CHAPTER 1

Linear Algebra

1 VECTORS IN $\mathbb{R}^n$

$\mathbb{R}^n$ will denote real Euclidean space of dimension n—that is, $\mathbb{R}^n$ is the set of all ordered n-tuples of real numbers, referred to as "points in $\mathbb{R}^n$" or "vectors in $\mathbb{R}^n$"—the two alternative terminologies "points" and "vectors" will be used quite interchangeably, and we in no way distinguish between them.[1] Depending on the context we may decide to write vectors (i.e., points) in $\mathbb{R}^n$ as columns or rows: in the latter case, $\mathbb{R}^n$ would denote the set of all ordered n-tuples $(x_1, \ldots, x_n)$ with $x_j \in \mathbb{R}$ for each $j = 1, \ldots, n$. In the former case, $\mathbb{R}^n$ would denote the set of all columns $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. In this case we often (for compactness of notation) write $(x_1, \ldots, x_n)^{\mathrm{T}}$ as an alternative notation for the column vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, and conversely $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^{\mathrm{T}} = (x_1, \ldots, x_n)$.

For the moment, we shall write our vectors (i.e., points) as rows. Thus, for the moment, points $\underline{x} \in \mathbb{R}^n$ will be written $\underline{x} = (x_1, \ldots, x_n)$. The real numbers $x_1, \ldots, x_n$ are referred to as *components* of the vector $\underline{x}$. More specifically, x_j is *the j-th component of $\underline{x}$*.

To start with, there are two important operations involving vectors in $\mathbb{R}^n$: we can (i) add them according to the definition

1.1 $$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and (ii) multiply a vector $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$ by a scalar (i.e., by a $\lambda \in \mathbb{R}$) according to the definition

1.2 $$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)\ .$$

Using these two definitions we can prove the following more general property: If $\underline{x} = (x_1, \ldots, x_n)$, $\underline{y} = (y_1, \ldots, y_n)$, and if $\lambda, \mu \in \mathbb{R}$, then

1.3 $$\lambda\underline{x} + \mu\underline{y} = (\lambda x_1 + \mu y_1, \ldots, \lambda x_n + \mu y_n)\ .$$

[1] Nevertheless, from the intuitive point of view it is sometimes helpful to think geometrically of vectors as directed arrows; for example, such a geometric interpretation is suggested in the discussion of *line vectors* below.

Notice, in particular, that, using the definitions 1.1, 1.2, starting with any vector $\underline{x} = (x_1, \dots, x_n)$ we can write

1.4 $$\underline{x} = \sum_{j=1}^{n} x_j \underline{e}_j \ ,$$

where $\underline{e}_j$ is the vector with each component $= 0$ except for the j-th component which is equal to 1. Thus, $\underline{e}_1 = (1, 0, \dots, 0)$, $\underline{e}_2 = (0, 1, 0, \dots, 0), \dots, \underline{e}_n = (0, \dots, 0, 1)$. The vectors $\underline{e}_1, \dots, \underline{e}_n$ are called *standard basis vectors for* $\mathbb{R}^n$.

The definitions 1.1, 1.2 are of course geometrically motivated—the definition of addition is motivated by the desire to ensure that the triangle (or parallelogram) law of addition holds (we'll discuss this below after we introduce the notion of "line vector" joining two points of $\mathbb{R}^n$) and the second is motivated by the natural desire to ensure that $\lambda\underline{x}$ should be a vector in either the same or reverse direction as $\underline{x}$ (according as λ is positive or negative) and that it should have "length" equal to $|\lambda|$ times the original length of $\underline{x}$. To make sense of this latter requirement, we need to first define length of a vector $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$: motivated by[2] Pythagoras, it is natural to define the length of $\underline{x}$ (also referred to as *the distance of* $\underline{x}$ *from* $\underline{0}$), denoted $\|\underline{x}\|$, by

