2 Vector Spaces

A set V is a **vector space** over $\mathbb{R}$ (field), and its elements are called **vectors**, if there are 2 operations defined on it:

1. Vector addition, that assigns to each pair of vectors $v_1, v_2 \in V$ another vector $w \in V$ (we write $v_1 + v_2 = w$)

2. Scalar multiplication, that assigns to each vector $v \in V$ and each scalar $r \in \mathbb{R}$ (field) another vector $w \in V$ (we write $rv = w$)

that satisfy the following 8 conditions $\forall\ v_1, v_2, v_3 \in V$ and $\forall\ r_1, r_2 \in \mathbb{R}$ (filed):

1. Commutativity of vector addition: $v_1 + v_2 = v_2 + v_1$

2. Associativity of vector addition: $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$

3. Identity element of vector addition: $\exists$ vector $0 \in V$, s.t. $v + 0 = v$, $\forall\ v \in V$

4. Inverse elements of vector addition: $\forall\ v \in V\ \exists\ -v = w \in V$ s.t. $v + w = 0$

5. Compatibility of scalar multiplication with (field) multiplication: $r_1(r_2 v) = (r_1 r_2)v$, $\forall\ v \in V$

6. Distributivity of scalar multiplication with respect to (field) addition: $(r_1 + r_2)v = r_1 v + r_2 v$, $\forall\ v \in V$

7. Distributivity of scalar multiplication with respect to vector addition: $r(v_1 + v_2) = rv_1 + rv_2$, $\forall\ r \in \mathbb{R}$

8. Identity element of scalar multiplication: $1v = v$, $\forall\ v \in V$

Vector spaces over fields other than $\mathbb{R}$ are defined similarly, with the multiplicative identity of the field replacing 1. We won't concern ourselves with those spaces, except for when we'll be needing complex numbers later on. Also, we'll be using the symbol 0 to designate both the number 0 and the vector 0 in V, and you should always be able to tell the difference from the context. Sometimes, we'll emphasize that we're dealing with, say, $n \times 1$ vector 0 by writing $0_{n\times 1}$.

Vector space is an elementary object considered in the linear algebra. Here are some concrete examples:

1. Vector space $\mathbb{R}^n$ with usual operations of element-wise addition and scalar multiplication. An example of these operations in $\mathbb{R}^2$ is illustrated above.

2. Vector space $F_{[-1,1]}$ of all functions defined on interval $[-1, 1]$, where we define $(f+g)(x) = f(x) + g(x)$ and $(rf)(x) = rf(x)$.

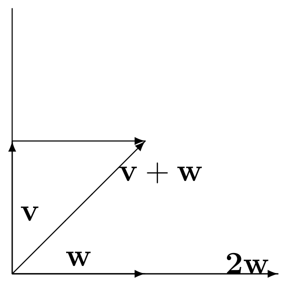

Figure 1: Vector Addition and Scalar Multiplication

2.1 Basic Concepts

Subspace and span We say that $S \subset V$ is a **subspace** of V, if S is closed under vector addition and scalar multiplication, i.e.

1. $\forall s_1, s_2 \in S,\ s_1 + s_2 \in S$
2. $\forall s \in S,\ \forall r \in \mathbb{R},\ rs \in S$

You can verify that if those conditions hold, S is a vector space in its own right (satisfies the 8 conditions above). Note also that S has to be non-empty; the empty set is not allowed as a subspace.

Examples:

1. A subset $\{0\}$ is always a subspace of a vectors space V.
2. Given a set of vectors $S \subset V$, $\mathbf{span}(S) = \{w : w = \sum_{i=1}^n r_i v_i, r_i \in \mathbb{R},\ \text{and}\ v_i \in S\}$, the set of all linear combinations of elements of S (see below for definition) is a subspace of V.
3. $S = \{(x, y) \in \mathbb{R}^2 : y = 0\}$ is a subspace of $\mathbb{R}^2$ (x-axis).
4. A set of all continuous functions defined on interval $[-1, 1]$ is a subspace of $F_{[-1,1]}$.

For all of the above examples, you should check for yourself that they are in fact subspaces.

Given vectors $v_1, v_2, \ldots, v_n \in V$, we say that $w \in V$ is a **linear combination** of $v_1, v_2, \ldots, v_n$ if for some $r_1, r_2, \ldots, r_n \in \mathbb{R}$, we have $w = r_1 v_1 + r_2 v_2 + \ldots + r_n v_n$. If every vector in V is a linear combination of $S = \{v_1, v_2, \ldots, v_n\}$, we have $\text{span}(S) = V$, then we say S **spans** V.

