PLANE TRIGONOMETRY AND VECTOR GEOMETRY

GABRIEL NAGY

September 2017

Contents

Chapter 1 Trigonometry for Acute Angles **1**
- 1.1 Measures of Physical Angles . . . 1
- 1.2 Right Triangle Trigonometry . . . 13
- 1.3 Solving Right Triangles . . . 23

Chapter 2 Trigonometry Beyond Acute Angles **35**
- 2.1 Basic Notions of Analytic and Vector Geometry . . . 35
- 2.2 The Analytic Construction of the Trigonometric Functions . . . 63
- 2.3 Trigonometric Identities . . . 80
- 2.4 Graphs of Trigonometric Functions . . . 88
- 2.5 General Sinusoidal Functions and Their Graphs . . . 108
- 2.6 The Inverse Trigonometric Functions . . . 114
- 2.7 Trigonometric Equations . . . 126

Chapter 3 Applications of Trigonometry in Geometry **141**
- 3.1 Applications to Vector Geometry . . . 141
- 3.2 Applications to Triangle Geometry . . . 155
- 3.3 Polar Coordinates . . . 172

Chapter 4 Trigonometry for Combinations of Angles **187**
- 4.1 Sums and Differences of Angles . . . 187
- 4.2 Multiples of Angles . . . 198
- 4.3 From Sums to Products and Back . . . 206

Chapter 5 Applications of Trigonometry in Algebra **213**
- 5.1 Complex Number Arithmetic . . . 213
- 5.2 Geometry of Complex Numbers . . . 219
- 5.3 Powers and Roots . . . 224

Chapter 6 Applications of Trigonometry in Analytic Geometry **237**
- 6.1 Parabolas . . . 238
- 6.2 Ellipses . . . 247
- 6.3 Hyperbolas . . . 260
- 6.4 Quadratic Curves in General Position . . . 274
- 6.5 Conics in Polar Coordinates . . . 288

Appendices **305**
A Algebra Review I: Equations 305
B Algebra Review II: Matrix Arithmetic 314

Chapter 1

Trigonometry for Acute Angles

Here beginneth TRIGONOMETRY!

1.1 Measures of Physical Angles

We start off by reviewing several concepts from Plane Geometry and set up some basic terminology.

A **geometric angle** is simply a *union of two rays that emanate from the same source*, which we call the *vertex* of the angle; the two rays are called the *sides* of the angle. When the rays are named, say, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, our angle will be denoted by $\angle \boldsymbol{r}_1 \boldsymbol{s}_2$. Depictions of geometric angles will look like the one shown in this figure:

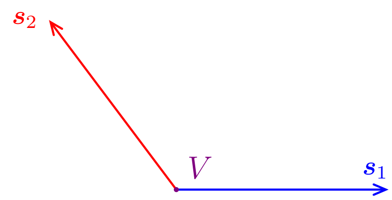


Figure 1.1.1

The above figure depicts a *non-flat* geometric angle $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$, with vertex V. The *flat* angles are:

- the **straight** angles, whose sides are *opposite "halves" of a line*;
- the **trivial** angles, whose sides *coincide*.

Any *non-flat* angle $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$ splits the plane in two *regions*, which we refer to as the **physical angles enclosed by** $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$.

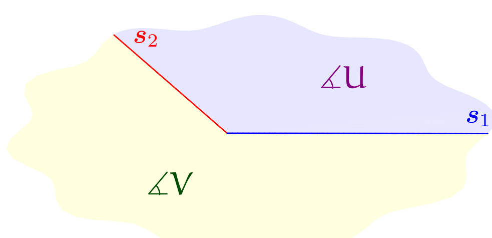


Figure 1.1.2

The figure above depicts these two physical angles, denoted by $\measuredangle \mathsf{U}$ and $\measuredangle \mathsf{V}$. To be a bit more specific, $\measuredangle \mathsf{U}$ is the *inner* physical angle enclosed by $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$; while $\measuredangle \mathsf{V}$ is the *outer* physical angle enclosed by $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$. Both $\measuredangle \mathsf{U}$ and $\measuredangle \mathsf{V}$ have same *vertex*, which is simply the vertex of the geometric angle $\angle \boldsymbol{s}_1 \boldsymbol{s}_2$ that encloses them.

However, when we are dealing with *flat* geometric angles, we agree that

- The physical angles enclosed by a *straight* geometric angle – which is just a line $\mathscr{L}$ – are simply the two *half-planes* determined by $\mathscr{L}$. These two physical angles will be called *straight physical angles*.
- The physical angles enclosed by a *trivial* geometric angle – which is just a ray $\boldsymbol{s}$ – is simply $\boldsymbol{s}$ (the ray itself) – which we call a *trivial physical angle*, and the *entire plane* – which we call a *complete physical angle*.

So, unless we are dealing with a straight physical angle, the two physical angles enclosed by $\angle \boldsymbol{s}_1\boldsymbol{s}_2$ are clearly distinguishable: one is "small" (the *inner* angle); the other is "big" (the *outer* angle).

In particular, any triangle $\triangle ABC$ has three (clearly defined) physical angles, $\measuredangle A$, $\measuredangle B$, and $\measuredangle C$, as depicted in the figure below.

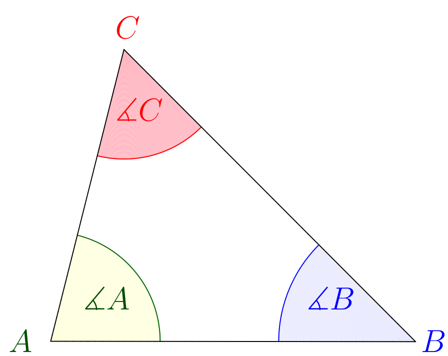


Figure 1.1.3

Turn Measure and the Protractor Principles

We measure a physical angle with the help of a *protractor*. To be precise, if we start with some physical angle $\measuredangle \mathfrak{U}$, what we call a **protractor suitable for** $\measuredangle \mathfrak{U}$ is nothing else but a *circle* C, *whose center is placed at the vertex of* $\measuredangle \mathfrak{U}$. With this set-up, we say that $\measuredangle \mathfrak{U}$ is *central in* C.

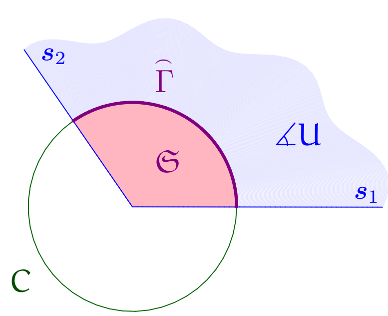


Figure 1.1.4

Once a protractor is set up like this (as shown in the picture above), two important geometric objects arise:

(A) the *arc* $\widehat{\Gamma}$ *on* C *subtended by* $\measuredangle \mathfrak{U}$;
(B) the *disk sector* $\mathfrak{S}$ *on* C *subtended by* $\measuredangle \mathfrak{U}$.

With all the set-up as above, the **turn measure of** $\measuredangle \mathfrak{U}$ is the ratio:

$$\mathfrak{t}(\measuredangle \mathfrak{U}) = \text{length}(\widehat{\Gamma}) : \text{circumference}(\mathsf{C}).$$

Furthermore, if we consider the entire disk D enclosed by C, then the turn measure can also be presented as:

$$\mathfrak{t}(\measuredangle \mathfrak{U}) = \text{Area}(\mathfrak{S}) : \text{Area}(\mathsf{D}).$$

CLARIFICATIONS AND ADDITIONAL NOTATION. The turn measure of a physical angle is always a number in the interval $[0, 1]$. To simply our notation a little bit, instead of writing "$\mathfrak{t}(\measuredangle \mathfrak{U}) = \boldsymbol{\tau}$" we are simply going to write: "$\widehat{\mathfrak{U}} = \boldsymbol{\tau}$ turn(s)."

Example 1.1.1. Certain physical angles can be clearly distinguished, based on their measures, as shown in the table below.

Angle type	*trivial*	*straight*	*complete*	*right*
Measure	0 turn(s)	$\frac{1}{2}$ turn	1 turn	$\frac{1}{4}$ turn

Table 1.1.1

The Protractor Principles

I. *The turn measure of an angle does not depend on the protractor radius.*

II. *Two physical angles are congruent, if and only if they have equal turn measures.*

III. *For any ray* $\boldsymbol{s}$ *and any number* $0 < \boldsymbol{\tau} < 1$, *there are exactly two physical angles* $\measuredangle \mathfrak{U}$ *and* $\measuredangle \mathfrak{U}'$, *such that:*
- *both* $\measuredangle \mathfrak{U}$ *and* $\measuredangle \mathfrak{U}'$ *have* $\boldsymbol{s}$ *as one of their sides, and*
- $\widehat{\mathfrak{U}} = \widehat{\mathfrak{U}}' = \boldsymbol{\tau}$ turn(s).

When $\boldsymbol{\tau} = 0$ *or* $\boldsymbol{\tau} = 1$, *only one such angle exists.*

NOTATION CONVENTION. The measures of the physical angles (see Figure 1.1.3) in a triangle $\triangle ABC$ are denoted by $\widehat{A}$ (the measure of $\measuredangle A$), $\widehat{B}$ (the measure of $\measuredangle B$), and $\widehat{C}$ (the measure of $\measuredangle C$).

Radian Measure

Besides the turn, another very important unit for angle measurement is the **radian**. When we want to measure an angle in *radians*, we still use protractors as above, but *instead of dividing* length$(\widehat{\Gamma})$ *by* circumference(C), *we divide* length$(\widehat{\Gamma})$ *by the radius of* C. Since the circumference of a circle is $2\pi \cdot$ *radius*, our conversion rule simply reads:

$$\mathbf{1}\ \text{turn} = \mathbf{2\pi}\ \text{radians}. \tag{1.1.1}$$

Using the identity (1.1.1), we can convert between our units, using:

Turn ↔ Radian Conversion Formulas

$$\boldsymbol{\tau} \text{ turn(s)} = \mathbf{2\pi} \cdot \boldsymbol{\tau} \text{ radian(s)}; \tag{1.1.2}$$

$$\boldsymbol{\theta} \text{ radians(s)} = \frac{\boldsymbol{\theta}}{\mathbf{2\pi}} \text{ turn(s)}. \tag{1.1.3}$$

NOTATION CONVENTION. When using *radians*, we ought to write angle measures like: "$\widehat{\mathrm{U}} = \boldsymbol{\theta}$ radian(s)." We are going to be "lazy" from now on, and **omit** the word "radian(s)" from our notation. In other words, whenever we see:

$$\widehat{\mathrm{U}} = \textit{number},$$

we understand that "*number*" stands for the *radian measure of* $\measuredangle \mathrm{U}$.

The "lazy" notation is **only used with radians!**. Thus, when other units are used (for example, turns), they **must be specified!**

Circle Measurements

The radian measure is very useful, particularly when computing lengths of arcs, areas of sectors, as seen in the following set of formulas.

Arc Length and Sector Area Formulas

Assume a circle C is given, along with some physical angle $\measuredangle \mathrm{U}$, which is central in C, and has radian measure: $\widehat{\mathrm{U}} = \boldsymbol{\theta}$. Assume also, some length **unit** is given, so that:

- the *radius* of C is $\boldsymbol{r}$ **units**;
- the *arc* $\widehat{\Gamma}$ subtended by $\measuredangle \mathrm{U}$ has length$(\widehat{\Gamma}) = \boldsymbol{\ell}$ **units**;
- the *sector* $\mathfrak{S}$ subtended by $\measuredangle \mathrm{U}$ has Area$(\mathfrak{S}) = \boldsymbol{A}$ **square units**.

Then the four numbers $\boldsymbol{\theta}$, $\boldsymbol{r}$, $\boldsymbol{\ell}$, and $\boldsymbol{A}$ are linked by the following identities:

$$\boldsymbol{\ell} = \boldsymbol{\theta}\, \boldsymbol{r} \tag{1.1.4}$$

$$\boldsymbol{A} = \tfrac{1}{2}\, \boldsymbol{\ell}\, \boldsymbol{r} = \tfrac{1}{2}\, \boldsymbol{\theta}\, \boldsymbol{r}^2. \tag{1.1.5}$$

In most applications involving arc or sectors determined by central angles in circles, we are facing the following

Four-Number Problem: *Given two of the numbers $\boldsymbol{\theta}$, $\boldsymbol{r}$, $\boldsymbol{\ell}$, $\boldsymbol{A}$, find the missing two.*

All instances of the Four-Number Problem (all in all, there are six possibilities) can be solved with the help of (1.1.4) and (1.1.5), and the following formulas which are derived from them:

Derived Angle-Radius-Length-Area Formulas. If the positive numbers $\boldsymbol{\theta}$, $\boldsymbol{r}$, $\boldsymbol{\ell}$, $\boldsymbol{A}$ satisfy (1.1.4) and (1.1.5), then they also satisfy:

$$\boldsymbol{r} = \frac{\boldsymbol{\ell}}{\boldsymbol{\theta}} = \sqrt{\frac{2\boldsymbol{A}}{\boldsymbol{\theta}}} = \frac{2\boldsymbol{A}}{\boldsymbol{\ell}}; \tag{1.1.6}$$

$$\boldsymbol{\theta} = \frac{\boldsymbol{\ell}}{\boldsymbol{r}} = \frac{2\boldsymbol{A}}{\boldsymbol{r}^2} = \frac{\boldsymbol{\ell}^2}{2\boldsymbol{A}}; \tag{1.1.7}$$

$$\ell = \frac{2A}{r} = \sqrt{2A\theta}; \qquad A = \frac{\ell^2}{2\theta}. \tag{1.1.8}$$

The second equality from (1.1.6) deserves a little explanation. When we look at (1.1.5), we can re-write it as $r^2 = 2A/\theta$. In principle an equation of the from "$r^2 = \textit{number}$" has two solutions: $r = \pm\sqrt{\textit{number}}$. However, since r is positive (as it represents a distance), only the solution that uses the $+$ sign will be of interest to us. All other equalities that involve a square root are treated the same way.