1.5 $$\|\underline{x}\| = \sqrt{\sum_{j=1}^{n} x_j^2} \ .$$

There is also a notion of "line vector" $\overrightarrow{AB}$ from one point $A = \underline{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ to another point $B = \underline{b} = (b_1, \dots, b_n)$ defined by

1.6 $$\overrightarrow{AB} = (b_1 - a_1, \dots, b_n - a_n) \quad (= \underline{b} - \underline{a} \text{ using the definition 1.3 }) \ .$$

Notice, that geometrically, we may think of $\overrightarrow{AB}$ as a directed line segment (or arrow) with tail at A and head at B, but mathematically, $\overrightarrow{AB}$ is nothing but $\underline{b} - \underline{a}$, i.e., the point with coordinates $(b_1 - a_1, \dots, b_n - a_n)$. By definition, we then do have the triangle identity for addition:

1.7 $$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC} \ ,$$

because if $A = \underline{a} = (a_1, \dots, a_n)$, $B = \underline{b} = (b_1, \dots, b_n)$, $C = \underline{c} = (c_1, \dots, c_n)$ then $\overrightarrow{AB} + \overrightarrow{BC} = (\underline{b} - \underline{a}) + (\underline{c} - \underline{b}) = \underline{c} - \underline{a} = \overrightarrow{AC}$.

[2] There is often confusion here in elementary texts, which are apt to claim, at least in the case $n = 2, 3$, that 1.5 is a consequence of Pythagoras, i.e., that 1.5 is proved by using Pythagoras' theorem; that is not the case—1.5 is *a definition, motivated by the desire to ensure that the Pythagorean theorem holds* when we later give the appropriate definition of angle between vectors.

2 DOT PRODUCT AND ANGLE BETWEEN VECTORS IN R^n

If $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$, we define the dot product of $\underline{x}$ and $\underline{y}$, denoted by $\underline{x} \cdot \underline{y}$, by $\underline{x} \cdot \underline{y} = \sum_{j=1}^{n} x_j y_j$. Notice that then the dot product has the properties:

(i) $$\underline{x} \cdot \underline{y} = \underline{y} \cdot \underline{x}$$

(ii) $$(\mu \underline{x} + \lambda \underline{y}) \cdot \underline{z} = \mu \underline{x} \cdot \underline{z} + \lambda \underline{y} \cdot \underline{z}$$

(iii) $$\underline{x} \cdot \underline{x} = \|\underline{x}\|^2 .$$

Using the above properties, we can check the identity

2.1 $$\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\underline{x} \cdot \underline{y}$$

as follows:

$$\begin{aligned} \|\underline{x} + \underline{y}\|^2 &= (\underline{x} + \underline{y}) \cdot (\underline{x} + \underline{y}) && \text{(by (iii))} \\ &= \underline{x} \cdot \underline{x} + \underline{y} \cdot \underline{y} + 2\underline{x} \cdot \underline{y} && \text{(by several applications of (i) and (ii))} \\ &= \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\underline{x} \cdot \underline{y} && \text{(by (iii) again)} . \end{aligned}$$

In particular, with $t\underline{y}$ in place of $\underline{y}$ (where $t \in \mathbb{R}$) we see that

2.2 $$\|\underline{x} + t\underline{y}\|^2 = t^2\|\underline{y}\|^2 + 2t\underline{x} \cdot \underline{y} + \|\underline{x}\|^2 .$$

Observe that if $\underline{y} \neq \underline{0}$ this is a quadratic in the variable t, and by using "completion of the square" $at^2 + 2bt + c = a(t + b/a)^2 + (ac - b^2)/a$ (valid if $a \neq 0$), we see that 2.2 can be written as:

2.3 $$\|\underline{x} + t\underline{y}\|^2 = \|\underline{y}\|^2(t + \underline{x} \cdot \underline{y}/\|\underline{y}\|)^2 + (\|\underline{y}\|^2\|\underline{x}\|^2 - (\underline{x} \cdot \underline{y})^2)/\|\underline{y}\| .$$