Some properties of subspaces:

1. Subspaces are closed under linear combinations.
2. A nonempty set S is a subspace if and only if every linear combination of (finitely many) elements of S also belongs to S.

Linear independence and dependence Given vectors $v_1, v_2, \ldots, v_n \in V$ we say that $v_1, v_2, \ldots, v_n$ are **linearly independent** if $r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = 0 \Longrightarrow r_1 = r_2 = \ldots = r_n = 0$, i.e. the only linear combination of $v_1, v_2, \ldots, v_n$ that produces 0 vector is the trivial one. We say that $v_1, v_2, \ldots, v_n$ are **linearly dependent** otherwise.

Theorem: Let $I, S \subset V$ be such that I is linearly independent, and S spans V. Then for every $x \in I$ there exists a $y \in S$ such that $\{y\} \cup I \backslash \{x\}$ is linearly independent.

Proof: This proof will be by contradiction, and use two facts that can be easily verified from the definitions above. First, if $I \subset V$ is linearly independent, then $I \cup \{x\}$ is linearly dependent if and only if (iff) $x \in \text{span}(I)$. Second, if $S, T \subset V$ with $T \subset \text{span}(S)$ then $\text{span}(T) \subset \text{span}(S)$.

If the theorem's claim does not hold. Then there exists a $x \in I$ such that for all $y \in S$ $\{y\} \cup I \backslash \{x\}$ is linearly dependent. Let $I' = I \backslash \{x\}$. By I linearly independent it follows that I' is also linearly independent. Then by the first fact above, $\{y\} \cup I'$ linearly dependent implies $y \in \text{span}(I')$. Moreover, this holds for all $y \in S$ so $S \subset \text{span}(I')$.

By the second fact we then have that $\text{span}(S) \subset \text{span}(I')$. Now since S spans V it follows that $x \in V = \text{span}(S) \subset \text{span}(I') = \text{span}(I \backslash \{x\})$. This means there exists $v_1, v_2, \ldots, v_n \in I \backslash \{x\}$ and $r_1, r_2, \ldots, r_n \in \mathbb{R}$ such that $0 = x - \sum_{i=1}^n r_i v_i$, contradicting I linearly independent. $\square$

Corollary: Let $I, S \subset V$ be such that I is linearly independent, and S spans V. Then $|I| \leq |S|$, where $|\cdot|$ denotes the number of elements of a set (possibly infinite).

Proof: If $|S| = \infty$ then the claim holds by convention, and if $I \subset S$ the claim holds directly. So assume $|S| = m < \infty$, and $I \not\subset S$.

Consider now the following algorithm. Select $x \in I, x \notin S$. By the theorem above, choose a $y \in S$ such that $I' = \{y\} \cup I \backslash \{x\}$ is linearly independent. Note that $|I'| = |I|$ and that $|I' \cap S| > |I \cap S|$. If $I' \subset S$ then the claim holds and stop the algorithm, else continue the algorithm with $I = I'$.

Now note that the above algorithm must terminate in at most $m < \infty$ steps. To see this, first note that after the m^{th} iteration $S \subset I'$. Next, if the algorithm does not terminate at this iteration $I' \not\subset S$, and there would exist a $x \in I', x \notin S$. But then since S spans V there would exist $v_1, v_2 \ldots, v_n \in S \subset I'$ and $r_1, r_2, \ldots, r_n \in \mathbb{R}$ such that $0 = x - \sum_{i=1}^n r_i v_i$ contradicting I' linearly independent. $\square$

Basis and dimension Now suppose that $v_1, v_2, \ldots, v_n$ span V and that, moreover, they are linearly independent. Then we say that the set $\{v_1, v_2, \ldots, v_n\}$ is a **basis** for V.

Theorem: Let S be a basis for V, and let T be another basis for V. Then $|S| = |T|$.

Proof: This follows directly from the above Corollary since S and T are both linearly independent, and both span V. $\square$

We call the unique number of vectors in a basis for V the **dimension** of V (denoted $\dim(V)$).

Examples:

1. $S = \{0\}$ has dimension 0.

2. Any set of vectors that includes 0 vector is linearly dependent (why?)

3. If V has dimension n, and we're given $k < n$ linearly independent vectors in V, then we can extend this set of vectors to a basis.