Example 1.1.2. Given $\widehat{\mathsf{U}} = 2\pi/3$ and *radius* $= 6$ inches, find length$(\widehat{\Gamma})$ and Area$(\mathfrak{S})$.

Solution. The given quantities in our Four-Number Problem are $\theta = 2\pi/3$ and $r = 6$ (with "inch" as our **unit**). Using (1.1.4) we get: length$(\widehat{\Gamma}) = \ell = (2\pi/3)\cdot 6 = 4\pi \simeq 12.56637061$ inches. Using (1.1.5) we get: Area$(\mathfrak{S}) = A = \frac{1}{2}\cdot(2\pi/3)\cdot 6^2 = 12\pi \simeq 37.69911184$ square inches.

Example 1.1.3. Given $\widehat{\mathsf{U}} = 5$ and length$(\widehat{\Gamma}) = 1.2$ miles, the *radius* of the circle and Area$(\mathfrak{S})$.

Solution. The given quantities in our Four-Number Problem are $\theta = 5$ and $\ell = 1.2$ (with "mile" as our **unit**). Using (1.1.6) we get: *radius* $= r = 1.2/5 = 0.24$ miles. Using (1.1.5) we also get: Area$(\mathfrak{S}) = A = \frac{1}{2}\cdot\ell\cdot r = \frac{1}{2}\cdot 1.2\cdot 0.24 = 1.44$ square miles.

▶ More examples of the Four-Number Problem are provides in Exercises 1–3, as well as in the K-State Online Homework System.

Degree Measurement

Besides the turn and the radian, another very popular unit for angle measurement is the **degree**, denoted with the help of the symbol °. So when we want to specify the *degree measure* of some physical angle $\measuredangle\mathsf{U}$, we will write:

$$\widehat{\mathsf{U}} = D^\circ.$$

The defining conversion rule for the degree measure is:

$$\mathbf{1}\ \text{turn} = \mathbf{360}^\circ. \tag{1.1.9}$$

As we continue through the rest of this section, we will gradually diminish the use of turn measures, and limit ourselves only to radians and degrees. Using (1.1.1), the above conversion rule now reads

$$\mathbf{2\pi}\ [\text{radians}] = \mathbf{360}^\circ. \tag{1.1.10}$$

(According to our Notation Convention stated earlier, we should omit "radian(s)" from all radian measure specifications. For pedagogical reasons, we will still keep "radian(s)" in certain formulas for awhile, but use square brackets to remind us that, once we get more and more comfortable with our Convention, we will eventually omit "[radian(s)]" from our notations.)

By taking halves on both sides, the identity (1.1.10) becomes

$$\boldsymbol{\pi}\ [\text{radians}] = \mathbf{180}^\circ, \tag{1.1.11}$$

so we can easily convert between radians and degrees, using:

Degree ↔ Radian Conversion Formulas

$$D^\circ = D \cdot \frac{\pi}{180} \text{ [radians]}; \tag{1.1.12}$$

$$\theta \text{ [radians]} = \left(\theta \cdot \frac{180}{\pi}\right)^\circ \tag{1.1.13}$$

Example 1.1.4. If we know that a physical angle $\measuredangle \mathsf{U}$ measures 105°, and we want to compute its radian measure, we use formula (1.1.12), which yields:

$$\widehat{\mathsf{U}} = 105 \cdot \frac{\pi}{180} = \frac{105 \cdot \pi}{180} = \frac{7\pi}{12} \simeq 1.832595715.$$

Notice that we omitted "radians" from our final answer. As for the way we present the radian measure, even though the numerical value given above is good enough to give us an idea of how large/small our angle is, the first answer, which presents the radian measure as a *fraction involving integers and integer multiples of* π is the preferred one, as we deem is as an *exact value*.

Example 1.1.5. If we know that a physical angle $\measuredangle \mathsf{U}$ measures $\frac{11\pi}{450}$ (in radians), and we want to compute its degree measure, we use formula (1.1.13), which yields:

$$\widehat{\mathsf{U}} = \left(\frac{11\pi}{450} \cdot \frac{180}{\pi}\right)^\circ = \left(\frac{11\pi \cdot 180}{450 \cdot \pi}\right)^\circ = \left(\frac{22}{5}\right)^\circ = 4.5^\circ.$$

For future reference, we now expand Table 1.1.1 which had the turn measures of several "special" angles, to also include the radian and degree measures:

Angle type	*trivial*	*straight*	*complete*	*right*
Turn Measure	0 turn(s)	$\frac{1}{2}$ turn	1 turn	$\frac{1}{4}$ turn
Radian Measure	0	π	2π	$\frac{\pi}{2}$
Degree Measure	0°	180°	360°	90°

Table 1.1.2

The D(egree)-M(inute)-S(econd) Measurement System

When doing computations using degrees, we can of course present them as *decimals*, as we have seen for instance in Example 1.1.5. However, when dealing with with a *non-integer* degree measure, say D°, we may convert the *fractional (or decimal) part* of D, we may use sub-divisions of the degree, which we call **minutes** (denoted using the symbol $'$), and **seconds** (denoted using the symbol $''$). The basic rules that explain how these sub-divisions are set up are:

$$1^\circ = 60' = 60''; \qquad 1' = 60''. \tag{1.1.14}$$

Fractional (Decimal) Unit $\leftrightarrow$ Units + Sub-Units Conversion Scheme

Assume we have one measurement **unit**, and a **sub-unit**, defined by:

$$\mathbf{1\ unit} = \boldsymbol{m}\ \mathbf{sub\text{-}units}, \tag{1.1.15}$$

so that the back and forth conversion formulas are:

$$\boldsymbol{X}\ \mathbf{units} = \boldsymbol{X} \cdot \boldsymbol{m}\ \mathbf{sub\text{-}units}, \tag{1.1.16}$$

$$\boldsymbol{X}\ \mathbf{sub\text{-}units} = \left(\frac{\boldsymbol{X}}{\boldsymbol{m}}\right)\ \mathbf{units} \tag{1.1.17}$$

A. Given some measurement of the form

$$\boldsymbol{\mu} = \boldsymbol{U}\ \mathbf{units}, \tag{1.1.18}$$

with $\boldsymbol{U}$ is *either a fractional, or a decimal number*, we can convert it to the form

$$\boldsymbol{\mu} = \boldsymbol{A}\ \mathbf{units} + \boldsymbol{B}\ \mathbf{sub\text{-}units}, \tag{1.1.19}$$

using the following procedure:

(i) *Split the number $\boldsymbol{U}$ into its integer part (which is what we define $\boldsymbol{A}$ to be), and its fractional part (which we call for the moment $\boldsymbol{X}$).* In other words, we write

$$\boldsymbol{U} = \boldsymbol{A} + \boldsymbol{X}$$

with $\boldsymbol{A}$ *integer*, and $0 \leq \boldsymbol{X} < 1$ *fractional (or decimal)*.

(ii) *Use* (1.1.16) *to convert the fractional part to sub-units, thus defining $\boldsymbol{B}$ to be the product $\boldsymbol{X} \cdot \boldsymbol{m}$.*

B. Conversely, given a measurement presented as (1.1.19), we can convert it to the form (1.1.18), using the formula:

$$\boldsymbol{A}\ \mathbf{units} + \boldsymbol{B}\ \mathbf{sub\text{-}units} = \left(\boldsymbol{A} + \frac{\boldsymbol{B}}{\boldsymbol{m}}\right)\ \mathbf{units} \tag{1.1.20}$$

When we specialize procedure B to the DMS system, we get:

DMS $\longrightarrow$ Degrees Conversion Formula

$$\boldsymbol{D}^\circ \boldsymbol{M}' \boldsymbol{S}'' = \left(\boldsymbol{A} + \frac{\boldsymbol{M}}{\mathbf{60}} + \frac{\boldsymbol{S}}{\mathbf{3600}}\right)^\circ$$

Example 1.1.6. Convert 12.3456° to degrees, minutes and seconds.

Solution. We start off by converting to degrees and minutes, using the above procedure A (with degree as the unit, and minute as the sub-unit):

$$12.3456^\circ = 12^\circ + 0.3456^\circ = 12^\circ + 0.3456 \cdot 60' = 12^\circ + 20.736'.$$

Then we convert $20.736'$ to minutes and seconds, again using the above procedure A (with minute as the unit, and second as the sub-unit), so our calculation continues:

$$\begin{aligned} 12.3456^\circ &= 12^\circ + 20.736' = 12^\circ + 20' + 0.736' = 12^\circ + 20' + 0.736 \cdot 60'' = \\ &= 12^\circ + 20' + 44.16'' \simeq 12^\circ\, 20'\, 44''. \end{aligned}$$

Example 1.1.7. Convert $12°30'45''$ to degrees.
Solution. Using the DMS $\to$ Degrees Formula we have:

$$12°30'67'' = \left(12 + \frac{30}{60} + \frac{67}{3600}\right)^{\circ} = \left(\frac{12 \cdot 3600 + 30 \cdot 60 + 67}{3600}\right)^{\circ} =$$
$$= \left(\frac{45067}{3600}\right)^{\circ} \simeq 12.51861111°.$$

When asked to convert to *degrees in decimal form*, the fraction calculation above is not necessary: we can do everything in "one shot" on a calculator by typing `12+30/60+67/3600`.

In the preceding two examples we worked with decimals. However, in many instances we want to work with *fractions*, instead of decimals. This is illustrated in the following two examples

Example 1.1.8. Convert $\frac{89\pi}{4050}$ [radians] to degrees, minutes and seconds.
Solution. We start off by converting from radians to degrees, using (1.1.13), which yields:

$$\frac{89\pi}{4050} = \left(\frac{89\pi}{4050} \cdot \frac{180}{\pi}\right)^{\circ} = \left(\frac{89\pi \cdot 180}{4050 \cdot \pi}\right)^{\circ} = \left(\frac{89 \cdot 2}{45}\right)^{\circ} = \left(\frac{178}{45}\right)^{\circ}.$$

Next we convert to degrees and minutes, using the above procedure A (with degree as the unit, and minute as the sub-unit):

$$\left(\frac{178}{45}\right)^{\circ} = 3° + \left(\frac{43}{45}\right)^{\circ} = 3° + \left(\frac{43}{45}\right) \cdot 60' = 3° + \left(\frac{43 \cdot 4}{3}\right)' = 3° + \left(\frac{172}{3}\right)'.$$

Then we convert $\left(\frac{172}{3}\right)'$ to minutes and seconds, again using the above procedure A (with minute as the unit, and second as the sub-unit), so our calculation continues:

$$3° + \left(\frac{172}{3}\right)' = 3° + 57' + \left(\frac{1}{3}\right)' = 3° + 57' + \left(\frac{1}{3}\right) \cdot 60'' = 3° + 57' + 20'' = 3°\,57'\,20''.$$

Example 1.1.9. Convert $14°7'12''$ to radians.
Solution. We start off by converting to degrees, using DMS $\to$ Degrees Formula. Since we want to work with fractions, throughout the entire computation, we will be careful and simplify as much as possible

$$14°31'12'' = \left(14 + \frac{31}{60} + \frac{12}{3600}\right)^{\circ} = \left(14 + \frac{31}{60} + \frac{1}{300}\right)^{\circ} =$$
$$= \left(\frac{14 \cdot 300 + 31 \cdot 5 + 1}{300}\right)^{\circ} = \left(\frac{4356}{300}\right)^{\circ} = \left(\frac{363}{25}\right)^{\circ}.$$

Next we convert to radians, using (1.1.12):

$$14°31'12'' = \left(\frac{363}{25}\right)^{\circ} = \left(\frac{363}{25}\right) \cdot \frac{\pi}{180} = \frac{363 \cdot \pi}{25 \cdot 180} = \frac{121\pi}{25 \cdot 60} = \frac{121\pi}{1500} \simeq 0.2534218074.$$

As pointed out earlier, if we only want the radian measure expressed in decimal form, we could have done it in "one shot' with our calculator by typing:

```
(14+31/60+12/3600)*π/180
```

However, the conversion $14°31'12'' = \dfrac{121\pi}{1500}$ is always the preferred one, as it gives an *exact value*.

Example 1.1.10. (Compare with Example 1.1.2.) Suppose we have a central angle $\widehat{\text{U}} = 100°20'$ in a circle of radius 12 cm, and we are asked to compute the length of the arc $\widehat{\Gamma}$ subtended by it.