Observe that the right side here evidently takes its minimum value when, and only when, $t = -\underline{x} \cdot \underline{y}/\|\underline{y}\|$ and the corresponding minimum value of $\|\underline{x} + t\underline{y}\|^2$ is $(\|\underline{y}\|^2\|\underline{x}\|^2 - (\underline{x} \cdot \underline{y})^2)/\|\underline{y}\|$ and the corresponding value of $\underline{x} + t\underline{y}$ is the point $\underline{x} - \|\underline{y}\|^{-2}(\underline{x} \cdot \underline{y})\underline{y}$. Thus,

2.4 $$\|\underline{x} + t\underline{y}\|^2 \geq (\|\underline{y}\|^2\|\underline{x}\|^2 - (\underline{x} \cdot \underline{y})^2)/\|\underline{y}\| ,$$

with equality if and only if $t = -\underline{x} \cdot \underline{y}/\|\underline{y}\|$, in which case $\underline{x} + t\underline{y} = \underline{x} - \|\underline{y}\|^{-2}(\underline{x} \cdot \underline{y})\underline{y}$. We can give a geometric interpretation of 2.4 as follows.

For $\underline{y} \neq \underline{0}$, the straight line ℓ through $\underline{x}$ parallel to $\underline{y}$ is by definition the set of points $\{\underline{x} + t\underline{y} : t \in \mathbb{R}\}$, so 2.4 says geometrically that the point $\underline{x} - \|\underline{y}\|^{-2}(\underline{x} \cdot \underline{y})\underline{y}$ is the point of ℓ which has least distance to the origin, and the least distance is equal to $\sqrt{(\|\underline{y}\|^2\|\underline{x}\|^2 - (\underline{x} \cdot \underline{y})^2)/\|\underline{y}\|}$.

Observe that an important particular consequence of the above discussion is that $\|\underline{y}\|^2\|\underline{x}\|^2 - (\underline{x} \cdot \underline{y})^2 \geq 0$, i.e.,

2.5 $$|\underline{x} \cdot \underline{y}| \leq \|\underline{x}\| \, \|\underline{y}\| \quad \forall \ \underline{x}, \underline{y} \in \mathbb{R}^n ,$$

and in the case $\underline{y} \neq 0$ equality holds in 2.5 if and only if $\underline{x} = \|\underline{y}\|^{-2}(\underline{x} \cdot \underline{y})\underline{y}$. (We technically proved 2.5 only in case $\underline{y} \neq \underline{0}$, but observe that 2.5 is also true, and equality holds, when $\underline{y} = \underline{0}$.) The inequality 2.5 is known as the Cauchy-Schwarz inequality.

An important corollary of 2.5 is the following "triangle inequality" for vectors in $\mathbb{R}^n$:

2.6 $$\underline{x}, \underline{y} \in \mathbb{R}^n \Rightarrow \|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\| \,.$$

This is proved as follows. By 2.1 and 2.5, $\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\underline{x} \cdot \underline{y} \leq \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\|\underline{x}\|\|\underline{y}\| = (\|\underline{x}\| + \|\underline{y}\|)^2$ and then 2.6 follows by taking square roots.

Now we want to give the definition of the angle between two nonzero vectors $\underline{x}, \underline{y}$. As a preliminary we recall that the function $\cos\theta$ is a 2π-periodic function which has maximum value 1, attained at $\theta = k\pi$, k any even integer, and minimum value -1 attained at $\theta = k\pi$ with k any odd integer. Also $\cos\theta$ is strictly decreasing from 1 to -1 as θ varies between 0 and π. Thus, for each value $t \in [-1, 1]$ there is a unique value of $\theta \in [0, \pi]$ (denoted $\arccos t$) with $\cos\theta = t$. We will give a proper definition of the function $\cos\theta$ and discuss these properties in detail later (in Lecture 6 of Appendix A), but for the moment we just assume them without further discussion. We are now ready to give the formal definition of the angle between two nonzero vectors. So suppose $\underline{x}, \underline{y} \in \mathbb{R}^n \setminus \{0\}$. Then we define[3] the angle θ between $\underline{x}, \underline{y}$ by

