1.3 Dot Product, Angles, and Orthogonal Projection

The dot product on $\mathbb{R}^n$ is an easy-to-calculate operation that you perform on pairs of vectors and which gives you back a real number, not a vector. The dot product is important because, in 2 and 3 dimensions, the dot product gives us an easy way of computing the angle between vectors. In higher dimensions, the dot product is used to define the angle between two vectors. A fundamental application of the dot product is in calculating the work done by a constant force as an object undergoes a displacement.

The dot product also arises when dealing with orthogonal projection. Orthogonal projection gives us a simple way to decompose a vector into a sum of two vectors, one of which is parallel to a prescribed vector **b** and the other of which is perpendicular, or orthogonal, to **b**.

Basics:

If we have two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, and we draw them as starting at the same point, then we may talk about the *angle between the vectors*, which could be anywhere from 0 degrees or radians, when the vectors point in the same direction, to 180 degrees, or π radians, when the vectors point in opposite directions.

Given two vectors in coordinates, we would like to find an easy way to calculate the angle between them.

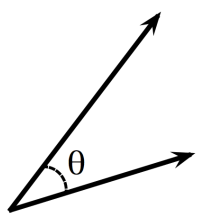

Figure 1.3.1: An acute angle between vectors.

Consider two vectors in $\mathbb{R}^2$, $\mathbf{a} = (x_1, y_1)$ and $\mathbf{b} = (x_2, y_2)$. We draw these vectors based at the origin, and we let $\mathbf{c} = \mathbf{b} - \mathbf{a}$; see Figure 1.3.3.

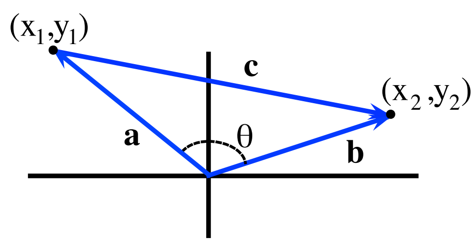

Figure 1.3.3: Vectors **a** and **b**, and the angle between them.

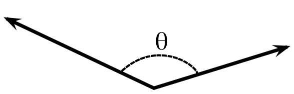

Figure 1.3.2: An obtuse angle between vectors.

Recall the *Law of Cosines* from trigonometry, which, in terms of our vectors, says

$$|\mathbf{c}|^2 \;=\; |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta.$$

Noting that

$$|\mathbf{a}|^2 = x_1^2 + y_1^2, \qquad |\mathbf{b}|^2 = x_2^2 + y_2^2, \qquad \text{and} \qquad |\mathbf{c}|^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2,$$

we find that the Law of Cosines becomes

$$x_2^2 - 2x_1x_2 + x_1^2 + y_2^2 - 2y_1y_2 + y_1^2 \;=\; x_1^2 + y_1^2 + x_2^2 + y_2^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta.$$

Subtracting $x_1^2 + y_1^2 + x_2^2 + y_2^2$ from both sides, and dividing by -2, we arrive at

$$x_1x_2 + y_1y_2 \;=\; |\mathbf{a}||\mathbf{b}|\cos\theta.$$

Note how simple the left-hand side of the above formula is; it's just the product of the x-coordinates of the vectors **a** and **b** added to the product of the y-coordinates.

You can perform the analogous calculations for two vectors $\mathbf{p} = (x_1, y_1, z_1)$ and $\mathbf{q} = (x_2, y_2, z_2)$ in $\mathbb{R}^3$, and what you find is

$$x_1x_2 + y_1y_2 + z_1z_2 \;=\; |\mathbf{p}||\mathbf{q}|\cos\theta,$$

where θ is the angle between **p** and **q**. Again, the left-hand side is simply the sum of the product of the corresponding coordinates of the two vectors.

This motivates us to define:

Definition 1.3.1. *Let* $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ *and* $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ *be vectors in* $\mathbb{R}^n$*. Then, the* **dot product** $\mathbf{v}\cdot\mathbf{w}$ *of* **v** *and* **w** *is the real number given by adding together the product to the corresponding coordinates of the two vectors, i.e.,*

$$\mathbf{v}\cdot\mathbf{w} \;=\; v_1w_1 + v_2w_2 + \cdots + v_nw_n.$$

It is a common, but horrible, mistake to think that the dot product of two vectors yields another vector. You add together the products of the corresponding coordinates, so **you end up with a number, a scalar, not a vector**.

The important properties of the dot product, which we shall use throughout the remainder of this book are:

Theorem 1.3.2. *Let* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *be vectors in* $\mathbb{R}^n$, *and let* r *and* s *be real numbers. Then,*

1. (**commutativity**) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$;
2. (**distributivity**) $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \cdot \mathbf{b}) + (\mathbf{a} \cdot \mathbf{c})$;
3. (**scalar extraction**) $(r\mathbf{a}) \cdot (s\mathbf{b}) = (rs)(\mathbf{a} \cdot \mathbf{b})$;
4. $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$;
5. (**Cauchy-Schwarz Inequality**) *the absolute value of* $\mathbf{a} \cdot \mathbf{b}$ *satisfies*
$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|,$$
and the equality holds if and only if $\mathbf{a}$ *and* $\mathbf{b}$ *are parallel; and*
6. *the dot product is related to the angles between vectors by*
$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta,$$
where θ *is the angle between the vectors* $\mathbf{a}$ *and* $\mathbf{b}$.