Solution. Of course, what we need here is formula (1.1.4), but before we use it, we must find the radian measure $\boldsymbol{\theta}$ of our angle. Working as in the preceding Example, we have

$$\widehat{\text{U}} = 100°20' = \left(100 + \frac{20}{60}\right)^{\circ} = \left(100 + \frac{1}{3}\right)^{\circ} = \left(\frac{301}{3}\right)^{\circ} = \left(\frac{301}{3}\right)^{\circ} \cdot \frac{\pi}{180} = \frac{301\pi}{540}$$

With this calculation in mind, the arc length formula (1.1.4) yields

$$\text{length}(\widehat{\Gamma}) = (301\pi/540) \cdot 12 = 301\pi/45 \simeq 21.01376419 \text{ cm}.$$

Angle Arithmetic

One key feature of physical angle measurements is the statement below, concerning *sub-angles*.

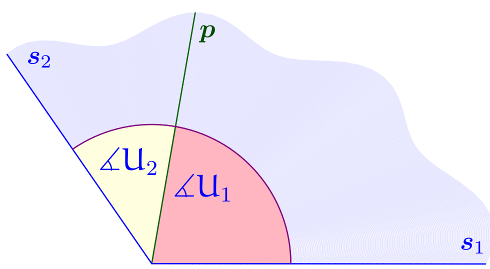


Figure 1.1.5

Given a physical angle $\measuredangle\text{U}$, any ray $\boldsymbol{p}$ that *emanates from the vertex of* $\measuredangle\text{U}$ *and sits inside* $\measuredangle\text{U}$, splits $\measuredangle\text{U}$ into two physical sub-angles, called the $\boldsymbol{p}$*-splits* (or $\boldsymbol{p}$*-sub-angles*) of $\measuredangle\text{U}$.

As a matter of convention, if the sides of $\measuredangle\text{U}$ are $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ (as shown for instance in Figure 1.1.5), the two $\boldsymbol{p}$-splits of $\measuredangle\text{U}$ are identified as follows:

- $\measuredangle\text{U}_1$ is the physical sub-angle enclosed by $\angle\boldsymbol{s}_1\boldsymbol{p}$;
- $\measuredangle\text{U}_2$ is the physical sub-angle enclosed by $\angle\boldsymbol{s}_2\boldsymbol{p}$.

Angle Addition Principle

With $\measuredangle\text{U}$ *and* $\boldsymbol{p}$ *as above, the measures of two* $\boldsymbol{p}$*-sub-angles* $\measuredangle\text{U}_1$, $\measuredangle\text{U}_2$ *add up to the measure of the angle* $\measuredangle\text{U}$:

$$\widehat{\text{U}}_1 + \widehat{\text{U}}_2 = \widehat{\text{U}} \tag{1.1.21}$$

This property *does not depend on the angle measure unit*.

CLARIFICATION. Based on the Angle Addition Principle, we see that if an angle $\measuredangle\text{U}$ is split into sub-angles $\measuredangle\text{U}_1$ and $\measuredangle\text{U}_2$, and if we know two out of the three measures $\widehat{\text{U}}_1$, $\widehat{\text{U}}_2$, $\widehat{\text{U}}$, then the third measure can obviously found either doing an *addition* as in (1.1.21), or a *subtraction*, for instance: $\widehat{\mathrm{u}}_2 = \widehat{\mathrm{u}} - \widehat{\mathrm{u}}_1$.

Of course, when the angle measures (whether they use turns, degrees, or radians) are expressed as decimals or fractions, additions and subtractions are easy to perform. The only type of arithmetic that is a little "tricky" is the one that uses the DMS system, in which case we use the following guidelines (see Examples 1.1.11, 1.1.12 and 1.1.13 below):

- Always try to add/subtract *like units* (degrees with degrees, minutes with minutes, seconds with second).
- When an addition results in large number, *carry* "chunks" of sub-units to units.
- When a subtraction results in a negative number of sub-units, *borrow* a "chunk" of sub-units from units.

Example 1.1.11. Add: $30°50'48'' + 89°42'15''$.

Solution. Set up our addition in columns:

$$\begin{array}{cccc} & 30° & 50' & 48'' \\ + & 89° & 42' & 15'' \\ \hline = & 119° & 92' & 63'' \end{array}$$

Although the final answer looks OK, some quantities exceed the usual cap: when we look at seconds, we see $63''$, which exceeds $60''$, so we can carry a $60''$ chunk, by replacing $63'' = 1' + 3''$, so now the above result reads: $119°(92+1)'3'' = 119°93'3''$. Likewise, when we look at minutes, we now see $93'$, which exceeds $60'$, so we can carry a $60'$ chunk, by replacing $93' = 1° + 33'$, so now the above result reads: $(119+1)°33'3'' = 120°33'3''$, which we regard as a *clean answer*.

Example 1.1.12. Subtract: $100°20'10'' - 65°43'21''$.

Solution. We would like to subtract seconds from seconds, and minutes from minutes. When we look at seconds, we see that we would get a negative number, so we will borrow $60'' = 1'$ from the minutes, so now our subtraction looks like:

$$100°20'10'' - 65°43'21'' = 100°(20-1)'(10+60)'' - 65°43'21'' = 100°19'70'' - 65°43'21''.$$

This looks fine, as far as the seconds are concerned, but with the minutes we have the same problem, so we will borrow $60' = 1°$ from the degrees, so now our subtraction looks like:

$$100°19'70'' - 65°43'21'' = (100-1)°(19+60)'70'' - 65°43'21'' = 99°79'70'' - 65°43'21''.$$

and all is fine, so we can subtract in columns:

$$\begin{array}{cccc} & 99° & 79' & 70'' \\ - & 65° & 43' & 21'' \\ \hline = & 34° & 36' & 49'' \end{array}$$

Supplements and Complements

Two special types of angle pairs are identified, depending on the sum of their measures.

Suppose two physical angles $\measuredangle \mathrm{U}_1$ and $\measuredangle \mathrm{U}_2$ are given.

(A) We say that $\measuredangle \mathrm{U}_1$ and $\measuredangle \mathrm{U}_2$ are **complementary**, if the *the sum* $\widehat{\mathrm{U}}_1 + \widehat{\mathrm{U}}_2$ *of their measures if equal to the measure of a right angle*.

(B) We say that $\measuredangle \mathrm{U}_1$ and $\measuredangle \mathrm{U}_2$ are **supplementary**, if the *the sum* $\widehat{\mathrm{U}}_1 + \widehat{\mathrm{U}}_2$ *of their measures if equal to the measure of a straight angle*.

CLARIFICATIONS AND ADDITIONAL TERMINOLOGY. In case (A), we say that $\measuredangle \mathrm{U}_1$ is **complementary to** $\measuredangle \mathrm{U}_2$ (and vice-versa, $\measuredangle \mathrm{U}_2$ is complementary to $\measuredangle \mathrm{U}_1$). Likewise, in case (B), we say that $\measuredangle \mathrm{U}_1$ is **supplementary to** $\measuredangle \mathrm{U}_2$ (and vice-versa, $\measuredangle \mathrm{U}_2$ is supplementary to $\measuredangle \mathrm{U}_1$).

In practical terms, finding complements or supplements amount to an angle subtraction:

(A) To find the measure of a *complement* of an angle $\measuredangle \mathrm{U}$, we *subtract* its measure $\widehat{\mathrm{U}}$ *from the measure of a right angle*.

(B) To find the measure of a *supplement* of an angle $\measuredangle \mathrm{U}$, we *subtract* its measure $\widehat{\mathrm{U}}$ *from the measure of a straight angle*.

(Depending on the unit we use, we can look up the measure of the right or straight angle in Table 1.1.2.)

Example 1.1.13. Find the measure of an angle that is complementary to an angle with $\widehat{\mathrm{U}} = 75°26'33''$.

Solution. What we need to compute is, of course, the difference

$$90° - 75°26'33''.$$

We will clearly need to borrow $60'$ and $60''$, so we will simply re-write (in "one shot") our subtraction as:

$$90° - 75°26'33'' = 89°59'60'' - 75°26'33'',$$

which we can then easily compute in columns:

$$\begin{array}{cccc} & 89° & 59' & 60'' \\ - & 75° & 26' & 33'' \\ \hline = & 14° & 33' & 27'' \end{array}$$

When computing complements, we must of course assume that the measure of the angle we start with does not exceed $90°$. Likewise, when computing supplements, we must of course assume that the measure of the angle we start with does not exceed $180°$. With these observations in mind, we isolate the following type of angles:

(A) A physical angle $\measuredangle \mathrm{U}$ is said to be **acute**, if its measure $\widehat{\mathrm{U}}$ is *positive and less than the measure of a right angle*.

(B) A physical angle $\measuredangle \mathrm{U}$ is said to be **obtuse**, if *its supplement is acute*.

This means that in terms of the measure $\widehat{\mathrm{U}}$, and the unit we use, the above two types can be chracterized as follows.

Unit used	"$\measuredangle \mathrm{U}$ is acute" means:	"$\measuredangle \mathrm{U}$ is obtuse" means:

Turn	0 turn(s) $< \widehat{\mathrm{U}} < \frac{1}{4}$ turn	$\frac{1}{4}$ turn $< \widehat{\mathrm{U}} < \frac{1}{2}$ turn
Radian	$0 < \widehat{\mathrm{U}} < \dfrac{\pi}{2}$	$\dfrac{\pi}{2} < \widehat{\mathrm{U}} < \pi$
Degree	$0^\circ < \widehat{\mathrm{U}} < 90^\circ$	$90^\circ < \widehat{\mathrm{U}} < 180^\circ$

Table 1.1.3

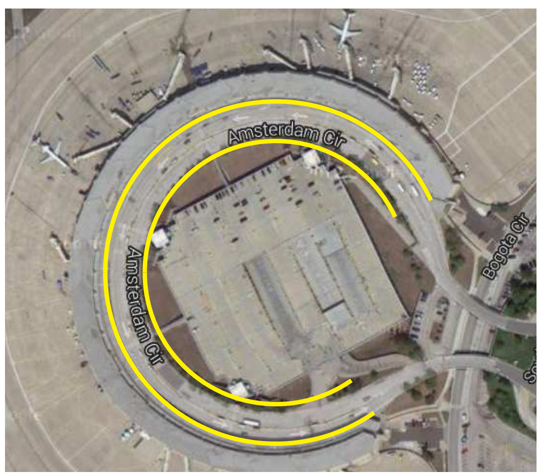
Exercises

The list of problems included here is quite short. An abundant supply of exercises is found in the K-STATE ONLINE HOMEWORK SYSTEM.

1. Find the area of a sector subtended by a central angle of 60° is a disk of radius 2 cm.

2. Given that a circle has circumference 120 ft., find the length of a circular arc subtended by a central angle that measures $\dfrac{\pi}{6}$ [radians].

3*. The figure below depicts one of the terminals at Kansas City International Airport. The two yellow circular arcs mark two sidewalks: the outer sidewalk is immediately outside the terminal; the inner sidewalk is across the (circular arc) street around the parking lot.



When walking from one terminal gate to another terminal gate (both positioned on the outer sidewalk), you have two options:
(A) walk along the outer sidewalk, or
(B) cross the street, walk on the inner sidewalk, then cross the street again.
Here "crossing the street" means to move along the radius of the circles, so whether you choose option (A) or option (B), both sidewalk arcs will be subtended by one and the same physical angle. Your task is to decide, given the radian measure θ of this physical angle, which is your shortest walk. (HINT: Let r and R be the radii of the inner and the outer circles. Compute the distance traveled in both scenarios, in terms of θ, r and R, and compare. You will find that your answer does not depend on r and R!)

The following group of exercises deals with the old fashioned clock, which has two hands: the hour hand and the minute hand.

4. How many degrees does the minute hand travel in one hour? How many degrees does the hour hand travel in one hour?

5. At 12:00 both hands overlap. What are the next three times when the two hands overlap?

6. At 3:00 both hands are pependicular. What are the next three times when the two hands are perpendicular?

7. At 6:00 both hands form a straight angle. What are the next three times when the two hands form a straight angle?

1.2 Right Triangle Trigonometry

A **right triangle** is a triangle in which one of the angles is is a *right angle*. The main features of right triangles are summarized as follows.

Basic Properties of Right Triangles

I. A right triangle has *exactly one right angle*.

II. The side *facing the right angle*, referred to as the **hypotenuse**, is the *longest side* of the triangle.

III. The *sides of the right angle* are called the **legs** of the right triangle.

IV. **Pythagoras' First Theorem.** No matter what units we use, the lengths leg_1, leg_2, hypotenuse, of the three sides of a right triangle satisfy the identity:

$$\text{leg}_1^2 + \text{leg}_2^2 = \text{hypotenuse}^2 \tag{1.2.1}$$

V. The two angles facing the legs are *complementary*. In particular, both these angles are *acute*.

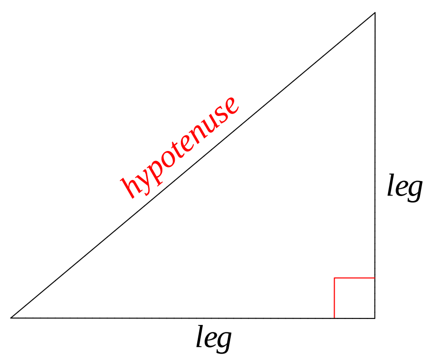


Figure 1.2.1

CLARIFICATION. Concerning property V above we have following important (and more complete) statement: *Any acute angle can be made part of a right triangle.*

The Trigonometric Ratios

Assume V is a vertex in a right triangle $\mathscr{T}$, so that $\measuredangle V$ is *one of the acute angles in* $\mathscr{T}$. In relation to $\measuredangle V$, the two *legs* of the triangle are identified as follows.