2.7 $$\theta = \arccos\bigl(\|x\|^{-1}\underline{x} \cdot \|\underline{y}\|^{-1}\underline{y}\bigr) \,.$$

Notice this makes sense because $\|x\|^{-1}\underline{x} \cdot \|\underline{y}\|^{-1}\underline{y} \in [-1, 1]$ by the Cauchy-Schwarz inequality 2.5, which says precisely that $-\|\underline{x}\|\|\underline{y}\| \leq \underline{x} \cdot \underline{y} \leq \|\underline{x}\|\|\underline{y}\|$.

Observe that by definition we then have

2.8 $$\underline{x} \cdot \underline{y} = \|\underline{x}\|\|\underline{y}\| \cos\theta, \quad \underline{x}, \underline{y} \in \mathbb{R}^n \setminus \{0\} \,,$$

where θ is the angle between $\underline{x}, \underline{y}$.

2.9 Definition: Two vectors $\underline{x}, \underline{y}$ are said to be orthogonal if $\underline{x} \cdot \underline{y} = 0$.

Observe that then the zero vector is orthogonal to every other vector, and, by 2.8, two nonzero vectors $\underline{x}, \underline{y} \in \mathbb{R}^n$ are orthogonal if and only if the angle between them is $\pi/2$ (because $\cos\theta = 0$ with $\theta \in [0, \pi]$ if and only if $\theta = \pi/2$).

SECTION 2 EXERCISES

2.1 Explain why the following "proof" of the Cauchy-Schwarz inequality $|\underline{x} \cdot \underline{y}| \leq \|\underline{x}\|\,\|\underline{y}\|$, where $\underline{x}, \underline{y} \in \mathbb{R}^n$, is not valid:

[3] Again, there is often confusion in elementary texts about what is a definition and what is a theorem. In the present treatment, 2.7 is a *definition* which relies only on the Cauchy-Schwarz inequality 2.5 (which ensures $\|\underline{x}\|^{-1}\|\underline{y}\|^{-1}\underline{x} \cdot \underline{y} \in [-1, 1]$) and the fact that arccos is a well defined function on $[-1, 1]$ (with values $\in [0, \pi]$). With this approach it becomes a theorem rather than a definition that $\cos\theta$ is the ratio (length of the edge adjacent to the angle θ)/(length of hypotenuse) in a right triangle—see problem 6.5 of real analysis Lecture 6 in the Appendix for further discussion.

Proof: If either $\underline{x}$ or $\underline{y}$ is zero, then the inequality $|\underline{x} \cdot \underline{y}| \le \|\underline{x}\|\,\|\underline{y}\|$ is trivially correct because both sides are zero. If neither $\underline{x}$ nor $\underline{y}$ is zero, then as proved above $\underline{x} \cdot \underline{y} = \|\underline{x}\|\,\|\underline{y}\| \cos\theta$, where θ is the angle between $\underline{x}$ and $\underline{y}$. Hence, $|\underline{x} \cdot \underline{y}| = \|\underline{x}\|\,\|\underline{y}\|\,|\cos\theta| \le \|\underline{x}\|\,\|\underline{y}\|$.

Note: We shall give a detailed discussion of the function $\cos x$ later (in Lecture 6 of the Appendix); for the moment you should of course assume that $\cos x$ is well defined and has its usual properties (e.g., it is 2π-periodic, has absolute value ≤ 1 and takes each value in $[-1, 1]$ exactly once if we restrict x to the interval $[0, \pi]$, etc.). Thus, $|\cos\theta| \le 1$ (used in the last step above) is certainly correct.