Note that, if one of the vectors in the dot product is the zero vector, then there is no "the" angle between the vectors, because we allow the zero vector to have **every** direction. Still, we go ahead and write that $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$, since if $\mathbf{a}$ or $\mathbf{b}$ equals $\mathbf{0}$, the equality holds for all θ.

Remark 1.3.3. We leave the verification of properties 1-4 of Theorem 1.3.2 as exercises; they follow easily from the definition and corresponding properties of real numbers.

We will prove the Cauchy-Schwarz Inequality in the More Depth portion of this section.

We derived the formula $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$ for vectors in $\mathbb{R}^2$; the proof in $\mathbb{R}^3$ is essentially identical. What happens in $\mathbb{R}^n$, where $n \geq 4$?

The answer may seem like cheating. If $\mathbf{a}$ and $\mathbf{b}$ are non-zero vectors in $\mathbb{R}^n$, then the Cauchy-Schwarz Inequality tells us that

$$-1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} \leq 1,$$

and we **define** the angle between $\mathbf{a}$ and $\mathbf{b}$ to be

$$\theta = \cos^{-1}\left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}\right).$$

Before we state an important corollary of Theorem 1.3.2, we should mention that there are three different terms which are all used to mean the same thing: *perpendicular*, *orthogonal*, and *normal*. For each of these synonyms, there are contexts in which one is the classically preferred term; you shall see this, and hopefully get used to it, throughout the textbook. In addition, an angle of $90^\circ = \pi/2$ radians is, as you probably know, called a *right angle*.

Corollary 1.3.4. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be vectors in* $\mathbb{R}^n$*. Then,* $\mathbf{a}$ *and* $\mathbf{b}$ *are perpendicular (or, orthogonal, or, normal) to each other if and only if*

$$\mathbf{a} \cdot \mathbf{b} = 0.$$

The angle θ *between non-zero vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is acute, i.e.,* $0 \leq \theta < \pi/2$ *radians, if and only if* $\mathbf{a} \cdot \mathbf{b} > 0$*. The angle* θ *between non-zero vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is obtuse, i.e.,* $\pi/2 < \theta \leq \pi$ *radians, if and only if* $\mathbf{a} \cdot \mathbf{b} < 0$*.*

Example 1.3.5. Consider the vectors

$$\mathbf{a} = (1, 2, 3), \qquad \mathbf{b} = (-2, -2, 2), \qquad \text{and} \qquad \mathbf{c} = (0, 4, 1).$$

a) For each pair of vectors, decide if the angle between them is right, acute, or obtuse.

b) For the acute and obtuse angles from part (a), determine the actual angle between the vectors.

Solution:

Part (a) is simple; you calculate the dot products and see whether you get 0, something positive, or something negative.

We find

$$\mathbf{a} \cdot \mathbf{b} \;= (1, 2, 3) \cdot (-2, -2, 2) \;=\; (1)(-2) + (2)(-2) + (3)(2) \;=\; 0,$$

so that the angle between $\mathbf{a}$ and $\mathbf{b}$ is a right angle, i.e. $\mathbf{a}$ and $\mathbf{b}$ are orthogonal;

$$\mathbf{a} \cdot \mathbf{c} \;= (1, 2, 3) \cdot (0, 4, 1) \;=\; (1)(0) + (2)(4) + (3)(1) \;=\; 11 \;>\; 0,$$

so that the angle between $\mathbf{a}$ and $\mathbf{c}$ is acute; and

$$\mathbf{b} \cdot \mathbf{c} \;= (-2, -2, 2) \cdot (0, 4, 1) \;=\; (-2)(0) + (-2)(4) + (2)(1) \;=\; -6 \;<\; 0,$$

so that the angle between $\mathbf{b}$ and $\mathbf{c}$ is obtuse.

Now, we want to find the actual angles between **a** and **c**, and between **b** and **c**. First, we calculate

$$|\mathbf{a}| = \sqrt{1^2+2^2+3^2} = \sqrt{14}, \qquad |\mathbf{b}| = |2(-1,-1,1)| = 2\sqrt{3}, \qquad \text{and} \qquad |\mathbf{c}| = \sqrt{17}.$$