4. Let $v_1, v_2, \ldots, v_n$ be a basis for V. Then if $v \in V$, $v = r_1v_1 + r_2v_2 + \ldots + r_nv_n$ for some $r_1, r_2, \ldots, r_n \in \mathbb{R}$. Moreover, these coefficients are unique, because if they weren't, we could also write $v = s_1v_1 + s_2v_2 + \ldots + s_nv_n$, and subtracting both sides we get $0 = v - v = (r_1 - s_1)v_1 + (r_2 - s_2)v_2 + \ldots + (r_n - s_n)v_n$, and since the v_i's form basis and are therefore linearly independent, we have $r_i = s_i \ \forall i$, and the coefficients are indeed unique.

5. $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -5 \\ 0 \end{bmatrix}$ both span x-axis, which is the subspace of $\mathbb{R}^2$. Moreover, any one of these two vectors also spans x-axis by itself (thus a basis is not unique, though dimension is), and they are not linearly independent since $5v_1 + 1v_2 = 0$

6. $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ form the standard basis for $\mathbb{R}^3$, since every vector $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in $\mathbb{R}^3$ can be written as $x_1e_1 + x_2e_2 + x_3e_3$, so the three vectors span $\mathbb{R}^3$ and their linear independence is easy to show. In general, $\mathbb{R}^n$ has dimension n.

7. Let $\dim(V) = n$, and let $v_1, v_2, \ldots, v_m \in V$, s.t. $m > n$. Then $v_1, v_2, \ldots, v_m$ are linearly dependent.

2.2 Special Spaces

Inner product space An **inner product** is a function $f : V \times V \to \mathbb{R}$ (which we denote by $f(v_1, v_2) = \langle v_1, v_2 \rangle$), s.t. $\forall\ v, w, z \in V$, and $\forall r \in \mathbb{R}$:

1. $\langle v, w + rz \rangle = \langle v, w \rangle + r\langle v, z \rangle$ (linearity)

2. $\langle v, w \rangle = \langle w, v \rangle$ (symmetry)

3. $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ iff $v = 0$ (positive-definiteness)

We note here that not all vector spaces have inner products defined on them. We call the vector spaces where the inner products are defined the **inner product space**.

Examples:

1. Given 2 vectors $x = [x_1, x_2, \cdots, x_n]'$ and $y = [y_1, y_2, \cdots, y_n]'$ in $\mathbb{R}^n$, we define their inner product $x'y = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$. You can check yourself that the 3 properties above are satisfied, and the meaning of notation $x'y$ will become clear from the next section.

2. Given $f, g \in C_{[-1,1]}$, we define $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$. Once again, verification that this is indeed an inner product is left as an exercise.

Cauchy-Schwarz Inequality: for v and w elements of V, the following inequality holds:

$$\langle v, w \rangle^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle$$

with equality if and only if v and w are linearly dependent.

Proof: Note that $\langle v, 0 \rangle = -\langle v, -0 \rangle = -\langle v, 0 \rangle \Rightarrow \langle v, 0 \rangle = 0, \forall v \in V$.

If $w = 0$, the equality obviously holds.

If $w \neq 0$, let $\lambda = \frac{\langle v, w \rangle}{\langle w, w \rangle}$. Since

$$\begin{aligned} 0 &\leq \langle v - \lambda w, v - \lambda w \rangle \\ &= \langle v, v \rangle - 2\lambda \langle v, w \rangle + \lambda^2 \langle w, w \rangle \\ &= \langle v, v \rangle - \frac{\langle v, w \rangle^2}{\langle w, w \rangle} \end{aligned}$$

□

we can show the result with equality if and only if $v = \lambda w$. Namely, the inequality holds and it's equality if and only if v and w are linearly dependent.

With Cauchy-Schwarz inequality, we can define the **angle** between two nonzero vectors v and w as:

$$angle(v, w) = \arccos \frac{\langle v, w \rangle}{\sqrt{\langle v, v \rangle \cdot \langle w, w \rangle}}$$

The angle is in $[0, \pi)$. This generates nice geometry for the inner product space.

Normed space The **norm**, or **length**, of a vector v in the vector space V is a function $g : V \to \mathbb{R}$ (which we denote by $g(v) = \|v\|$), s.t. $\forall\ v, w \in V$, and $\forall r \in \mathbb{R}$:

1. $\|rv\| = |r|\|v\|$

2. $\|v\| \geq 0$, with equality if and only if $v = 0$

3. $\|v + w\| \leq \|v\| + \|w\|$ (triangle inequality)

Examples:

1. In $\mathbb{R}^n$, let's define the **length of a vector** $x := \|x\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{x'x}$, or $\|x\|^2 = x'x$. This is called the **Euclidian norm**, or the L_2 norm (denote by $\|x\|_2$). (verify it by yourself)