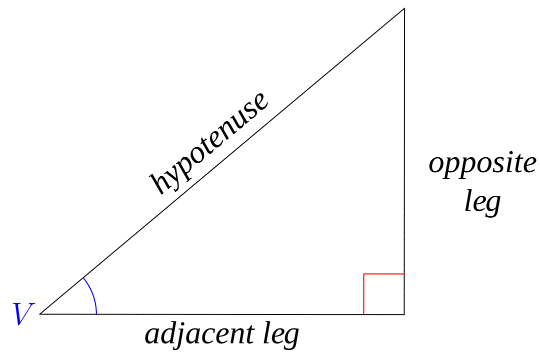


Figure 1.2.2

(i) The leg *facing* V is referred to as the **leg opposite to** $\measuredangle V$.

(ii) The other leg of the triangle – the leg that *contains the vertex* V, is referred to as the **leg adjacent to** $\measuredangle V$.

With this set-up in mind, the *trigonometric ratios of* $\measuredangle V$ are defined as follows:

Assume, as above, that V is a vertex in a right triangle $\mathscr{T}$, so that $\measuredangle V$ is *one of the acute angles in* $\mathscr{T}$.

(A) The **sinus ratio of** $\measuredangle V$ is: $\dfrac{\text{leg opposite to } \measuredangle V}{\text{hypotenuse}}$.

(B) The **secant ratio of** $\measuredangle V$ is: $\dfrac{\text{hypotenuse}}{\text{leg adjacent to } \measuredangle V}$.

(C) The **tangent ratio of** $\measuredangle V$ is: $\dfrac{\text{leg opposite to } \measuredangle V}{\text{leg adjacent to } \measuredangle V}$.

The Trigonmetric Ratios as Functions

Using *similarity* properties of right triangles, one can easily derive the following two important statements.

I. *If two acute angles (in two right triangles) are congruent, then their matching trigonometric ratios are equal.*

II. *Conversely, if two acute angles (in two right triangles) have a pair of equal matching trigonometric ratios, then they are congruent.*

When reading carefully statement I, we see that the trigonometric ratios in effect *only depend on the measures of the acute angle*. For this reason, we can view them as *numerical functions*. To be more precise, we can set up the following definitions and notations.

FACT A. If $\measuredangle V$ is some acute angle (in some right triangle $\mathscr{T}$), then:

(A) the numerical value of the *sinus ratio of* $\measuredangle V$ only depends on the measure $\widehat{V}$, and we denote it by $\sin(\widehat{V})$;

(B) the numerical value of the *secant ratio of* $\measuredangle V$ only depends on the measure $\widehat{V}$, and we denote it by $\sec(\widehat{V})$;

(C) The numerical value of the *tangent ratio of* $\measuredangle V$ only depends on the measure $\widehat{V}$, and we denote it by $\tan(\widehat{V})$.

CLARIFICATIONS. Assuming that the angle measure is presented numerically $\widehat{V} = \boldsymbol{\theta}$, we will feel free to write $\sin(\boldsymbol{\theta})$, $\sec(\boldsymbol{\theta})$, $\tan(\boldsymbol{\theta})$, instead of $\sin(\widehat{V})$, $\sec(\widehat{V})$, $\tan(\widehat{V})$. For instance, if we use radians and say, we present $\boldsymbol{\theta} = \frac{\pi}{4}$, we will simply write $\sin\left(\frac{\pi}{4}\right)$, $\sec\left(\frac{\pi}{4}\right)$, $\tan\left(\frac{\pi}{4}\right)$, but when we use degrees, and we present the same angle measure as $\boldsymbol{\theta} = 45°$, then we will write $\sin(45°)$, $\sec(45°)$, $\tan(45°)$.

Statement II above can now be re-phrased as follows

FACT B. *If $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are two acute angle measures, then the following equalities are equivalent:*

(A) $\sin(\boldsymbol{\theta}) = \sin(\boldsymbol{\theta}')$; (B) $\sec(\boldsymbol{\theta}) = \sec(\boldsymbol{\theta}')$;

(C) $\tan(\boldsymbol{\theta}) = \tan(\boldsymbol{\theta}')$; (D) $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

The Trigonometric Co-Functions

So far, we have only introduced three trigonometric functions: sin, sec, and tan. There are three more functions, which are built according to the following scheme:

If **function** is any one of the functions sin, sec, or tan, its associated **cofunction** is defined as

$$\textbf{cofunction}(\text{Angle}) = \textbf{function}(\textbf{complement of } \text{Angle}).$$

The short-hand notations for the co-functions associated to our three trigonometric functions are as follows: (A) cosin is denoted by cos; (B) cosec is denoted by csc; (C) cotan is denoted by cot.

CLARIFICATION. If we have a right triangle $\triangle VWR$, with the right angle at R, then according to the above rule, the trigonometric cofunctions of $\widehat{V}$ are simply:

$$\cos(\widehat{V}) = \sin(\widehat{W});$$
$$\csc(\widehat{V}) = \sec(\widehat{W});$$
$$\cot(\widehat{V}) = \tan(\widehat{W}).$$

We now have six trigonometric functions of any acute angle measure. Wrapping up all we have learned so far, these trigonometric functions are defined as follow.

Geometric Definitions of Trigonometric Functions for Acute Angles

Assume $\measuredangle V$ is an *acute* physical angle, in some right triangle $\mathscr{T}$, which has measure $\widehat{V} = \boldsymbol{\theta}$. The six trigonometric functions of $\boldsymbol{\theta}$ are given by the following ratios:

$$\sin(\theta) = \frac{\text{leg opposite to } \measuredangle V}{\text{hypotenuse}}; \qquad \cos(\theta) = \frac{\text{leg adjacent to } \measuredangle V}{\text{hypotenuse}};$$

$$\sec(\theta) = \frac{\text{hypotenuse}}{\text{leg adjacent to } \measuredangle V}; \qquad \csc(\theta) = \frac{\text{hypotenuse}}{\text{leg opposite to } \measuredangle V};$$

$$\tan(\theta) = \frac{\text{leg opposite to } \measuredangle V}{\text{leg adjacent to } \measuredangle V}; \qquad \cot(\theta) = \frac{\text{leg adjacent to } \measuredangle V}{\text{leg opposite to } \measuredangle V}.$$

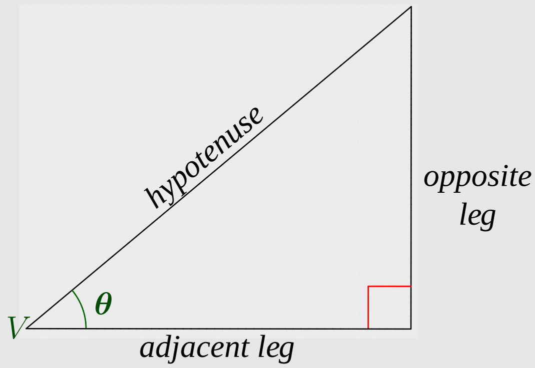


Figure 1.2.3

It is pretty clear that Fact B stated earlier in this section is also true if we include co-functions. In other words, we now have the following statement.

The Fundamental Theorem of Trigonometry for Acute Angles

If θ is an acute angle measure, then the value of one trigonometric function of θ completely determines θ, and consequently the values of the other five trigonometric functions of θ.

"Familiar" Values of Trigonometric Functions

Given some acute angle measure θ, how do we go about to *compute* its six trigonometric functions? Using the Geometric definition, the most reasonable way to do this is to *build a right triangle, in which one of the acute angles has measure θ*. (This is always possible, with the help of the Acute Angle Theorem.) In the examples below we will follow this method, for computing the trigonometric functions of three "familiar" angles.

Example 1.2.1. Consider the acute angle which measures $45°$. When we build a right triangle which has one of its acute angles of measure $45°$, the other acute angle will also measure $45°$, so any such triangle will be an *isosceles right triangle*. Using Pythagoras' First Theorem, it is easy to derive the following well known fact:

In an isosceles right triangle the legs have equal length, and furthermore,

$$\text{hypotenuse} = \sqrt{2} \cdot \text{leg}_1 = \sqrt{2} \cdot \text{leg}_2.$$

In particular, if we denote the leg(s) by ℓ, then an easy calculation of the six ratios yields

$$\sin(45^\circ) = \frac{\ell}{\sqrt{2}\cdot\ell} = \frac{1}{\sqrt{2}}; \qquad \cos(45^\circ) = \frac{\ell}{\sqrt{2}\cdot\ell} = \frac{1}{\sqrt{2}};$$
$$\sec(45^\circ) = \frac{\sqrt{2}\cdot\ell}{\ell} = \sqrt{2}; \qquad \csc(45^\circ) = \frac{\sqrt{2}\cdot\ell}{\ell} = \sqrt{2};$$
$$\tan(45^\circ) = \frac{\ell}{\ell} = 1; \qquad \cot(45^\circ) = \frac{\ell}{\ell} = 1.$$

Example 1.2.2. Another right triangle that is very dear to us is the so-called *half of an equilateral triangle*, in which the acute angles measure 30° and 60°. It is not hard to see that, the sides of such a triangle obey the following pattern:

> *If we denote the leg facing the 30° angle by ℓ, then:*
> - *the leg facing the 60° angle is $\sqrt{3}\cdot\ell$, and*
> - *the hypotenuse is $2\cdot\ell$.*

So if we fix such a triangle, and its acute angles are denoted by $\measuredangle V$ and $\measuredangle W$, labeled such that $\widehat{V} = 30^\circ$ and $\widehat{W} = 60^\circ$, then:

$$\begin{aligned} \text{leg opposite to } \measuredangle V &= \text{leg adjacent to } \measuredangle W = \ell; \\ \text{leg adjacent to } \measuredangle V &= \text{leg opposite to } \measuredangle W = \sqrt{3}\cdot\ell, \end{aligned}$$

and then an easy calculation of the six ratios yields

$$\sin(30^\circ) = \cos(60^\circ) = \frac{\ell}{2\cdot\ell} = \frac{1}{2}; \qquad \cos(30^\circ) = \sin(60^\circ) = \frac{\sqrt{3}\cdot\ell}{2\cdot\ell} = \frac{\sqrt{3}}{2};$$
$$\sec(30^\circ) = \csc(60^\circ) = \frac{2\cdot\ell}{\sqrt{3}\cdot\ell} = \frac{2}{\sqrt{3}}; \qquad \csc(30^\circ) = \sec(60^\circ) = \frac{2\cdot\ell}{\ell} = 2;$$
$$\tan(30^\circ) = \cot(60^\circ) = \frac{\ell}{\sqrt{3}\cdot\ell} = \frac{1}{\sqrt{3}}; \qquad \tan(60^\circ) = \cot(30^\circ) = \frac{\sqrt{3}\cdot\ell}{\ell} = \sqrt{3}.$$

The "Holy Grail" of Trigonometry

We use the above title to designate a collection of formula packages that will ultimately allow us to compute values of trigonometric functions. Starting from this section, we are going to adopt the following:

"Lazy" Notation Conventions

Whenever **there is no danger of confusion**:

I. Parentheses are removed from the notations involving trigonometric functions, so instead of writing "$\sin(\boldsymbol{\theta})$" we simply write "$\sin\boldsymbol{\theta}$."

II. Whenever we raise the value of a trigonometric function to a integer power $p \geq 2$, we put the exponent immediately after the function, so instead of writing "$(\tan\boldsymbol{\theta})^p$" we simply write "$\tan^p\boldsymbol{\theta}$."

The “lazy” power notation II is only reserved for positive exponents.

The first formula package is derived directly from the Geometric Definitions.

The Reciprocal and Ratio Identities

$$\begin{aligned}
\sec\theta &= \frac{1}{\cos\theta}; & \csc\theta &= \frac{1}{\sin\theta};\\
\sin\theta &= \frac{1}{\csc\theta}; & \cos\theta &= \frac{1}{\sec\theta};\\
\sin\theta &= \frac{\tan\theta}{\sec\theta}; & \cos\theta &= \frac{\cot\theta}{\csc\theta};\\
\tan\theta &= \frac{\sin\theta}{\cos\theta}; & \cot\theta &= \frac{\cos\theta}{\sin\theta};\\
\tan\theta &= \frac{1}{\cot\theta}; & \cot\theta &= \frac{1}{\tan\theta}.
\end{aligned}$$

Using these identities, we can easily derive our next formula package which expressed various functions as *products*.

Product Identities

$$\begin{aligned}
\sin\theta &= \tan\theta \cdot \cos\theta; & \cos\theta &= \cot\theta \cdot \sin\theta;\\
\tan\theta &= \sin\theta \cdot \sec\theta; & \cot\theta &= \cos\theta \cdot \csc\theta.
\end{aligned}$$

Our next formula package comes directly from Pythagoras’ Theorem, and is named after him.

The Pythagorean Identities

$$\sin^2\theta + \cos^2\theta = 1; \tag{1.2.2}$$

$$1 + \tan^2\theta = \sec^2\theta; \tag{1.2.3}$$

$$1 + \cot^2\theta = \csc^2\theta. \tag{1.2.4}$$

CLARIFICATIONS. The identity (1.2.2) follows immediately from Pythagoras’ First Theorem by dividing both sides of (1.2.1), which gives:

$$\left[\frac{\text{leg}_1}{\text{hypotenuse}}\right]^2 + \left[\frac{\text{leg}_2}{\text{hypotenuse}}\right]^2 = 1. \tag{1.2.5}$$

Upon multiplying the above identity by $\left[\dfrac{\text{hypotenuse}}{\text{leg}_1}\right]^2$ we get:

$$1 + \left[\frac{\text{leg}_2}{\text{leg}_1}\right]^2 = \left[\frac{\text{hypotenuse}}{\text{leg}_1}\right]^2, \tag{1.2.6}$$

which will clearly give (1.2.3) and (1.2.4).