Therefore, we find

$$\theta_{\mathbf{a},\mathbf{c}} \;=\; \cos^{-1}\left(\frac{\mathbf{a}\cdot\mathbf{c}}{|\mathbf{a}||\mathbf{c}|}\right) \;=\; \cos^{-1}\left(\frac{11}{\sqrt{14}\sqrt{17}}\right) \;\approx\; 0.777 \;\text{ radians } \;\approx\; 44.52^\circ,$$

and

$$\theta_{\mathbf{b},\mathbf{c}} \;=\; \cos^{-1}\left(\frac{\mathbf{b}\cdot\mathbf{c}}{|\mathbf{b}||\mathbf{c}|}\right) \;=\; \cos^{-1}\left(\frac{-6}{2\sqrt{3}\sqrt{17}}\right) \;\approx\; 2.004 \;\text{ radians } \;\approx\; 114.8^\circ.$$

Remark 1.3.6. Note that the standard basis vectors (recall Definition 1.2.21) in $\mathbb{R}^n$ are pairwise-orthogonal. This is easy to see; in $\mathbb{R}^2$,

$$\mathbf{i}\cdot\mathbf{j} = (1,0)\cdot(0,1) = 0.$$

In $\mathbb{R}^3$, it's just as easy:

$$\mathbf{i}\cdot\mathbf{j} = (1,0,0)\cdot(0,1,0) = 0, \quad \mathbf{i}\cdot\mathbf{k} = (1,0,0)\cdot(0,0,1) = 0, \quad \text{and} \quad \mathbf{j}\cdot\mathbf{k} = (0,1,0)\cdot(0,0,1) = 0.$$

More generally, in $\mathbb{R}^n$, $\mathbf{e}_l \cdot \mathbf{e}_m = 0$, if $l \neq m$.

In fact, the standard basis is what's known as an *orthonormal basis*, which means that the vectors are pairwise-orthogonal **and** that each basis vector is a unit vector, which, in terms of the dot product, means that $|\mathbf{e}_m| = \sqrt{\mathbf{e}_m\cdot\mathbf{e}_m} = 1$ or, simply, $\mathbf{e}_m\cdot\mathbf{e}_m = 1$.

Example 1.3.7. Calculate the dot product

$$(3\mathbf{i}-\mathbf{j}+2\mathbf{k})\cdot(\mathbf{i}+\mathbf{j}).$$

Solution:

Of course, we can eliminate the explicit references to **i**, **j**, and **k**, and calculate

$$(3\mathbf{i}-\mathbf{j}+2\mathbf{k})\cdot(\mathbf{i}+\mathbf{j}) \;=\; (3,-1,2)\cdot(1,1,0) \;=\; (3)(1)+(-1)(1)+(2)(0) \;= 3-1+0 = 2.$$

However, it's instructive to use the algebraic properties of the dot product, along with the fact that the standard basis is orthonormal:

$$(3\mathbf{i}-\mathbf{j}+2\mathbf{k})\cdot(\mathbf{i}+\mathbf{j}) \;=\; 3(\mathbf{i}\cdot\mathbf{i})+3(\mathbf{i}\cdot\mathbf{j})-(\mathbf{j}\cdot\mathbf{i})-(\mathbf{j}\cdot\mathbf{j})+2(\mathbf{k}\cdot\mathbf{i})+2(\mathbf{k}\cdot\mathbf{j}) \;=$$

$$3+0-0-1+0+0 \;=\; 2.$$

Example 1.3.8. Sometimes, given a non-zero vector (a, b) in $\mathbb{R}^2$, it is desirable to produce a non-zero vector which is perpendicular to (a, b). How do you do this? It's easy.

Just swap a and b and negate one them. That is, take the vector $(b, -a)$ or $(-b, a)$. It is trivial to verify that the dot product of either of these with (a, b) is 0:

$$(a, b) \cdot (b, -a) \;=\; ab + (b)(-a) \;=\; 0 \qquad \text{and} \qquad (a, b) \cdot (-b, a) \;=\; (a)(-b) + ba \;=\; 0.$$

In many physical problems, you are given a vector $\mathbf{F}$, and a non-zero vector $\mathbf{v}$, and you want to consider the "part of $\mathbf{F}$ that is parallel to $\mathbf{v}$". What does this mean?

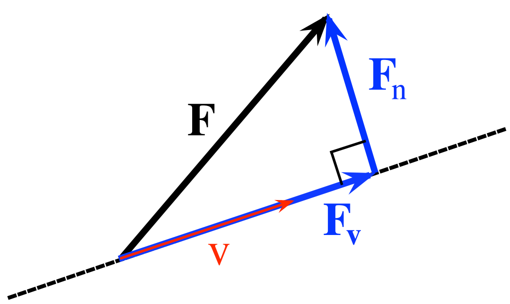

Figure 1.3.4: Writing $\mathbf{F}$ as a sum of vectors parallel and normal to $\mathbf{v}$.

What we mean is that we want to write $\mathbf{F}$ as the sum of two vectors $\mathbf{F_v}$ and $\mathbf{F}_n$, where $\mathbf{F_v}$ is a scalar multiple of $\mathbf{v}$ and $\mathbf{F}_n$ is orthogonal, or normal, to $\mathbf{v}$. Thus, we want

$$\mathbf{F} \;=\; \mathbf{F_v} + \mathbf{F}_n \;=\; t\mathbf{v} + \mathbf{F}_n,$$

for some t, where $\mathbf{F}_n$ is orthogonal to $\mathbf{v}$.

