_1 and V_2 are vector spaces of n-vectors, then $V_1 + V_2$ is a vector space.

Definition. If V_1 and V_2 are essentially disjoint vector spaces of n-vectors, then the vector space $V_1 + V_2$ is the *direct sum* of V_1 and V_2, denoted $V_1 \oplus V_2$.

Theorem 2.1.4. If vector spaces V_1 and V_2 are essentially disjoint then every element of $V_1 \oplus V_2$ can be written as $v_1 + v_2$, where $v_1 \in V_1$ and $v_2 \in V_2$, in a unique way.

Definition. A set of n-vectors that contains all positive scalar multiples of any vector in the set and contains the zero vector is a *cone*. A set of vectors C is a *convex cone* if for all $v_1, v_2 \in C$ and all $a, b \geq 0$, we have $av_1 + bv_2 \in C$.

Example. In $\mathbb{R}^2$, $C_1 = \{[x,0],[0,y] \mid x > 0, y > 0\} \cup \{[0,0]\}$ is a cone but not a convex cone (see the figure below). The set $C_2 = \{v \in \mathbb{R}^2 \mid v > 0\} \cup \{[0,0]\}$ (which includes all vectors in $\mathbb{R}^2$ which when in standard position with their tail at the origin, have their head in the first quadrant of the Cartesian plane) is a convex cone. Similarly, a "single cone" (not double-cone) with its vertex at the origin in $\mathbb{R}^n$ is a cone in the sense defined here; this follows from the parallelogram law of addition of vectors.



The cones C_1, C_2, and a "single cone" in $\mathbb{R}^3$, mentioned above.

Definition. Let G be a subset of vector space V of n-vectors. If each vector in V is some linear combination of elements of G, then G is a *spanning set* (or *generating set*) of V. The set of all linear combinations of elements of G is the *span* of G, denoted $\text{span}(G)$. If a spanning set G of V is linearly independent, then it is a *basis* of V.

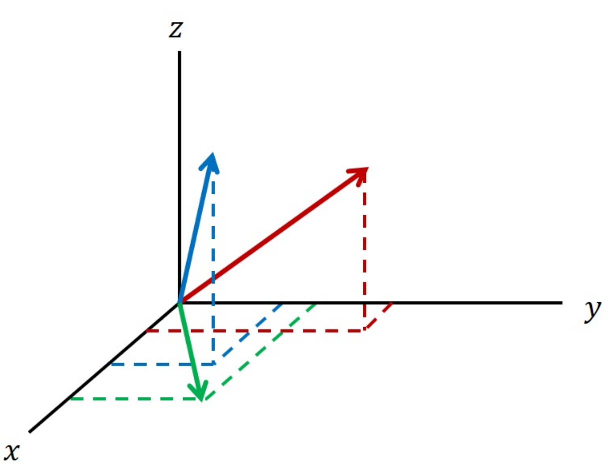
Theorem 2.1.5. If $\{v_1, v_2, \ldots, v_k\}$ is a basis for a vector space V, then each element can be uniquely represented as a linear combination of the basis vectors.

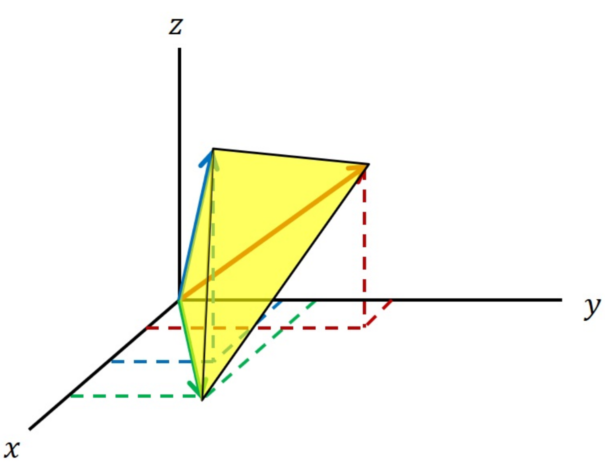
Note. If B_1 is a basis for V_1 and B_2 is a basis for V_2 where V_1 and V_2 are essentially disjoint vector space of n-vectors, then $B_1 \cup B_2$ is a basis for $V_1 \oplus V_2$. This follows from Theorems 2.1.4 and 2.1.5.