The Pythagorean Identities are often employed in the direct computations that are based on the following formula package.

The Derived Pythagorean Identities for Acute Angles

Assuming θ is an **acute angle** measurement, the following equalities hold:

$$\cos\theta = \sqrt{1-\sin^2\theta}; \qquad \sin\theta = \sqrt{1-\cos^2\theta};$$
$$\sec\theta = \sqrt{1+\tan^2\theta}; \qquad \csc\theta = \sqrt{1+\cot^2\theta};$$
$$\tan\theta = \sqrt{\sec^2\theta - 1}; \qquad \cot\theta = \sqrt{\csc^2\theta - 1};$$

The Derived Identities are obtained from the Pythagorean Identities by adding or subtracting terms from both sides. For example, if we subtract $\sin^2\theta$ from both sides of (1.2.2) we would get

$$\cos^2\theta = 1 - \sin^2\theta. \tag{1.2.7}$$

If we want to solve for $\cos\theta$, we see that the type of equation we are dealing with is a power equation of the form

$$?^2 = \textit{number}, \tag{1.2.8}$$

which in general has two solutions: $? = \pm\sqrt{\textit{number}}$, so in principle, the correct way to solve (1.2.7) would be:

$$\cos\theta = \pm\sqrt{1-\sin^2\theta}.$$

However, since we only limit ourselves to acute angles, $\cos\theta$ will be positive, thus the only valid solution will be the one that uses the $+$ sign. BOTTOM LINE: As presented here (without $\pm$), the Derived Pythagorean Identities are valid only for acute angles!

The same phenomenon (when the only valid solution of an equation like (1.2.8) is the one with $+$ sign) occurs when we deal with the following geometric versions of the above identities.

The Derived Geometric Pythagorean Identities

$$\text{hypotenuse} = \sqrt{\text{leg}_1^2 + \text{leg}_2^2}; \tag{1.2.9}$$
$$\text{leg}_1 = \sqrt{\text{hypotenuse}^2 - \text{leg}_2^2}; \tag{1.2.10}$$
$$\text{leg}_2 = \sqrt{\text{hypotenuse}^2 - \text{leg}_1^2}. \tag{1.2.11}$$

Computing Values of Trigonometric Functions

The main reason the “Holy Grail” is so significant for us is the fact that it allows us the solve the following type of question.

Basic Trigonometry Problem. *Given one of the values* $\sin\theta$, $\cos\theta$, $\tan\theta$, $\cot\theta$, $\sec\theta$, $\csc\theta$, *find the other five values* .

Presently, we will address this question under the additional assumption that θ is an acute angle measure.

CLARIFICATIONS. Of course, by the Fundamental Theorem of Trigonometry for Acute Angles, we know in fact that one of the six values listed above does in fact *determine θ completely*, but at this time we do not know exactly how this is done. The way we should understand this statement at this point is to simply say that: *knowing one of the values* $\sin\theta$, $\cos\theta$, $\tan\theta$, $\cot\theta$, $\sec\theta$, $\csc\theta$, *for an acute angle measure* θ, *is as good as knowing* θ *itself*. We regard this type of thinking as describing angles implicitly. So if you are told for instance that $\cos\theta = 0.7$, you know pretty much everything about θ, except possibly for the value of θ itself. However, what you know about θ is enough for computing all other five values of the trigonometric functions of θ.

Depending on personal taste and preference, the Basic Trigonometry Problem (for acute angle measures) can be solved by two methods.

Algebraic Method for Solving the Basic Trigonometry Problem for Acute Angles

I. The value of one of the unknown five functions is the reciprocal of the value of the given function. Compute it!
II. Using either the given value, or the one computed in the previous step, compute the value of another one of the unknown functions, using one of the Derived Pythagorean Identities.
III. Upon completing steps I and II you would have the values of three (of the six) trigonometric functions. The remaining three values are obtained using either the Reciprocal/Ratio Identities, or the Product Identities.

Example 1.2.3. Suppose θ is an acute measure angle, and $\sin\theta = \frac{5}{13}$. We will find the remaining five values, using the three steps from the Algebraic Method.

I. Using reciprocals, we immediately find $\csc\theta = \frac{1}{\sin\theta} = \frac{13}{5}$.
II. Using the derived Pythagorean Identities, we can compute

$$\cos\theta = \sqrt{1-\sin^2\theta} = \sqrt{1-\left[\frac{5}{13}\right]^2} = \sqrt{1-\frac{25}{169}} =$$
$$= \sqrt{\frac{169}{169}-\frac{25}{169}} = \sqrt{\frac{169-25}{169}} = \sqrt{\frac{144}{169}} = \frac{\sqrt{144}}{\sqrt{169}} = \frac{12}{13}.$$

III. Find the remaining three values using reciprocals and products:

$$\sec\theta = \frac{1}{\cos\theta} = \frac{13}{12};$$
$$\tan\theta = \sin\theta \cdot \sec\theta = \frac{5}{13}\cdot\frac{13}{12} = \frac{5\cdot 13}{13\cdot 12} = \frac{5}{12};$$
$$\cot\theta = \frac{1}{\tan\theta} = \frac{12}{5}.$$

The second method for solving our problem is based on the geometric approach, which requires that you “cook up” a triangle.

Geometric Method for Solving the Basic Trigonometry Problem for Acute Angles

I. Use the given value of one trigonometric function of $\boldsymbol{\theta}$ to build a right triangle $\mathscr{T}$, in which one trigonometric ratio of one acute angle $\measuredangle V$ matches the given value. Your triangle will have two sides already given by matching. (By the Fundamental Theorem of Trigonometry, this angle is guaranteed to have measure $\widehat{V} = \boldsymbol{\theta}$!)
II. Find the third side in the right triangle $\mathscr{T}$.
III. Compute all other five missing values using the Geometric Definition, which presents each one of them as a trigonometric ratio.

Example 1.2.4. Let us redo Example 1.2.3 using the Geometric Method.

I. We need to build some right triangle $\mathscr{T}$ which has an acute angle $\measuredangle V$ with *sinus ratio* equal to $\frac{5}{13}$. By the definition of the sinus ratio, we can certainly build such a triangle, by prescribing

$$\begin{cases} \text{leg opposite to } \measuredangle V = 5 \\ \text{hypotenuse} = 13 \\ \text{leg adjacent to } \measuredangle V = \text{ (unknown)} \end{cases}$$

II. Using the Derived Geometric Pythagorean Identities it follows immediately that

$$\text{leg adjacent to } \measuredangle V = \sqrt{13^2 - 5^2} = \sqrt{169 - 25} = \sqrt{144} = 12.$$

III. Using the Geometric Definitions, the missing values are:

$$\cos(\boldsymbol{\theta}) = \frac{\text{leg adjacent to } \measuredangle V}{\text{hypotenuse}} = \frac{12}{13}; \qquad \sec(\boldsymbol{\theta}) = \frac{\text{hypotenuse}}{\text{leg adjacent to } \measuredangle V} = \frac{13}{12};$$

$$\csc(\boldsymbol{\theta}) = \frac{\text{hypotenuse}}{\text{leg opposite to } \measuredangle V} = \frac{13}{5}; \qquad \tan(\boldsymbol{\theta}) = \frac{\text{leg opposite to } \measuredangle V}{\text{leg adjacent to } \measuredangle V} = \frac{5}{12};$$

$$\cot(\boldsymbol{\theta}) = \frac{\text{leg adjacent to } \measuredangle V}{\text{leg opposite to } \measuredangle V} = \frac{12}{5}.$$

YOUR CALL! When comparing the two methods, the only difference is that one method used fractions, while the other one did not. Other than that, with bot methods we ended playing with the same numbers: 25, 169, 144 and their square roots. The author of this text prefers the Algebraic Method.

In the preceding Example(s) all answers were left in fraction form. These are the preferred forms, since they give *exact values*. As we have seen already when we computed the trigonometric functions of the "familiar" angles, we also used exact values, in that case using radicals.

Example 1.2.5. Suppose $\boldsymbol{\theta}$ is an acute measure angle, and $\sec\boldsymbol{\theta} = 3$. We will find the remaining five values, using the three steps from the Algebraic Method. (If you are so inclined, try it also with the Geometric Method!)

I. Using reciprocals, we immediately find $\cos\boldsymbol{\theta} = \frac{1}{\sec\boldsymbol{\theta}} = \frac{1}{3}$.
II. Using the derived Pythagorean Identities, we can compute

$$\tan\boldsymbol{\theta} = \sqrt{\sec^2\boldsymbol{\theta} - 1} = \sqrt{3^2 - 1} = \sqrt{9 - 1} = \sqrt{8} = 2\sqrt{2}.$$

III. Find the remaining three values using reciprocals and ratios:

$$\cot\theta = \frac{1}{\tan\theta} = \frac{1}{2\sqrt{2}};$$
$$\sin\theta = \frac{\tan\theta}{\sec\theta} = \frac{2\sqrt{2}}{3};$$
$$\csc\theta = \frac{1}{\sin\theta} = \frac{3}{2\sqrt{2}}.$$

Exercises

In Exercises 1–6 we assume that $\triangle ABC$ is a right triangle, with the right angle at A. You are asked to find all six trigonometric functions of $\widehat{B}$ and $\widehat{C}$, depending on the given information in each Exercise. Use *exact values*. (We also assume a length unit is fixed, so all lengths are given in "lazy" notation, without specifying this unit.)

1. $AB = 3$, $AC = 4$.

2. $AB = 3$, $BC = 4$.

3. $AC = 3$, $BC = 4$.

4*. $AB = 1$, $BC = x$, with $x > 1$. Your answers should be algebraic expressions in x.

5*. $BC = 1$, $AB = x$, with $0 < x < 1$. Your answers should be algebraic expressions in x.

6*. $AB = 3$, $AC = x$, with $x > 0$. Your answers should be algebraic expressions in x.

In Exercises 7–14 we assume that θ is an acute angle measure. Based on the given information in each Exercise, you are asked to find all six trigonometric functions of θ. Use *exact values*.

7. $\sin\theta = \frac{2}{5}$.

8. $\cos\theta = \frac{8}{17}$.

9. $\tan\theta = \frac{24}{7}$.

10. $\cot\theta = 5$.

11. $\sec\theta = \frac{5}{3}$.

12. $\csc\theta = \sqrt{5}$.

13*. $\sec\theta - \tan\theta = \frac{1}{2}$. (HINT: Factor $\sec^2\theta - \tan^2\theta$. and "play" with the Pythogorean Identities. You should be able to get the value for $\sec\theta + \tan\theta$.)

14*. $\sin \boldsymbol{\theta} + \csc \boldsymbol{\theta} = \dfrac{11}{10}$. (HINT: Write the given information as an equation in $\sin \boldsymbol{\theta}$.)

1.3 Solving Right Triangles

In this section we illustrate how Trigonometry provides us with effective tools for solving a key problem in triangle geometry. With what we have learned so far, at this point we are only able to treat right triangles. The case of general triangles will be treated in Chapter 3.

The Problem

What we call the *six elements* of a triangle are:

(A) the *physical lengths* of the three sides;

(B) the *measures* of the three angles.

With this terminology, to **solve a triangle** means to *find all its six elements*. What will be made pretty clear here (as well as in Chapter 3, where we treat *general* triangles) is that the following statement is always true.

Triangle Solving Principle

A triangle can always be solved, given three of its elements, at least one being the length of a side.

CLARIFICATION. When dealing with *right* triangles, one element – the right angle measure – is already given. As it turns out, in this case, it is not in fact necessary to specify all three sides, because using Pythagoras' Theorem, knowing two of them automatically gives the third one. So, when dealing with right triangles, the above statement admits the following concrete formulation.

Right Triangle Solving Principle

A right triangle can be solved, in either one of the following two cases.

I. *The length of one side, together with the measure of one acute angle are given.*

II. *The lengths of two sides are given.*

Case I will be referred to using the phrase Side-Angle; case II will be referred to using the phrase Side-Side.

The methods employed in solving right triangles depend on the case (Side-Angle or Side-Side), as we shall see shortly. However, there are a few common threads, which we summarize in the following list of suggestions.

TIPS:

(A) *Once two sides are found (or given), the third one can be found using the Derived Geometric Pythagorean Identities.* (See formulas (1.2.9), (1.2.10) and (1.2.11) discussed in Section 1.2.)

(B) *Once the measure of one acute angle is found (or given), the measure of the second* *angle can be found using complements*. (If we use degrees, we subtract from $90°$; if we use radians, we subtract from $\frac{\pi}{2}$.)

(C) *For all other calculations involving trigonometric functions, we use their Geometric Definitions as ratios. The preferred functions are* $\sin$, $\cos$, *and* $\tan$.

(D) *For all numerical computations, we use the six basic calculator functions:* $\boxed{\texttt{sin}}$, $\boxed{\texttt{sin}^{-1}}$, $\boxed{\texttt{cos}}$, $\boxed{\texttt{cos}^{-1}}$, $\boxed{\texttt{tan}}$, $\boxed{\texttt{tan}^{-1}}$.