2. Again in $\mathbb{R}^n$, if we define $\|x\| = |x_1| + \ldots + |x_n|$, it's also a norm called the L_1 norm (denote by $\|x\|_1$). (verify it by yourself)

3. Given $f \in C_{[-1,1]}$, we define $\|f\|_p = \left(\int_{-1}^{1} |f(x)|^p dx\right)^{\frac{1}{p}}$, which is also a norm. (see Minkowski Inequality)

4. For any inner product space V, $\|x\|^2 = \langle x, x \rangle$ defines a norm.

Again, not all vector spaces have norms defined in them. For those with defined norms, they are called the **normed spaces**.

In general, we can naturally obtain a norm from a well defined inner product space. Let $\|v\| = \sqrt{\langle v, v \rangle}$ for $\forall v \in V$, where $\langle \cdot, \cdot \rangle$ is the inner product on the space V. It's not hard to verify all the requirements in the definition of norm (verify it by yourself). Thus, for any defined inner product, there is a naturally derived norm. However, in most cases, the opposite (i.e. to obtain inner products from norms) is not obvious.

Metric Space A more general definition on the vector space is the **metric**. The metric is a function $d : V \times V \to \mathbb{R}$ such that for $x, y, z \in V$ it satisfies:

1. $d(x, y) = d(y, x)$

2. $d(x, y) \geq 0$, with equality if and only if $x = y$

3. $d(x, y) \leq d(x, z) + d(y, z)$ (triangle inequality)

A vector space equipped with a metric is called **metric space**. Many analytic definitions (e.g. completeness, compactness, continuity, etc) can be defined under metric space. Please refer to the analysis material for more information.

For any normed space, we can naturally derive a metric as $d(x, y) = \|x - y\|$. This metric is said to be induced by the norm $\|\cdot\|$. However, the opposite is not true. For example, assuming we define the discrete metric on the space V, where $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ if $x \neq y$; it is not obvious what kind of norm should be defined in this space.

If a metric d on a vector space V satisfies the properties: $\forall x, y \in V$ and $\forall r \in \mathbb{R}$,

1. $d(x, y) = d(x + r, y + r)$ (translation invariance)

2. $d(rx, ry) = |r| d(x, y)$ (homogeneity)

then we can define a norm on V by $\|x\| := d(x, 0)$.

To sum up, the relation between the three special spaces is as follows. Given a vector space V, if we define an inner product in it, we can naturally derived a norm in it; if we have a norm in it, we can naturally derived a metric in it. The opposite is not true.

2.3 Orthogonality

We say that vectors v, w in V are **orthogonal** if $\langle v, w\rangle = 0$, or equivalently, $angle(v, w) = \pi/2$. It is denoted as $v \perp w$.

Examples:

1. In $\mathbb{R}^n$ the notion of orthogonality agrees with our usual perception of it. If x is orthogonal to y, then **Pythagorean theorem** tells us that $\|x\|^2 + \|y\|^2 = \|x - y\|^2$. Expanding this in terms of inner products we get:

$$x'x + y'y = (x - y)'(x - y) = x'x - y'x - x'y + y'y \text{ or } 2x'y = 0$$

   and thus $\langle x, y\rangle = x'y = 0$.

2. **Nonzero orthogonal vectors are linearly independent**. Suppose we have $q_1, q_2, \ldots, q_n$, a set of nonzero mutually orthogonal vectors in V, i.e., $\langle q_i, q_j\rangle = 0\ \forall i \neq j$, and suppose that $r_1q_1 + r_2q_2 + \ldots + r_nq_n = 0$. Then taking inner product of q_1 with both sides, we have $r_1\langle q_1, q_1\rangle + r_2\langle q_1, q_2\rangle + \ldots + r_n\langle q_1, q_n\rangle = \langle q_1, 0\rangle = 0$. That reduces to $r_1\|q_1\|^2 = 0$ and since $q_1 \neq 0$, we conclude that $r_1 = 0$. Similarly, $r_i = 0\ \forall\ 1 \leq i \leq n$, and we conclude that $q_1, q_2, \ldots, q_n$ are linearly independent.

3. Suppose we have a $n \times 1$ vector of observations $x = [x_1, x_2, \cdots, x_n]'$. Then if we let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, we can see that vector $e = [x_1 - \bar{x}, x_2 - \bar{x}, \cdots x_n - \bar{x}]'$ is orthogonal to vector $\hat{x} = [\bar{x}, \bar{x}, \cdots, \bar{x}]'$, since $\sum_{i=1}^{n} \bar{x}(x_i - \bar{x}) = \bar{x}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} \bar{x} = n\bar{x}^2 - n\bar{x}^2 = 0$.