How do we determine t? We take the dot product of both sides of the previous equation with $\mathbf{v}$ and use that we are requiring that $\mathbf{F}_n$ is orthogonal to $\mathbf{v}$, so that $\mathbf{F}_n \cdot \mathbf{v} = 0$; we find

$$\mathbf{F} \cdot \mathbf{v} \;=\; t(\mathbf{v} \cdot \mathbf{v}) \;+\; 0.$$

Therefore, t would have to equal $(\mathbf{F} \cdot \mathbf{v})/(\mathbf{v} \cdot \mathbf{v})$, and so, if there exist $\mathbf{F_v}$ and $\mathbf{F}_n$ with the desired properties, we must have that

$$\mathbf{F_v} \;=\; t\mathbf{v} \;=\; \left(\frac{\mathbf{F} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\right)\mathbf{v} \;=\; \left(\frac{\mathbf{F} \cdot \mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v},$$

and

$$\mathbf{F}_n \;=\; \mathbf{F} \;-\; \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v}.$$

Moreover, it is easy to check that these $\mathbf{F}_\mathbf{v}$ and $\mathbf{F}_n$ do, indeed, satisfy the properties that we wanted; for clearly, $\mathbf{F} = \mathbf{F}_\mathbf{v} + \mathbf{F}_n$, $\mathbf{F}_\mathbf{v}$ is parallel to $\mathbf{v}$, and $\mathbf{F}_n$ is orthogonal to $\mathbf{v}$ because

$$\left(\mathbf{F} \;-\; \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v}\right)\cdot\mathbf{v} \;=\; \mathbf{F}\cdot\mathbf{v} - \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)(\mathbf{v}\cdot\mathbf{v}) \;=\; \mathbf{F}\cdot\mathbf{v} - \mathbf{F}\cdot\mathbf{v} \;=\; 0.$$

We give names to $\mathbf{F}_\mathbf{v}$ and $\mathbf{F}_n$.

Definition 1.3.9. *Given a vector* $\mathbf{F}$ *and a non-zero vector* $\mathbf{v}$*, both in* $\mathbb{R}^n$*, we define the* **orthogonal projection of F onto v** *to be the vector*

$$\mathbf{F}_\mathbf{v} \;=\; \mathrm{proj}_\mathbf{v}\mathbf{F} \;=\; \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v} \;=\; \left(\mathbf{F}\cdot\frac{\mathbf{v}}{|\mathbf{v}|}\right)\frac{\mathbf{v}}{|\mathbf{v}|}.$$

This is also referred to as the **component of F, parallel to v**.

In this context, the vector $\mathbf{F}_n = \mathbf{F} - \mathbf{F}_\mathbf{v}$ *is referred to as the* **component of F, normal to v**

Remark 1.3.10. We shall usually use the notation $\mathbf{F}_\mathbf{v}$ for the orthogonal projection. However, the notation $\mathrm{proj}_\mathbf{v}\mathbf{F}$ is better if you're going to project multiple vectors onto $\mathbf{v}$, and so want to discuss the *orthogonal projection function* $\mathrm{proj}_\mathbf{v} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\mathrm{proj}_\mathbf{v}(\mathbf{F}) \;=\; \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v}.$$

The vector $\mathbf{F}_n$, the component of $\mathbf{F}$ normal to $\mathbf{v}$ that we discussed earlier, is not usually given special notation or a special name; you simply write $\mathbf{F}-\mathbf{F}_\mathbf{v}$ for this normal component.

Note that the vector $\mathbf{u} \;=\; \mathbf{v}/|\mathbf{v}|$ is the unit vector in the direction of $\mathbf{v}$, so that

$$\mathbf{F}_\mathbf{v} \;=\; \left(\mathbf{F}\cdot\frac{\mathbf{v}}{|\mathbf{v}|}\right)\frac{\mathbf{v}}{|\mathbf{v}|} \;=\; (\mathbf{F}\cdot\mathbf{u})\mathbf{u} \;=\; \mathbf{F}_\mathbf{u},$$

which makes it clear that only the direction of $\mathbf{v}$ matters when calculating the orthogonal projection. In fact,

$$\mathbf{F}_{-\mathbf{v}} \;=\; \mathbf{F}_{-\mathbf{u}} \;=\; (\mathbf{F}\cdot -\mathbf{u})(-\mathbf{u}) \;=\; (\mathbf{F}\cdot\mathbf{u})\mathbf{u} \;=\; \mathbf{F}_\mathbf{u} \;=\; \mathbf{F}_\mathbf{v}$$

and, thus, you get the same orthogonal projection if you project onto parallel vectors. Of course, this makes perfect sense from our original discussion, where we simply required that $\mathbf{F_v}$ be parallel to $\mathbf{v}$.