Both here, as well as in Chapter 3, we will use the following:

LABELING CONVENTION. The vertices of a triangle are labeled using uppercase letters, such as A, B, C, etc., thus the angle measures are written as $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, etc. The lengths of the sides are labeled using lowercase letters, such as a, b, c, etc., according to the following letter-matching rule: *Any lowercase letter designates the length of the side that faces the vertex labeled by the matching uppercase letter*.

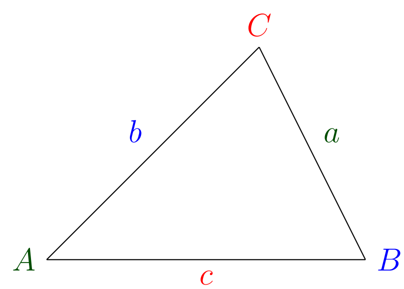


Figure 1.3.1

For example, if we have a triangle $\triangle ABC$, the lowercase letters denote the physical lengths $a = BC$, $b = AC$ and $c = AB$.

The Side-Angle Case

Given one side and one acute angle, we solve a right triangle using the following steps.

Method for Solving the Side-Angle Problem for Right Triangles

I. *Find the other acute angle using complements.*

II. *Find the unknown sides, using a proportion equations constructed using the Geometric Definition of two (preferred) trigonometric functions of one of the two acute angles which involve the given side and an unknown side.*

II'. *If desired, find only one unknown side using step II, then find the second unknown side using Pythagoras' Theorem (see formulas* (1.2.9), (1.2.10) *and* (1.2.11) *from section 1.2.).*

CLARIFICATION. When working on Step II, we will encounter *proportion equations* of the form "*ratio* = *number*," in which the unknown quantity may be either the numerator, or the denominator. In order to solve such equations, we will use one of the following "recipes" that use *equivalent forms of proportions*:

A. To solve an equation of the form $\dfrac{?}{\#} = \textit{number}$, we *multiply*: $? = \# \cdot \textit{number}$.

B. To solve an equation of the form $\dfrac{\#}{?} = \textit{number}$, we *divide*: $? = \dfrac{\#}{\textit{number}}$.

Example 1.3.1. Suppose $\triangle ABC$ has $\widehat{A} = 90°$, $\widehat{B} = 40°$ and $a = 2$ cm.

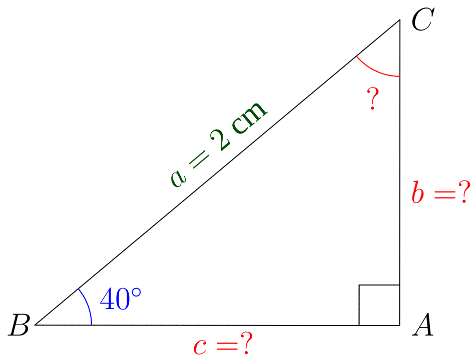


Figure 1.3.2

We solve the triangle, following the two-step method.

I. The other acute angle $\measuredangle C$, is *complementary to* $\measuredangle B$, so its measure is: $\widehat{C} = 90° - \widehat{B} = 90° - 40° = 50°$.

II. To find the remaining two sides b and c, we seek two *trigonometric functions of* $\widehat{B}$ (among the *preferred* ones), to which the given side a "contributes." At this point, we note that:

- the given side $a = 2$ cm is facing the right angle $\measuredangle A$, so it is the *hypotenuse*;
- the unknown side b is the *leg opposite to* $\measuredangle B$;
- the unknown side c is the *leg adjacent to* $\measuredangle B$.

Based on these observations, the functions we identify are

$$\sin \widehat{B} = \frac{b}{a} \quad \text{and} \quad \cos \widehat{B} = \frac{c}{a},$$

so our proportion equations are:

$$\frac{b}{2} = \sin 40° \quad \text{and} \quad \frac{c}{2} = \cos 40°.$$

Using equivalent forms of these proportion equations, we immediately get

$$b = 2 \cdot \sin 40° \simeq 1.285575219 \text{ cm},$$
$$c = 2 \cdot \cos 40° \simeq 1.532088886 \text{ cm}.$$

These calculations were done on a TI-84 calculator, on which we typed:

```
2*sin(40)
      1.285575219
2*cos(40)
      1.532088886
```

Make sure you set your calculator to work in degrees!

Example 1.3.2. Suppose $\triangle ABC$ has $\widehat{A} = 31°23'$, $\widehat{B} = 90°$ and $a = 5.6$ cm.

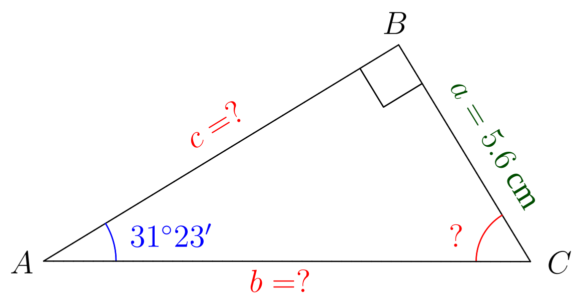


Figure 1.3.3

We solve the triangle, again following the two-step method.

I. The other acute angle $\measuredangle C$, is *complementary to* $\measuredangle A$, so its measure is: $\widehat{C} = 90° - \widehat{A} = 90° - 31°23' = 58°37'$.

II. To find the remaining two sides b and c, we seek two *trigonometric functions of* $\widehat{A}$ (among the *preferred* functions), to which the given side a "contributes." Note that:

- the given side $a = 2$ cm is the *leg opposite to* $\measuredangle A$
- the unknown side b is facing the right angle $\measuredangle A$, so it is the *hypotenuse*;
- the unknown side c is the *leg adjacent to* $\measuredangle A$.

Based on these observations, the functions we can use are

$$\sin \widehat{A} = \frac{a}{b} \quad \text{and} \quad \tan \widehat{A} = \frac{a}{c},$$

so our proportion equations are:

$$\frac{5.6}{b} = \sin(31°23') \quad \text{and} \quad \frac{5.6}{c} = \tan(31°23').$$

Using equivalent forms of these proportion equations, we immediately get

$$b = \frac{5.6}{\sin(31°23')} \simeq 10.7534868 \text{ cm},$$
$$c = \frac{5.6}{\tan(31°23')} \simeq 9.180276592 \text{ cm},$$

When doing these calculations on a TI-84 calculator, we typed:

```
5.6/sin(31+23/60
)
        10.7534868
5.6/tan(31+23/60
)
       9.180276592
```

Notice that, in order to avoid error compounding, each calculation was carried on in "one shot," so for the trigonometric functions ($\sin$ and $\tan$) of $31°23'$ were we typed `sin(31+23/60)` and `tan(31+23/60)`.

The Side-Side Case

When solving a right triangle given two sides, we will have to use three new calculator functions:[1] $\boxed{\texttt{sin}^{-1}}$, $\boxed{\texttt{cos}^{-1}}$, and $\boxed{\texttt{tan}^{-1}}$. The way these calculator functions work (which is consistent with the Fundamental Theorem of Trigonometry for Acute Angles) is summarized as follows:

THE INVERSE TRIGONOMETRIC CALCULATOR FUNCTIONS. Assume ***number*** is some positive number.

(A) **If *number* < 1**, the result of $\boxed{\texttt{sin}^{-1}(\textit{number})}$ is the *unique acute angle measure, which is a solution of the equation:*

$$\sin ? = \textit{number}. \tag{1.3.1}$$

(B) **If *number* < 1**, the result of $\boxed{\texttt{cos}^{-1}(\textit{number})}$ is the *unique acute angle measure, which is a solution of the equation:*

$$\cos ? = \textit{number}. \tag{1.3.2}$$

(C) The result of $\boxed{\texttt{tan}^{-1}(\textit{number})}$ is the *unique acute angle measure, which is a solution of the equation:*

$$\tan ? = \textit{number}. \tag{1.3.3}$$

Statements (A) and (B) **are only true, if *number* < 1**. This issue will be revisited in Chapter 2, where these calculator functions will be thoroughly investigated, along with the basic trigonometric equations (1.3.1), (1.3.2), (1.3.3).

Given two sides, we solve a right triangle using the following steps.

Method for Solving the Side-Side Problem for Right Triangles

I. *Find one of the unknown acute angles, by:*
   - *computing one of its (preferred) trigonometric function as a ratio put together using the given sides, then*
   - *solve the associated basic trigonometric equation – one of the form* (1.3.1), *or* (1.3.2), *or* (1.3.3) *– using the appropriate inverse trigonometric calculator function.*

II. *Find the remaining elements (the other acute angle and the third side) using the steps given in the Side-Angle Problem.*

Example 1.3.3. Suppose $\triangle ABC$ has $\widehat{A} = 90°$, $a = 12.3$ cm and $b = 7.5$ cm.

[1] On a TI-84, these functions are accessed using "$\boxed{\texttt{2ND}}\boxed{\texttt{sin}}$," "$\boxed{\texttt{2ND}}\boxed{\texttt{cos}}$," and "$\boxed{\texttt{2ND}}\boxed{\texttt{tan}}$." On other brands instead of $\boxed{\texttt{2ND}}$, one uses the $\boxed{\texttt{INV}}$ key.

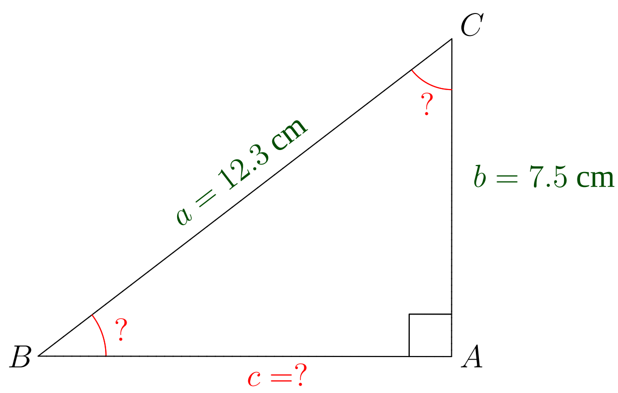


Figure 1.3.4

We solve the triangle, following the method outlined above.

I. The first given side is a, which faces the right angle, so $a = 12.3$ cm is the *hypotenuse*. The second given side is $b = 7.5$ cm, which is the *leg opposite the acute angle $\measuredangle B$*, so using these two sides, we can set

$$\sin \widehat{B} = \frac{b}{a} = \frac{7.5}{12.3},$$

and then, using the `sin⁻¹` function on the calculator (see below) we get $\widehat{B} \simeq 37.57186932°$.

II. The other acute angle $\measuredangle C$, is *complementary to $\measuredangle B$*, so its measure is:

$$\widehat{C} = 90° - \widehat{B} \simeq 90° - 37.57186932° \simeq 52.42813068°.$$

To find the third side c, which is a *leg*, we can use (for a nice computation flow) the sine of the angle $\widehat{C}$ which we just found

$$\sin \widehat{C} = \frac{c}{a},$$

which becomes

$$\sin 52.42813068° = \frac{c}{12.3},$$

from which by cross-multiplication we get:

$$c = 12.3 \cdot \sin 52.42813068° \simeq 9.748846086 \text{ cm}.$$

,

The three calculations above were done on a TI-84 by typing:

```
sin⁻¹(7.5/12.3)
     37.57186932
90-Ans
     52.42813068
12.3*sin(Ans)
     9.748846086
```

Example 1.3.4. Suppose $\triangle ABC$ has $\widehat{C} = 90°$, $a = 6$ cm and $b = 8$ cm.

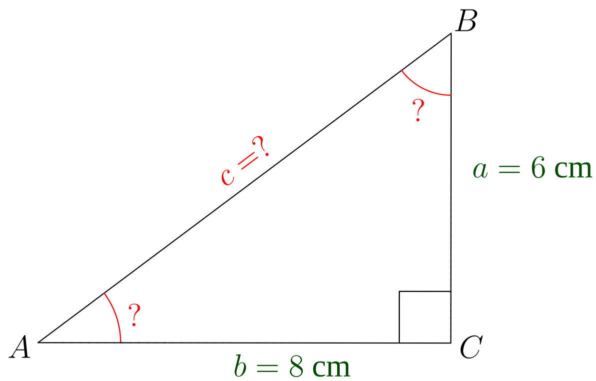


Figure 1.3.5

Again, follow the method outlined above.

I. The two given sides a and b are both *legs*. Since $a = 6$ cm, which is the *leg opposite to* $\measuredangle A$, and $b = 8$ cm, which is the *leg adjacent to* $\measuredangle A$, so using these two sides, we can set

$$\tan \widehat{A} = \frac{a}{b} = \frac{6}{8},$$

and then, using the $\boxed{\texttt{tan}^{-1}}$ function on the calculator, we get $\widehat{A} \simeq 36.86989765°$.

II. For a change, we choose to find the missing side c using the Pythagorean formula (1.2.9):

$$c = \text{hypotenuse} = \sqrt{\text{leg}_1{}^2 + \text{leg}_2{}^2} = \sqrt{6^2 + 8^2} = \sqrt{100} = 10 \text{ cm}.$$

The other acute angle $\measuredangle B$, is *complementary to* $\measuredangle A$, so its measure is:

$$\widehat{B} = 90° - \widehat{A} \simeq 90° - 36.86989765° \simeq 53.13010235°.$$

Applications

With the techniques we developed up to this point, we are now able to solve a variety of problems, many of which have real life applications. The types of problems we are going to explore are what young students describe as "word problems." Since the phrase "word problem" appears a bit inappropriate to describe what is going on, we prefer to replace it with the phrase: *"practical problem."*

TIPS/STEPS FOR SOLVING PRACTICAL PROBLEMS.