Definition. A set of vectors $S = \{v_i \mid i = 1, 2, \ldots\}$ such that for any vector v in a cone C there exists scalars $a_i \geq 0$ for $i = 1, 2, \ldots, k$ so that $v = \sum_{i=1}^{k} a_i v_i$ and if for scalars $b_i \geq 0$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{n} b_i v_i = 0$ then $b_i = 0$ for $i = 1, 2, \ldots, k$ is a *spanning set* (or *generating set*) for cone C. If a spanning set of a cone has a finite number of elements, the cone is a *polyhedron*. A spanning set of a cone consisting of a minimum number of vectors of any spanning set for that cone is a *basis set* for the cone.

Example. In $\mathbb{R}^2$, the cones C_1 and C_2 both have as a basis $\{[1, 0], [0, 1]\}$. Therefore, both are polyhedra.

Note. A polyhedron determined by three vectors is given here (on the left):





Definition. For two n-vectors $x, y \in \mathbb{R}^n$, define the *inner product* of $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_n]$ (also called *dot product* or *scalar product*) as $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$. A vector space with an inner product is an *inner product space*.

Theorem 2.1.6. Properties of Inner Products.

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

1. If $x \neq 0$ then $\langle x, x\rangle > 0$ and $\langle 0, x\rangle = \langle x, 0\rangle = \langle 0, 0\rangle = 0$ (Nonnegativity and Mapping of the Identity),

2. $\langle x, y\rangle = \langle y, x\rangle$ (Commutivity of Inner Products),

3. $a\langle x, y\rangle = \langle ax, y\rangle$ (Factoring of Scalar Multiplication in Inner Products),

4. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ (Relation of Vector Addition to Addition of Inner Products).

Note. We can combine the claim of Theorem 2.1.6 to show that the Inner Product is linear in the first and second entries:

$$\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle \text{ and } \langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle.$$

Note. An important property of inner products is the Schwarz Inequality (or "Cauchy-Schwarz Inequality"). It is used in the proof of the Triangle Inequality once we address norms.

Theorem 2.1.7. Schwarz Inequality.

For any $x, y \in \mathbb{R}^n$ we have $|\langle x, y\rangle| \leq \langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}$.

Definition. A *norm*, $\| \cdot \|$, on a set $S \subset \mathbb{R}^n$ is a function from S to $\mathbb{R}$ satisfying:

1. If $x \neq 0$ then $\|x\| > 0$, and $\|0\| = 0$ (Nonnegativity and Mapping of the Identity),

2. $\|ax\| = |a|\|x\|$ for all $x \in S$ and $a \in \mathbb{R}$ (Relation of Scalar Multiplication to Real Multiplication),

3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in S$ (Triangle Inequality).

A vector space together with a norm is a *normed linear space* (or simply *normed space*).

Example. The *Euclidean norm* on $\mathbb{R}^n$ is the familiar norm $\|x\| = \|[x_1, x_2, \ldots, x_n]\|$ $= \left\{\sum_{k=1}^n (x_k)^2\right\}^{1/2} = \langle x, x\rangle^{1/2}$. If $x \neq 0$ then some $x_i \neq 0$ and so $\|x\| \geq |x_i| > 0$, and $\|0\| = 0$. So "Nonnegativity and Mapping of the Identity" holds. For $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$,

$$\|ax\| = \|[ax_1, ax_2, \ldots, ax_n]\| = \left\{\sum_{k=1}^n (ax_k)^2\right\}^{1/2}$$

$$= \left\{a^2 \sum_{k=1}^n (x_k)^2\right\}^{1/2} = |a| \left\{\sum_{k=1}^n (x_k)^2\right\}^{1/2} = |a|\|x\|,$$

so "Relation of Scalar Multiplication to Real Multiplication" holds. For any $x, y \in \mathbb{R}^n$,

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\
&\leq \langle x, x\rangle + 2\langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2} + \langle y, y\rangle \text{ by the Schwarz Inequality} \\
&= \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,
\end{aligned}$$

and so $\|x + y\| \leq \|x\| + \|y\|$ and the Triangle Inequality holds. Therefore, the Euclidean norm is indeed a norm or $\mathbb{R}^n$.