Finally, when $\mathbf{u}$ is a unit vector, the formula

$$\mathbf{F_u} \;=\; (\mathbf{F}\cdot\mathbf{u})\mathbf{u}$$

is so simple to remember that many people begin the problem of calculating $\mathbf{F_v}$ by first producing $\mathbf{u} = \mathbf{v}/|\mathbf{v}|$, and then using that

$$\mathbf{F_v} \;=\; \mathbf{F_u} \;=\; (\mathbf{F}\cdot\mathbf{u})\mathbf{u}.$$

Example 1.3.11. Suppose that $\mathbf{F} = (1, 2, 5)$. Find the component of $\mathbf{F}$ parallel to $\mathbf{v} = (-1, 0, 1)$, i.e., calculate the orthogonal projection of $(1, 2, 5)$ onto $(-1, 0, 1)$. Also, write $\mathbf{F}$ as the sum of two vectors, one parallel to $\mathbf{v}$ and one orthogonal/normal to $\mathbf{v}$.

Solution:

We find:

$$\mathbf{F_v} \;=\; \left(\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\right)\mathbf{v} \;=\; \left(\frac{(1,2,5)\cdot(-1,0,1)}{|(-1,0,1)|^2}\right)(-1,0,1) \;=\; \frac{4}{2}(-1,0,1) \;=\; 2(-1,0,1).$$

The component of $\mathbf{F}$, normal to $\mathbf{v}$, $\mathbf{F}_n$ is simply the difference

$$\mathbf{F}_n \;=\; \mathbf{F} - \mathbf{F_v} \;=\; (1,2,5) - (-2,0,2) \;=\; (3,2,3).$$

Thus,

$$(1,2,5) \;=\; 2(-1,0,1) \;+\; (3,2,3)$$

is the required decomposition.

Example 1.3.12. In physics and engineering problems, a vector $\mathbf{F}$ is frequently specified, not by giving components, but rather by giving the magnitude and the angle made (in a given plane) with respect to a fixed line or straight object. See Figure 1.3.5.

In this case, the orthogonal projection of $\mathbf{F}$ onto the line or object is denoted in some intuitive way, like $\mathbf{F}_{\text{rod}}$ in Figure 1.3.5, and isn't computed by explicitly using the dot product in its coordinate form.

Consider, for instance, a force of 12 Newtons acting at a 30° angle at one end of a metal rod; referring to Figure 1.3.5, this means that $|\mathbf{F}| = 12$ Newtons and $\theta = 30°$.

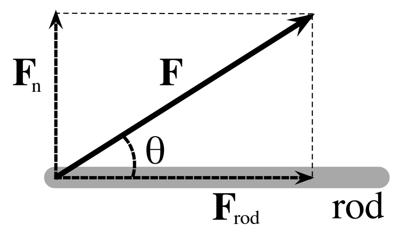

Figure 1.3.5: The components of **F** parallel and normal to a straight rod.

Then, the magnitude of $\mathbf{F}_{\text{rod}}$, the component of **F** which is parallel to the rod, is obtained simply from trigonometry:

$$|\mathbf{F}_{\text{rod}}| \;=\; |\mathbf{F}|\cos\theta \;=\; 12\cos(30^\circ) \;=\; 6\sqrt{3}\;\text{ N.}$$

While the magnitude of the normal component $\mathbf{F}_n$ is

$$|\mathbf{F}_n| \;=\; |\mathbf{F}|\sin\theta \;=\; 12\sin(30^\circ) \;=\; 6\;\text{ N.}$$

There is no question about the directions in such a physical set-up; one component is parallel, and one is normal to the rod, and you see from the physical diagram which way the vectors point.

Of course, if you want to write things in terms of coordinates, you can. Consider the whole situation as taking place in xy-plane, with the rod lying along the positive x-axis, with the origin being at the end where the force is acting. Then, the unit vector **u** which points in the direction of the rod is simply $\mathbf{u} = \mathbf{i}$.

Then,

$$\mathbf{F}_{\mathbf{u}} \;=\; (\mathbf{F}\cdot\mathbf{u})\mathbf{u} \;=\; \big(|\mathbf{F}|\cos\theta\big)\mathbf{i},$$

which, of course, is what we obtained without referring to the dot product.

Now, recall that, if an object undergoes a displacement of magnitude d along a line, and a force of magnitude F acts on the object in a direction parallel to the line, then the *work* done by the force on the object is $\pm Fd$, where the work is positive if the force and displacement are in the same direction, and negative if the force and displacement are in opposite directions.

More generally, if a force vector **F** acts in one direction and the displacement vector **d** is in a (possibly) different direction, then the work is calculated how it's calculated when the force acts parallel to the displacement, except that you use the magnitude of the component of the force that's parallel to **d**. Thus, the work is $\pm\,|\mathbf{F}_{\mathbf{d}}|\,|\mathbf{d}|$, where you pick + if $\mathbf{F_d}$ is in the direction of $\mathbf{d}$, and pick − if the direction of $\mathbf{F_d}$ is opposite that of $\mathbf{d}$.