2.2 (a) The Cauchy-Schwarz inequality proved above guarantees $|\underline{x} \cdot \underline{y}| \le \|\underline{x}\|\|\underline{y}\|$ and hence $\underline{x} \cdot \underline{y} \le \|\underline{x}\|\|\underline{y}\|$ for all $\underline{x}, \underline{y} \in \mathbb{R}^n$. If $\underline{y} \ne \underline{0}$, prove that equality holds in the first inequality if and only if $\underline{x} = \lambda\underline{y}$ for some $\lambda \in \mathbb{R}$ and equality holds in the second inequality if and only if $\underline{x} = \lambda\underline{y}$ for some $\lambda \ge 0$.

(b) By examining the proof of the triangle inequality $\|\underline{x} + \underline{y}\| \le \|\underline{x}\| + \|\underline{y}\|$ given above (recall that proof began with the identity $\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 + 2\underline{x} \cdot \underline{y}$), prove that *equality* holds in the triangle inequality $\iff$ either at least one of $\underline{x}, \underline{y}$ is $\underline{0}$ or $\underline{x}, \underline{y} \ne \underline{0}$ and $\underline{y} = \lambda\underline{x}$ with $\lambda > 0$.

2.3 (Another proof of the Cauchy-Schwarz inequality.) If $\underline{a} = (a_1, \ldots, a_n), \underline{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$, prove the identity $\frac{1}{2}\sum_{i,j=1}^{n}(a_i b_j - a_j b_i)^2 = \|\underline{a}\|^2\|\underline{b}\|^2 - (\underline{a} \cdot \underline{b})^2$, and hence prove $|\underline{a} \cdot \underline{b}| \le \|\underline{a}\|\|\underline{b}\|$.

2.4 Using the dot product, prove, for any vectors $\underline{x}, \underline{y} \in \mathbb{R}^n$:

(a) The parallelogram law: $\|\underline{x} - \underline{y}\|^2 + \|\underline{x} + \underline{y}\|^2 = 2(\|x\|^2 + \|y\|^2)$.

(b) The law of cosines: $\|\underline{x} - \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 - 2\|\underline{x}\|\,\|\underline{y}\| \cos\theta$, assuming $\underline{x}, \underline{y}$ are nonzero and θ is the angle between $\underline{x}$ and $\underline{y}$.

(c) Give a geometric interpretation of identities (a),(b) (i.e., describe what (a) is saying about the parallelogram determined by $\underline{x}, \underline{y}$—i.e., $OACB$ where $\overrightarrow{OA} = \underline{x}$, $\overrightarrow{OB} = \underline{y}$, $\overrightarrow{OC} = \underline{x} + \underline{y}$, and what (b) is saying about the triangle determined by $\underline{x}$ and $\underline{y}$—i.e., OAB, where $\overrightarrow{OA} = \underline{x}$, $\overrightarrow{OB} = \underline{y}$).

3 SUBSPACES AND LINEAR DEPENDENCE OF VECTORS

We say that a subset V of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ (or more specifically a *linear subspace* of $\mathbb{R}^n$) if V has the properties that it contains at least the zero vector $\underline{0} = (0, \ldots, 0)$, i.e.,

3.1 $$\underline{0} \in V$$

and if it is *closed under addition and multiplication by scalars*, i.e.,

3.2 $$\underline{x}, \underline{y} \in V \text{ and } \lambda, \mu \in \mathbb{R} \Rightarrow \lambda\underline{x} + \mu\underline{y} \in V\,.$$

For example, the subset V consisting of just the zero vector (i.e., $V = \{\underline{0}\}$) is a subspace (usually referred to as the *trivial subspace*) and the subset V consisting of all vectors in $\mathbb{R}^n$ (i.e., $V = \mathbb{R}^n$) is a subspace.