(A) Draw a picture or diagram.

(B) Name (using symbols) all unknown quantities. Don't be modest! Use a lot of letters!

(C) Write down *all algebraic relations that link the given and the unknown quantities*. Many of these relations are obtained by *identifying all the right triangles* that the picture/diagram produces. Upon completing this step, you should have only an ALGEBRA PROBLEM to "play with."

(D) *Make a plan on how to solve your* ALGEBRA PROBLEM, *and execute your plan!*

(E) Carefully write down you answer, *as demanded by the problem*. (For instance, if you are asked to find a *length*, specify the units!)

Example 1.3.5. A radio 100 ft tower tower is anchored with several wires, which are attached to the top and form a 35° angle, as shown in the picture below (where only one anchor is shown; in real life, at least three such anchors are needed):

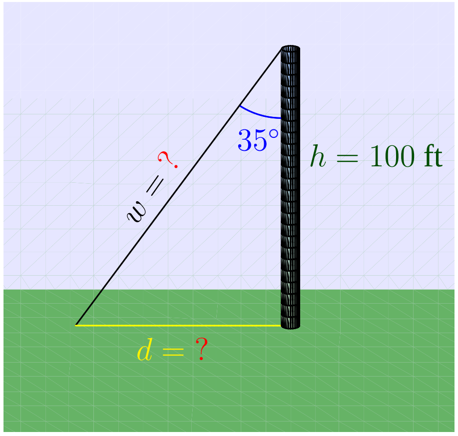


Figure 1.3.6

We are asked to determine the length of each wire, as well as the distance from the base of the tower to the point where the wire(s) are anchored in the ground.

Solution. As the tower is vertical, the above diagram reveals a right triangle, which has an acute angle (which we may call $\measuredangle V$, if we like) measuring 35°, to which the three sides are related as follows:

- the side $h = 100$ ft (the height of the tower) represents the *leg adjacent to* $\measuredangle V$;
- the side labeled d (the distance from the base of the tower to the anchor point on the ground) represents the *leg opposite to* $\measuredangle V$;
- the side labeled w (the length of the wire) represents the *hypotenuse*.

Clearly, this looks like a Side-Angle Problem, so we can set two trigonometric functions of $\widehat{V} = 35^\circ$ as

$$\cos\widehat{V} = \frac{h}{w} \quad \text{and} \quad \tan\widehat{V} = \frac{d}{h},$$

so when we replace all given numbers we get our ALGEBRA PROBLEM:

$$\begin{cases} \dfrac{100}{w} = \cos 35^\circ \\[2ex] \dfrac{d}{100} = \tan 35^\circ \end{cases} \tag{1.3.4}$$

All we need now is to solve the ALGEBRA PROBLEM, which should be pretty easy in this situation: we find w by division; we find d by multiplication:

$$w = \frac{100}{\cos 35^\circ} \simeq 122.0774589 \text{ ft}.$$
$$d = 100 \cdot \tan 35^\circ \simeq 70.02075382 \text{ ft}.$$

So our complete answer is: we need approximately 122 ft. for each wire anchor, and each wire is anchored at about 70 ft from the base of the tower.

Trigonometry has numerous applications in *surveying*. Essentially what a surveyor is capable to do is to *measure angles between various lines of sight and the level line*.

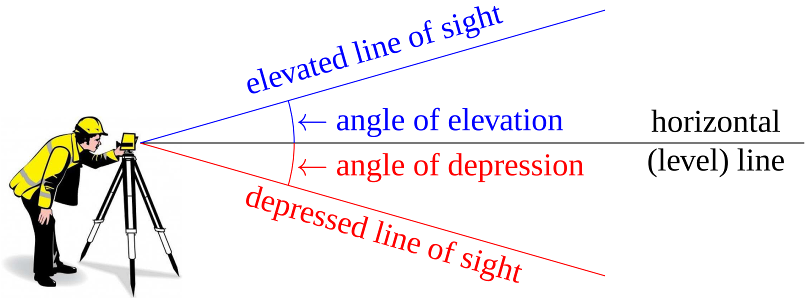


Figure 1.3.7

Depending on the position of the line of sight with respect to the level line, the angles measured are either called

- *angles of elevation*, if the line of sight is *above* the level line, or
- *angles of depression*, if the line of sight is *below* the level line.

Example 1.3.6. Assume we have tower, on top of which a 20 ft antenna sits, as shown in the picture below. From an observation point on ground level, we measure the angles of elevation to the bottom and to the top of the antenna, and our measurements read $26.6°$ and $32°$.

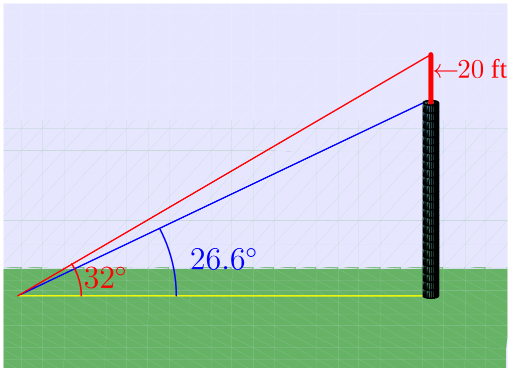


Figure 1.3.8

Based on these two measurements, we can in fact compute both the height x of the tower and the distance y from the observation point to the base of the tower.

To see how we do this, we start off by breaking up our original figure into two separate right triangles, depicted below.

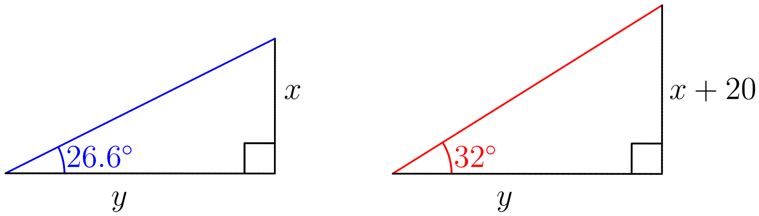


Figure 1.3.9

The left triangle includes the elevation angle to the bottom of the antenna, so we can write:

$$\frac{x}{y} = \tan 26.6^\circ. \tag{1.3.5}$$

The triangle on the right includes the elevation angle to the top of the antenna, so the leg opposite to the 32° angle is longer than x, because we must add 20 (ft) that accounts for the height of the antenna, so now we can write:

$$\frac{x+20}{y} = \tan 32^\circ. \tag{1.3.6}$$

By transforming now both equations (1.3.5) and (1.3.6) using multiplication, we reach our ALGEBRA PROBLEM, which now looks like a *system of equations:*

$$\begin{cases} x = y \cdot \tan 26.6^\circ \\ \\ x + 20 = y \cdot \tan 32^\circ \end{cases} \tag{1.3.7}$$

To solve this system we plan to do the following:

(i) Subtract first equation from the second, so we will eliminate x. This will produce and equation with only one unknown y, which we can solve.

(ii) After we find y, we use the first equation to compute x.

Let us now execute the above plan. Upon subtracting first equation from the second equation, we will get

$$20 = y \cdot \tan 32^\circ - y \cdot \tan 26.6^\circ = y \cdot (\tan 32^\circ - \tan 26.6^\circ),$$

which looks like: $20 = y \cdot$ *number*, so we can find y by division:

$$y = \frac{20}{\tan 32^\circ - \tan 26.6^\circ} \simeq 161.1517137 \text{ ft}.$$

Using this value, we can compute x:

$$x = y \cdot \tan 26.6^\circ = \frac{20}{\tan 32^\circ - \tan 26.6^\circ} \cdot \tan 26.6^\circ \simeq 80.6987669 \text{ ft}.$$

These calculations above were done on a TI-84 by typing:

```
20/(tan(32)-tan(
26.6))
     161.1517137
Ans*tan(26.6)
      80.6987669
```

(Notice that we did our computations in “one shot,” which is always a good practice, when we need to worry about precision.)

So our complete answer is: the tower is approximately 161 ft. tall, and the observer is at about 81 ft from the base of the tower.

Exercises

The list of problems included here is quite short. An abundant supply of exercises is found in the K-STATE ONLINE HOMEWORK SYSTEM. Except for Exercise 10, *round all your answers to three decimal places*.

1. From an observation point 10 meters above ground level, a surveyor measures the depression angle to an object on the ground and the measurement reads $20.5°$. Find the distance from the object to the point directly beneath the observation point.

2. Suppose an amateur radio enthusiast wants to build a small radio tower, which needs to be anchored at an angle of $40°$. As in Example 1.3.5, the angle referred to here is the angle formed by the tower and the anchor wire. Given that, three anchors are needed, and 345 ft of wire are available, how tall can the radio tower be?

3. From an observation point on the ground, 500 yards from a launching site of a weather balloon, an observer measures the angles of elevation of the balloon at two different times: 1 minute after launch, and 2 minutes after launch. At the 1-minute mark, the measurement reads $19.2°$; at the 2-minute mark, the measurement reads $24.7°$. Estimate the vertical distance traveled by the balloon between the 1- and the 2-minute marks.

4. From an observation point on the ground, the angle of elevation to the top of a very tall tree measures $50°$.

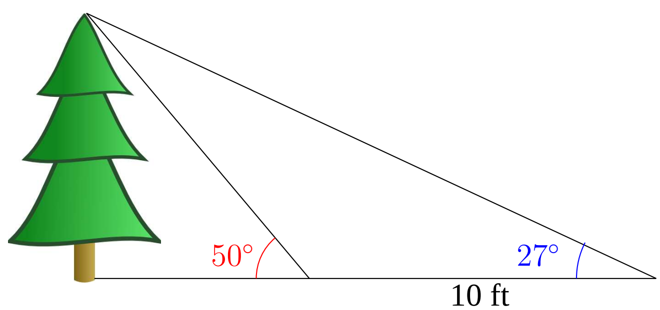
   

   Figure 1.3.10

   We move 10 ft further away from the tree and measure again the angle of elevation to the top of the tree, which now shows $27°$. How tall is the tree?

5. Suppose you live in Manhattan KS and one day in June at 10 a.m. you look at the Sun (with some special protective glasses!) and measure its angle of elevation, which reads $39°$. (Since the Sun is very very far, this reading will be the same for everybody in Manhattan.) Assuming you have a 20 ft flag-pole, find the length of its shadow.

Right triangles appear naturally as "halves" of isosceles triangles, so they can be used for solving such triangles. Exercises 6-8 illustrate this technique.

6. Solve the triangle $\triangle ABC$, given $a = b = 10$ cm, and $\widehat{C} = 40°$. (HINT: Let M be the midpoint of $\overline{AB}$. Solve the *right* triangle $\triangle AMC$.)

7. Solve the triangle $\triangle ABC$, given $a = b = 7$ cm, and $\widehat{A} = 62°$. (Same hint as above.)

8. Solve the triangle $\triangle ABC$, given $a = b = 10$ cm, and $c = 8$ cm. (Same hint as above.)

9. From an observation point P on the ground, you are looking at a spherical balloon of 1 foot radius flying in the air, and you are able to measure the angle between the lines of sight of both "ends" as shown below.

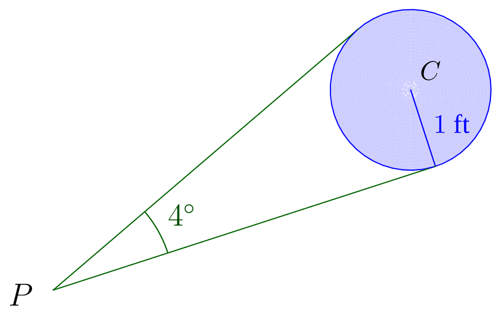


Figure 1.3.11

If your measurement reads $4°$, how close is the balloon to you? Compute first the distance PC from the observation point to the center of the balloon, then subtract the radius.

With very few exceptions, when computing values of trigonometric functions of angles, we always have to rely on a calculator. At the same time, we know there are certain angles, such as $30°$, $45°$, and $60°$, for which we can compute the values of the trigonometric functions *by hand*, thus we can compute *exact values*. The Exercise below outlines the calculation of the *exact values* of the trigonometric functions of $15°$.

10*. Start with a square $ABCD$ with side 1 inch. Then build an equilateral triangle $\triangle ABP$, where P sits inside the square. Let M be the midpoint of $\overline{AB}$ and let N be the midpoint of $\overline{CD}$.

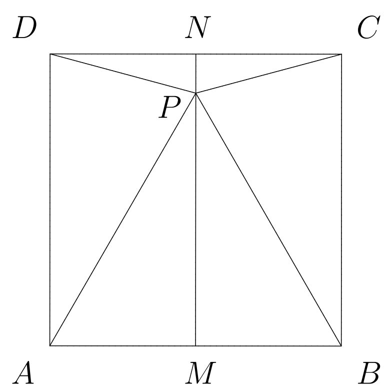


Figure 1.3.12

The three points M, N and P all sit on a line, which is perpendicular to both $\overline{AB}$ and $\overline{CD}$, so we have four right triangles: $\triangle AMP$, $\triangle BMP$ (both being halves of the equilateral triangle $\triangle APB$), and $\triangle CNP$, $\triangle DNP$.