Therefore, the absolute value of the work is given by:

$$|\text{work}| \;=\; \left|\left(\frac{\mathbf{F}\cdot\mathbf{d}}{|\mathbf{d}|^2}\right)\mathbf{d}\right|\cdot|\mathbf{d}| \;=\; |\mathbf{F}\cdot\mathbf{d}|.$$

But notice that $\mathbf{F_d}$ points in the direction of $\mathbf{d}$ if and only if $\mathbf{F}\cdot\mathbf{d}\geq 0$, and points in the direction opposite $\mathbf{d}$ if and only if $\mathbf{F}\cdot\mathbf{d}\leq 0$. It follows that the work done, with the appropriate $\pm$ sign is simply given by $\mathbf{F}\cdot\mathbf{d}$.

The above discussion was our intuitive lead-in to making the following definition:

Definition 1.3.13. *The work done by a (constant) force $\mathbf{F}$ in $\mathbb{R}^n$, acting on an object, as the object is displaced along a line by a displacement vector $\mathbf{d}$ in $\mathbb{R}^n$ is given by*

$$\text{work} \;=\; \mathbf{F}\cdot\mathbf{d} \;=\; |\mathbf{F}||\mathbf{d}|\cos\theta,$$

where θ is the angle between $\mathbf{F}$ and $\mathbf{d}$.

Later, in Section 1.6, we shall see that the object does not need to move along a straight line and yet, still, to calculate the work, you dot the constant force with the total, straight, displacement vector. However, we make this more basic definition here, and show that the general case follows from integrating the case where you look at infinitesimal displacement in a straight line.

We will wait until Section 4.2 to address the case where the force is not constant.

Example 1.3.14. Suppose that a force of $\mathbf{F}=(-1,3,2)$ Newtons acts on an object as it is displaced along a line from the point $(0,2,5)$ to $(4,0,-7)$, where all coordinates are measured in meters. How much work does the force do on the object?

Solution:

The displacement vector is

$$\mathbf{d} \;=\; (4,0,-7)-(0,2,5) \;=\; (4,-2,-12) \;\;\text{meters}.$$

Thus, the work done by $\mathbf{F}$ is

$$\mathbf{F}\cdot\mathbf{d} \;=\; (-1,3,2)\cdot(4,-2,-12) \;=\; -4-6-24 \;=\; -34 \;\;\text{joules}.$$

More Depth:

Example 1.3.15. Find the angles in the triangle with vertices $(1,1)$, $(2,5)$, and $(4,0)$.

Solution:

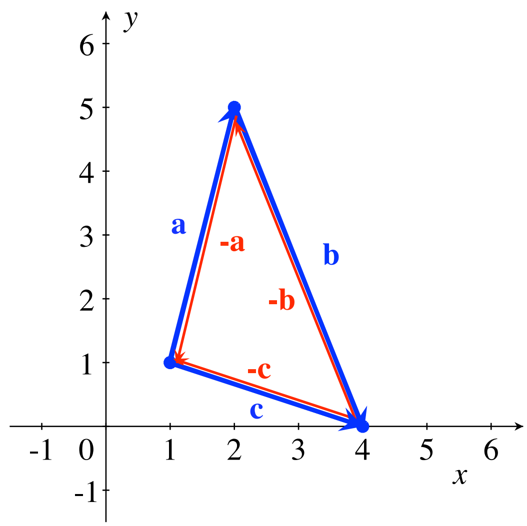

Figure 1.3.6: Vectors, and their negatives, between the vertices of a triangle.

We denote the angle inside the triangle, at a given vertex, by subscripting θ by the vertex, and thus want to calculate $\theta_{(1,1)}$, $\theta_{(2,5)}$, and $\theta_{(4,0)}$.

We calculate the displacement vectors between the vertices, as indicated in Figure 1.3.6. We find

$$\mathbf{a} = (2,5)-(1,1) = (1,4), \;\; \mathbf{b} = (4,0)-(2,5) = (2,-5), \;\text{ and }\; \mathbf{c} = (4,0)-(1,1) = (3,-1).$$

Now, remembering that we want the vectors to start at the same base point to determine the angle, we calculate

$$\theta_{(1,1)} = \cos^{-1}\left(\frac{\mathbf{a}\cdot\mathbf{c}}{|\mathbf{a}||\mathbf{c}|}\right) = \cos^{-1}\left(\frac{-1}{\sqrt{17}\sqrt{10}}\right) \approx 1.647568 \text{ radians} \approx 94.40^\circ;$$

$$\theta_{(2,5)} = \cos^{-1}\left(\frac{-\mathbf{a}\cdot\mathbf{b}}{|-\mathbf{a}||\mathbf{b}|}\right) = \cos^{-1}\left(\frac{18}{\sqrt{17}\sqrt{29}}\right) \approx 0.625485 \text{ radians} \approx 35.84^\circ;$$

and

$$\theta_{(4,0)} = \cos^{-1}\left(\frac{-\mathbf{b}\cdot -\mathbf{c}}{|-\mathbf{b}||-\mathbf{c}|}\right) = \cos^{-1}\left(\frac{11}{\sqrt{29}\sqrt{10}}\right) \approx 0.868539 \text{ radians} \approx 49.76^\circ.$$

Note that, as they should, the angles add up to 180°.