Orthogonal subspace and complement Suppose S, T are subspaces of V. Then we say that they are **orthogonal subspaces** if every vector in S is orthogonal to every vector in T. We say that S is the **orthogonal complement** of T in V, if S contains ALL vectors orthogonal to vectors in T and we write $S = T^\perp$.

For example, the x-axis and y-axis are orthogonal subspaces of $\mathbb{R}^3$, but they are not orthogonal complements of each other, since y-axis does not contain $[0, 0, 1]'$, which is perpendicular to every vector in x-axis. However, y-z plane and x-axis ARE orthogonal complements of each other in $\mathbb{R}^3$. You should prove as an exercise that if dim(V) = n, and dim(S) = k, then dim($S^\perp$) = $n - k$.

2.4 Gram-Schmidt Process

Suppose we're given linearly independent vectors $v_1, v_2, \ldots, v_n$ in V, and there's an inner product defined on V. Then we know that $v_1, v_2, \ldots, v_n$ form a basis for the subspace which they span (why?). Then, the **Gram-Schmidt process** can be used to construct an orthogonal basis for this subspace, as follows:

Let $q_1 = v_1$ Suppose v_2 is not orthogonal to v_1. then let rv_1 be the **projection** of v_2 on v_1, i.e. we want to find $r \in \mathbb{R}$ s.t. $q_2 = v_2 - rq_1$ is orthogonal to q_1. Well, we should have $\langle q_1, (v_2 - rq_1)\rangle = 0$, and we get $r = \frac{\langle q_1, v_2\rangle}{\langle q_1, q_1\rangle}$. Notice that the span of q_1, q_2 is the same as the span of v_1, v_2, since all we did was to subtract multiples of original vectors from other original vectors.

Proceeding in similar fashion, we obtain $q_i = v_i - \left(\frac{\langle q_1, v_i\rangle}{\langle q_1, q_1\rangle} q_1 + \ldots + \frac{\langle q_{i-1}, v_i\rangle}{\langle q_{i-1}, q_{i-1}\rangle} q_{i-1}\right)$, and we thus end up with an orthogonal basis for the subspace. If we furthermore divide each of the resulting vectors $q_1, q_2, \ldots, q_n$ by its length, we are left with **orthonormal basis**, i.e. $\langle q_i, q_j\rangle = 0 \ \forall i \neq j$ and $\langle q_i, q_i\rangle = 1, \forall i$ (why?). We call these vectors that have length 1 **unit** vectors.

You can now construct an orthonormal basis for the subspace of $F_{[-1,1]}$ spanned by $f(x) = 1, g(x) = x$, and $h(x) = x^2$ (Exercise 2.6 (b)). An important point to take away is that given any basis for finite-dimensional V, if there's an inner product defined on V, we can always turn the given basis into an orthonormal basis.

Exercises

2.1 Show that the space F_0 of all differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $\frac{df}{dx} = 0$ defines a vector space.

2.2 Verify for yourself that the two conditions for a subspace are independent of each other, by coming up with 2 subsets of $\mathbb{R}^2$: one that is closed under addition and subtraction but NOT under scalar multiplication, and one that is closed under scalar multiplication but NOT under addition/subtraction.

2.3 *Strang, section 3.5 #17b* Let V be the space of all vectors $v = [c_1\, c_2\, c_3\, c_4]' \in \mathbb{R}^4$ with components adding to 0: $c_1 + c_2 + c_3 + c_4 = 0$. Find the dimension and give a basis for V.

2.4 Let $v_1, v_2, ..., v_n$ be a linearly independent set of vectors in V. Prove that if $n = dim(V)$, $v_1, v_2, ..., v_n$ form a basis for V.

2.5 If $F_{[-1,1]}$ is the space of all continuous functions defined on the interval $[-1, 1]$, show that $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx$ defines an inner product of $F_{[-1,1]}$.

2.6 Parts (a) and (b) concern the space $F_{[-1,1]}$, with inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx$.

(a) Show that $f(x) = 1$ and $g(x) = x$ are orthogonal in $F_{[-1,1]}$

(b) Construct an orthonormal basis for the subspace of $F_{[-1,1]}$ spanned by $f(x) = 1, g(x) = x$, and $h(x) = x^2$.

2.7 If a subspace S is contained in a subspace V, prove that $S^\perp$ contains $V^\perp$.