Note. Another norm on $\mathbb{R}^n$ is the ℓ^p-norm, $\|\cdot\|_p$, for $p \geq 1$ defined as

$$\|x\|_p = \|[x_1, x_2, \ldots, x_n]\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}.$$

We omit the details of the proof that the ℓ^p-norm is indeed a norm. Notice that the ℓ^2-norm is the same as the Euclidean norm. You are likely to encounter the ℓ^p-norm in the future in the setting of ℓ^p-spaces (as opposed to the space $\mathbb{R}^n$) where the ℓ^p-spaces are examples of *Banach spaces*. The ℓ^∞-norm on $\mathbb{R}^n$ is defined as $\|x\|_\infty = \max\{|x_k| \mid k = 1, 2, \ldots, n\}$. Gentle gives a (not very rigorous) argument on page 18 that $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$. Statisticians sometimes call the ℓ^∞-norm the "Chebyshev norm."

Definition. Let $\{v_1, v_2, \ldots, v_k\}$ be a basis for vector space V. For any $x = c_1v_1 + c_2v_2 + \cdots + c_kv_k \in V$ we define the *basis norm* $\rho(x) = \left\{\sum_{j=1}^{k} c_j^2\right\}^{1/2}$.

Theorem 2.1.8. The basis norm is indeed a norm for any basis $\{v_1, v_2, \ldots, v_k\}$ of vector space V.

Definition. Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be norms on a vector space V. Then $\|\cdot\|_a$ is *equivalent* to $\|\cdot\|_b$, denoted $\|\cdot\|_a \cong \|\cdot\|_b$, if there are $r > 0$ and $s > 0$ such that for all $x \in V$ we have $r\|x\|_b \leq \|x\|_a \leq s\|x\|_b$.

Note. The "equivalence" of $\|\cdot\|_a$ and $\|\cdot\|_b$ does not reflect the fact that the norms give the same *value* for any particular vector. Equivalence is ultimately related to the "topology" on the vector space and the behavior of convergence sequences.

Theorem 2.1.9. Equivalence of norms is an equivalence relation. That is, it is reflexive, symmetric, and transitive.

Proof. The proof is left as Exercise 2.1.D. □

Theorem 2.1.10. Every norm on (finite dimensional vector space) V is equivalent to the basis norm ρ for any given basis $\{v_1, v_2, \ldots, v_k\}$. Therefore, any two norms on V are equivalent.

Note. We are only considering finite dimensional spaces, so Theorem 2.1.10 applies to all vector spaces we consider. The result does not hold in infinite dimensions and a given infinite dimensional vector space can have two unequivalent norms.

Definition. A sequence of vectors $x_1, x_2, \ldots$ in a normed vector space V (with norm $\|\cdot\|$) *converges* to vector $x \in V$ if for all $\varepsilon > 0$ there is $M \in \mathbb{N}$ such that if $n \geq M$ then $\|x - x_n\| < \varepsilon$.

Definition. For nonzero vector x in normed vector space V (with norm $\|\cdot\|$), the vector $\tilde{x} = x/\|x\|$ is the *normalized vector* associated with x.

Definition. For S any set of elements, $\Delta : S \times S \to \mathbb{R}$ is a *metric* if for all $x, y, z \in S$ we have:

1. $\Delta(x, y) > 0$ if $x \neq y$ and $\Delta(x, y) = 0$ if $x = y$,

2. $\Delta(x, y) = \Delta(y, x)$,

3. $\Delta(x, y) \leq \Delta(x, z) + \Delta(z, y)$.