Example 1.3.16. Suppose that a constant force $\mathbf{F}$, in $\mathbb{R}^2$ or $\mathbb{R}^3$ (or, really, in $\mathbb{R}^n$), acts on an object as the object is displaced along a straight line from a point $\mathbf{p}_0$ to a point $\mathbf{p}_1$, then along a straight line from $\mathbf{p}_1$ to $\mathbf{p}_2$, then along a straight line from $\mathbf{p}_2$ to $\mathbf{p}_3$, and so on, and then finally along a straight line from $\mathbf{p}_{k-1}$ to $\mathbf{p}_k$.

Show that the total work done by the force is equal to simply the force dotted with the net displacement $\mathbf{d} = \mathbf{p}_k - \mathbf{p}_0$.

Solution

This is actually quite easy. For each i, where $1 \leq i \leq k$, let $\mathbf{d}_i = \mathbf{p}_i - \mathbf{p}_{i-1}$ be the displacement vector from $\mathbf{p}_{i-1}$ to $\mathbf{p}_i$. Then, the work done by $\mathbf{F}$, as the object is displaced by $\mathbf{d}_i$ is $W_i = \mathbf{F} \cdot \mathbf{d}_i$.

Thus, the total work is

$$W = \sum_{i=1}^{k} W_i \;=\; \sum_{i=1}^{k} \mathbf{F} \cdot \mathbf{d}_i \;=\; \mathbf{F} \cdot \left(\sum_{i=1}^{k} \mathbf{d}_i \right) \;=\; \mathbf{F} \cdot \left(\sum_{i=1}^{k} (\mathbf{p}_i - \mathbf{p}_{i-1}) \right) \;=$$

$$\mathbf{F} \cdot \big[(\mathbf{p}_1 - \mathbf{p}_0) + (\mathbf{p}_2 - \mathbf{p}_1) + (\mathbf{p}_3 - \mathbf{p}_2) + \cdots + (\mathbf{p}_k - \mathbf{p}_{k-1}) \big],$$

which "telescopes" to $\mathbf{F} \cdot (\mathbf{p}_k - \mathbf{p}_0) \;=\; \mathbf{F} \cdot \mathbf{d}$.

We would now like to prove the Cauchy-Schwarz Inequality from Theorem 1.3.2.

Let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbb{R}^n$. Note that, if either $\mathbf{a}$ or $\mathbf{b}$ is the zero vector, then the equality holds, and the vectors are parallel, since the zero vector has every direction.

So, assume that $\mathbf{b} \neq \mathbf{0}$. Then, for all real numbers t

$$(\mathbf{a} + t\mathbf{b}) \cdot (\mathbf{a} + t\mathbf{b}) \;=\; |\mathbf{a} + t\mathbf{b}|^2 \geq 0,$$

and the equality holds if and only if $\mathbf{a} + t\mathbf{b} = \mathbf{0}$. Note that $\mathbf{a} + t\mathbf{b} = \mathbf{0}$ implies that $\mathbf{a}$ and $\mathbf{b}$ are parallel.

Now, expanding algebraically, using properties 1-4 in Theorem 1.3.2, we find

$$(\mathbf{a} + t\mathbf{b}) \cdot (\mathbf{a} + t\mathbf{b}) \;=\; \mathbf{a} \cdot \mathbf{a} + t(\mathbf{a} \cdot \mathbf{b}) + t(\mathbf{b} \cdot \mathbf{a}) + t^2(\mathbf{b} \cdot \mathbf{b}) \;=$$

$$|\mathbf{a}|^2 + 2t(\mathbf{a} \cdot \mathbf{b}) + t^2|\mathbf{b}|^2 \;=\; |\mathbf{b}|^2 \left[t^2 + 2\left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right) t + \frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} \right],$$

where $|\mathbf{b}|^2 \neq 0$, since $\mathbf{b} \neq \mathbf{0}$.

This shows that, for all t,

$$t^2 + 2\left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right) t + \frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} \geq 0,$$

where equality holds if and only if $\mathbf{a} + t\mathbf{b} = \mathbf{0}$.

We are now going to complete the square in the t variable, and then see what special value of t gives us the Cauchy-Schwarz Inequality.

Completing the square, we obtain, for all t

$$t^2 + 2\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)t + \frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} \;=\; \left[t + \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right]^2 + \frac{|\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a}\cdot\mathbf{b})^2}{|\mathbf{b}|^4} \geq 0.$$

Therefore, if we let $t = -(\mathbf{a}\cdot\mathbf{b})/|\mathbf{b}|^2$, we conclude that

$$|\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a}\cdot\mathbf{b})^2 \geq 0,$$

and, if equality holds, then $\mathbf{a} = -t\mathbf{b}$, i.e., $\mathbf{a}$ is parallel to $\mathbf{b}$. Now note that, after taking square roots, the previous inequality is equivalent to the Cauchy-Schwarz Inequality: $|\mathbf{a}\cdot\mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|$.

It remains to be shown that, if $\mathbf{a}$ and $\mathbf{b}$ are parallel, then $|\mathbf{a}\cdot\mathbf{b}| = |\mathbf{a}||\mathbf{b}|$. However, this is easy; if $\mathbf{a} = s\mathbf{b}$, then both sides of the equality equal $|s||\mathbf{b}|^2$.

We can now prove a "geometrically obvious" theorem; one which effectively says that the sum of the lengths of two sides of a triangle is greater than the length of the remaining side.

Theorem 1.3.17. (**Triangle Inequality**) *Suppose that* $\mathbf{a}$ *and* $\mathbf{b}$ *are vectors in* $\mathbb{R}^n$. *Then,*

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|,$$

and equality holds if and only if $\mathbf{a}$ *and* $\mathbf{b}$ *have the same direction.*

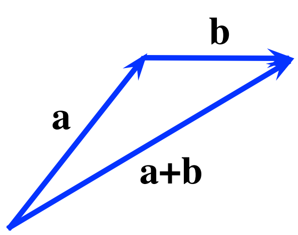

Figure 1.3.7: Geometric representation of the Triangle Inequality.

Proof. The inequality is equivalent to:

$$|\mathbf{a} + \mathbf{b}|^2 \leq |\mathbf{a}|^2 + 2|\mathbf{a}||\mathbf{b}| + |\mathbf{b}|^2.$$

Now, we use that

$$|\mathbf{a} + \mathbf{b}|^2 \;=\; (\mathbf{a} + \mathbf{b})\cdot(\mathbf{a} + \mathbf{b}) \;=\; (\mathbf{a}\cdot\mathbf{a}) + 2(\mathbf{a}\cdot\mathbf{b}) + (\mathbf{b}\cdot\mathbf{b}) \;=\; |\mathbf{a}|^2 + 2(\mathbf{a}\cdot\mathbf{b}) + |\mathbf{b}|^2$$

to conclude that the Triangle Inequality is equivalent to

$$\mathbf{a}\cdot\mathbf{b} \leq |\mathbf{a}||\mathbf{b}|,$$

which follows at once from the Cauchy-Schwarz Inequality.

In addition, the equality $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|$ implies that $\mathbf{a}$ and $\mathbf{b}$ are parallel by the Cauchy-Schwarz Theorem; but the equality also implies that $\mathbf{a}\cdot\mathbf{b} \geq 0$, so that $\mathbf{a}$ and $\mathbf{b}$ must, in fact, have the same direction. $\square$

+ Linear Algebra:

In the + Linear Algebra portion of the previous section, we discussed that a vector $\mathbf{v}$ in $\mathbb{R}^n$ is frequently written as a column vector $[\mathbf{v}]$, a matrix with n rows and 1 column.

There is an operation on matrices called *transpose*. The transpose of an $n \times m$ matrix A is denoted A^{T}, and is the $m \times n$ matrix whose rows are the columns of A, and whose columns are the rows of A.

For instance,

$$\begin{bmatrix} 3 & -5 & 7 \\ -1 & 0 & 2 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 3 & -1 \\ -5 & 0 \\ 7 & 2 \end{bmatrix}.$$

In particular, the transpose of a column vector with n entries is a *row vector* with the same n entries, e.g.,

$$\begin{bmatrix} 2 \\ 5 \\ -1 \\ 7 \end{bmatrix}^{\mathrm{T}} = [\, 2 \;\; 5 \;\; -1 \;\; 7 \,].$$

In terms of *matrix multiplication*, the dot product of two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$, written as column vectors, is defined to be (the unique entry of) the 1×1 matrix given by

$$[\mathbf{v}] \cdot [\mathbf{w}] \;=\; [\mathbf{v}]^{\mathrm{T}}[\mathbf{w}].$$

Technically, there is a difference between a 1×1 matrix and the unique entry of the matrix; we shall not distinguish between these two objects.

If $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^n$, and A is an $n \times n$ matrix, then the matrix product $A[\mathbf{v}]$ is a column vector in $\mathbb{R}^n$, and $(A[\mathbf{v}])^{\mathrm{T}} \;=\; [\mathbf{v}]^{\mathrm{T}} A^{\mathrm{T}}$.

Therefore, we arrive at the important formula for how matrix multiplication interacts with the dot product:

$$A[\mathbf{v}] \cdot [\mathbf{w}] \;=\; (A[\mathbf{v}])^{\mathrm{T}}\,[\mathbf{w}] \;=\; [\mathbf{v}]^{\mathrm{T}}\left(A^{\mathrm{T}}[\mathbf{w}]\right) \;=\; [\mathbf{v}] \cdot A^{\mathrm{T}}[\mathbf{w}].$$