Note. We can think of a norm on a vector space as a way to associate a "length" with a vector. A metric on a vector space gives a way to associate a "distance" between two vectors. In a vector space, norms and metrics are intimately related. In fact, if $\|\cdot\|$ is a norm on vector space V then defining $\Delta(x, y) = \|x - y\|$ for all $x, y \in V$ implies that Δ is a metric on V. Conversely, if Δ is a metric on vector space V, then defining $\|x\| = \Delta(0, x)$ is a norm on V. These claims are to be justified in a modified version of Exercise 2.7.

Definition. Two vectors $v_1, v_2 \in V$, where V is a vector space with an inner product, are *orthogonal* if $\langle v_1, v_2 \rangle = 0$. A set of vectors $S = \{v_1, v_2, \ldots, v_k\}$ is an *orthonormal* set if for all $v_i, v_j \in S$ we have $\langle v_i, v_j \rangle = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j. \end{cases}$

Theorem 2.1.11. A set of nonzero vectors $\{v_1, v_2, \ldots, v_k\}$ in a vector space with an inner product for which $\langle v_i, v_j \rangle = 0$ for $i \neq j$ (the vectors are said to be *mutually orthogonal*) is a linearly independent set.

Definition. Two spaces of n-vectors, V_1 and V_2, are *orthogonal*, denoted $V_1 \perp V_2$, if $\langle v_1, v_2 \rangle = 0$ for all $v_1 \in V_1$ and $v_2 \in V_2$. If $V_1 \perp V_2$ and $V_1 \oplus V_2 = \mathbb{R}^n$, then V_2 is the *orthogonal complement* of V_1, denoted $V_2 = V_1^\perp$ (sometimes called the "perp space" of V_1).

Note. By Exercise 2.9, if $V_1 \perp V_2$ then $V_1 \cap V_2 = \{0\}$. That is, if $V_1 \perp V_2$ then V_1 and V_2 are essentially disjoint and so $V_1 \oplus V_2$ is defined.

Note. If B_1 is a basis for V_1 and B_2 is a basis for V_2 where $V_1 \perp V_2$ (and V_1, V_2 are finite dimensional), then $B_1 \cup B_2$ is a basis for $V_1 \oplus V_2$, by Exercise 2.1.F. Since $B_1 \cap B_2 = \varnothing$ ($B_1 \subset V_1$, $B_2 \subset V_2$, and $V_1 \cap V_2 = \{0\}$, but $0 \notin B_1$ and $0 \notin B_2$). So the dimension of $V_1 \oplus V_2$ is $|B_1 \cup B_2| = |B_1| + |B_2|$. That is, $\dim(V_1 \oplus V_2) = \dim(V_1) + \dim(V_2)$.

Note. We introduced the vector $1_n \in \mathbb{R}^n$ as $1_n = [1, 1, \ldots, 1]^T$ above and referred to it as the "summing vector." For any $x \in \mathbb{R}^n$ we have, by treating 1_n^T ("1_n transpose") as a $1 \times n$ matrix and x as a $n \times 1$ matrix, that as a matrix product $1_n^T x = \sum_{i=1}^n x_i$. (Technically, the right hand side is a 1×1 matrix.)

Definition. For $x \in \mathbb{R}^n$, the *arithmetic mean* (or simply *mean*) $\overline{x}$ is defined as $\overline{x} = \dfrac{1}{n} 1_n^T x$. The *mean vector* associated with $x \in \mathbb{R}^n$ is $\overline{x} = [\overline{x}, \overline{x}, \ldots, \overline{x}] \in \mathbb{R}^n$.

Note. We have $\|[\overline{x}, \overline{x}, \ldots, \overline{x}]\|^2 = n\overline{x}^2$. Gentle also uses the symbol "$\overline{x}$" for both the mean of the entries in x and the mean vector of x. We follow his lead, but try to verbally express whether $\overline{x}$ represents a scalar or vector.

Revised: 6/10/2020