(i) Find the angles of the isosceles $\triangle PBC$. (You already know the angle $\widehat{B} = 30°$.)
(ii) Find the angles in the right triangle $\triangle PCN$.
(iii) Solve the right triangle $\triangle PMB$. Use *exact values*.
(iv) Find the *exact value* of $PN = MN - PM$. Since you also know that $NC = \frac{1}{2}$, you can now also find the *exact value* of PC.
(v) Write down the *exact values* of the trigonometric functions of the angle $\widehat{C}$ from $\triangle PCN$.

Chapter 2

Trigonometry Beyond Acute Angles

In this Chapter we extend the trigonometric functions beyond acute angle measures, and we study them in detail, by focusing on their algebraic features.

2.1 Basic Notions of Analytic and Vector Geometry

In preparation for our next development of Trigonometry, we need to set up the adequate Geometry framework, which in our case is what we call *Analytic Geometry*. In a nutshell, our goal is to do Geometry in coordinates.

Rectangular and Square Coordinate Systems

We construct a **rectangular coordinate system** in the plane as follows:

- we fix two perpendicular lines, the intersection of which we call the *origin*;
- we "coordinatize" each line[2] with the help a *unit length*, such that *the origin has coordinate zero on both lines*.

With this set-up, our two lines are referred to as the *coordinate axes*. It is customary to designate one of the lines as the x-*axis*, and the other one as the y-*axis*. (It is up to us to decide what symbols we use: we do not always have to stick to x and y. It is also up to us to decide which line is responsible for which coordinate.)

Assuming all this set-up, the *coordinates* of a point P in the plane are obtained as follows:

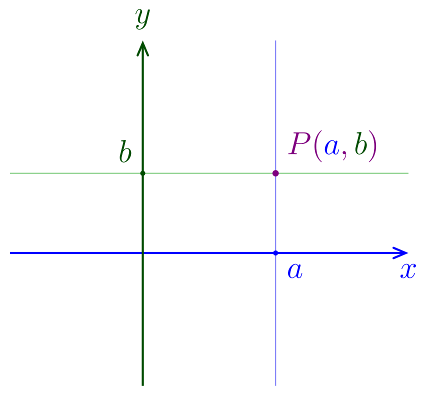


Figure 2.1.1

(A) We take the (unique) line through P, which is *perpendicular to the* x-*axis*, and pick up the point where this line intersects the x-axis. The coordinate on this point (on the x-axis) is what we call the x-*coordinate of* P.

[2] To "coordinatize" a line $\mathscr{L}$ using a length unit means to establish a *1-1 correspondence* between the points on $\mathscr{L}$ and the set of all real numbers, called the *coordinate correspondence*, so that for any two points A and B on $\mathscr{L}$, we have the equlity: $\text{dist}(A, B) = \left|(\textit{coordinate of } A) - (\textit{coordinate of } B)\right|$ units.

(B) We take the (unique) line through P, which is *perpendicular to the y-axis*, and pick up the point where this line intersects the y-axis. The coordinate on this point (on the y-axis) is what we call the *y-coordinate of P*.

As shown in Figure 2.1.1, we specify both coordinates of P using *ordered pairs* of real numbers. Conversely, given any ordered pair (a, b), there is exactly one point that has this pair as its coordinates.

The length units used for "coordinatizing" the axes need not be the same! For example, when we use a graphing calculator such as the TI-84, in the "standard" graphing mode, the length unit on the y-axis is slightly smaller than the one used on the x-axis. (Roughly, in "standard" mode the y-axis unit is $2/3$ of the x-axis unit.) In case when we use the *same length unit on both axes*, we say that our coordinate system is a **square coordinate system**. (On a graphing calculator, this can be obtained using the "square" display mode.)

Square coordinate systems are particularly useful, because they allow us to compute *distances*.

The Distance Formula in Square Coordinates

Assume a common length **unit** *is used for building a square coordinate system. Given two points, written in coordinates as $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$, the physical distance between them is:*

$$P_1P_2 = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \text{ units}. \tag{2.1.1}$$

Using coordinates, we can describe various sets of points in the plane using *equations in two variables*, which are presented in the form

$$\textit{Expression in } x \textit{ and } y = \textit{Expression in } x \textit{ and } y.$$

For example, using the Distance Formula in Square Coordinates, we have the following coordinate equation for *circles*.

The Equation of a Circle in Square Coordinates

Assume a common length **unit** *is used for building a square coordinate system. Given a point written in coordinates $Z(a, b)$, and some positive real number r, the equation of the circle of radius r* **units**, *centered at Z, is:*

$$(x - a)^2 + (y - b)^2 = r^2. \tag{2.1.2}$$

Example 2.1.1. Assume we work in a square coordinate system. Consider the equation:

$$4x^2 - 8x + 4y^2 + 6y = 6. \tag{2.1.3}$$

As it turns out, this equation does represent a circle. To see how this comes about, we are going to form two groups in the left-hand side of the given equation:

$$\big[\underbrace{4x^2 - 8x}_{\text{(I)}}\big] + \big[\underbrace{4y^2 + 6y}_{\text{(II)}}\big] = 6, \tag{2.1.4}$$

and to work on each of the two expressions individually, using the *easy square completion identity*:

$$\boldsymbol{a}\boldsymbol{t}^2+\boldsymbol{b}\boldsymbol{t}+\boldsymbol{c}=\boldsymbol{a}\left(\boldsymbol{t}+\frac{\boldsymbol{b}}{2\boldsymbol{a}}\right)^2-\frac{\boldsymbol{D}}{4\boldsymbol{a}}. \tag{2.1.5}$$

In formula (2.1.5), the symbol $\boldsymbol{t}$ designates some *variable*, and $\boldsymbol{D}=\boldsymbol{b}^2-4\boldsymbol{a}\boldsymbol{c}$ is the *discrimininant*.

In (I), the leading coefficient is 4, the middle coefficient is -8 (and the constant term is 0), so the discriminant is $D=(-8)^2=64$, so after completing the square, this expression is

$$4x^2-8x=4\left(x+\frac{-8}{2\cdot 4}\right)^2-\frac{64}{4\cdot 4}=4(x-1)^2-4. \tag{2.1.6}$$

In (II), the leading coefficient is 4, the middle coefficient is 6 (and the constant term is 0), so the discriminant is $D=6^2=36$, so after completing the square, this expression is

$$4y^2+6y=4\left(y+\frac{6}{2\cdot 4}\right)^2-\frac{36}{4\cdot 4}=4\left(y-\frac{3}{4}\right)^2-\frac{9}{4}. \tag{2.1.7}$$

When go back to (2.1.4) and replace (I) and (II) using the above two identities, the equation becomes:

$$\underbrace{4(x-1)^2-4}_{\text{(I)}}+\underbrace{4\left(y+\frac{3}{4}\right)^2-\frac{9}{4}}_{\text{(II)}}=6.$$

We now add $4+\dfrac{9}{4}$ to both sides, so our equation becomes:

$$4(x-1)^2+4\left(y+\frac{3}{4}\right)^2=4+\frac{9}{4}+6. \tag{2.1.8}$$

Will continue to transform the above equation, first by simplifying the right-hand side

$$4+\frac{9}{4}+6=10+\frac{9}{4}=\frac{40}{4}+\frac{9}{4}=\frac{49}{4},$$

so now the equation (2.1.8) reads:

$$4(x-1)^2+4\left(y+\frac{3}{4}\right)^2=\frac{49}{4}. \tag{2.1.9}$$

Finally, when we divide all terms in this equation by 4, we get an equation that matches perfectly with (2.1.2):

$$\begin{array}{ccccc} (x-1)^2 & + & \left(y+\dfrac{3}{4}\right)^2 & = & \dfrac{49}{16} \\ \updownarrow & & \updownarrow & & \updownarrow \\ (x-a)^2 & + & (y-b)^2 & = & r^2 \end{array} \tag{2.1.10}$$

The above match yields $a = 1$, $b = -\frac{3}{4}$, and $r^2 = \frac{49}{16}$, thus[3] $r = \sqrt{\frac{49}{16}} = \frac{\sqrt{49}}{\sqrt{16}} = \frac{7}{4}$.

We conclude that the equation (2.1.3) represents a *circle of radius $\frac{7}{4}$ units, centered at the point* $Z\left(1, -\frac{3}{4}\right)$.

▶ Now work Exercises 1-6.

Directions and Vectors

Suppose you are driving in a big flat desert, where there are no roads, but you have a good vehicle that allows you to go from any point straight to any other point. To know "where you are" at any given time during your trip, you use a GPS device that will show your position on a map. The map, of course, is nothing else than a coordinate system! However, when you want to go from one point to another, you need a sense of *direction*, for which you would need a *compass*. Although these notions are intuitively quite clear, the actual mathematical definitions are quite elaborate.

Assume we have two rays $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, which emanate from points P_1 and P_2. We say that these two rays are **parallel**, if they *sit on two parallel lines $\mathscr{L}_1$ and $\mathscr{L}_2$*. If this is the case, then consider the line $\mathscr{M}$ that passes through P_1 and P_2, and depending on the positions of the two rays relative to this line, we say that

- $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are **directly parallel**, if they *sit on the same side of $\mathscr{M}$* (see Figure 2.1.2);
- otherwise, if the two rays *sit on opposite sides of $\mathscr{M}$*, we say that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are **opposite parallel** (see Figure 2.1.3).

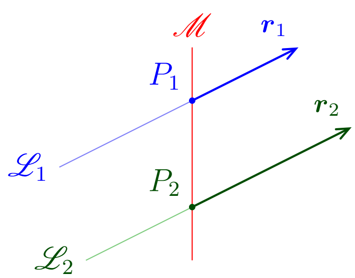


Figure 2.1.2

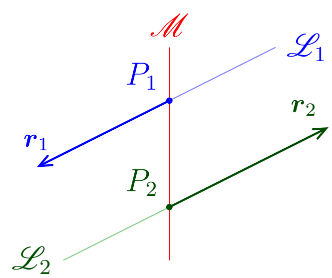


Figure 2.1.3

We adapt the above definition to the case when the two rays sit on the *same line*, as follows.

Assume we have two rays $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, which *both sit on one line $\mathscr{L}$*.

- We say that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are **directly aligned**, if *one of them is contained in the other*;
- otherwise, we say that the two rays are **opposite aligned**.

CLARIFICATION. The figure below depicts four rays $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$, $\boldsymbol{r}_4$, all sitting on a line $\mathscr{L}$, which can be paired as follows

(A) The rays $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are *directly aligned*; likewise, the rays $\boldsymbol{r}_3$ and $\boldsymbol{r}_4$ are *directly aligned*.

[3] Of course, the equation $r^2 = \frac{49}{16}$ has two solutions $r = \pm\sqrt{\frac{49}{16}}$. However, since we are dealing with *positive* quantities, we will only choose the $+$ sign.

(B) The following four pairs of rays are *opposite aligned*: (i) $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$; (ii) $\boldsymbol{r}_1$ and $\boldsymbol{r}_4$; (iii) $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$; (iv) $\boldsymbol{r}_2$ and $\boldsymbol{r}_4$.

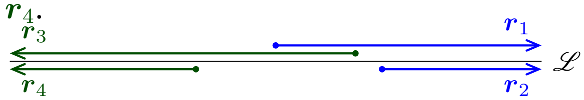


Figure 2.1.4

Of course, any ray $\boldsymbol{r}$ is directly aligned to itself, and opposite aligned to the opposite ray $\boldsymbol{r}^{\text{op}}$.

Assume two rays $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are given.

(A) We say that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ **point in the same direction**, if either
- *$\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are directly parallel*, or
- *$\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are directly aligned*.

(A) We say that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ **point in opposite directions**, if either
- *$\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are opposite parallel*, or
- *$\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are opposite aligned*.

We have not really defined what *direction* actually means! The correct way to define this notion is to take what mathematicians call *equivalence classes*. Since this concept is beyond the scope of a traditional Trigonometry course, we are going to circumvent it using the following gadget.

A **compass** is a *circle* sitting somewhere in the plane. The main feature of the compass is that, *for any ray $\boldsymbol{r}$ in the plane, there is exactly one ray $\boldsymbol{r}'$ on the compass, which points in the same direction as $\boldsymbol{r}$*.

CLARIFICATIONS. A *ray on the compass* means, of course, a *ray that emanates from the center of the compass*. Any such ray, is completely determined by a *point on the circle*.

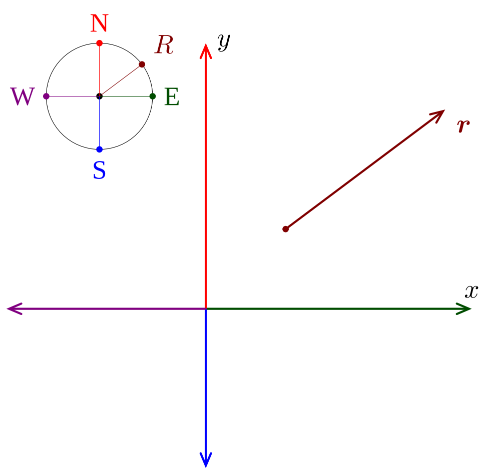


Figure 2.1.5

The figure above depicts, in the presence of a *square coordinate system*, a ray $\boldsymbol{r}$, and the point (denoted by R) which determines the direction of $\boldsymbol{r}$ on the compass.

Following the traditional conventions used in topography, in the presence of a *square coordinate system*, four particular directions are given special names as follows: the direction of the *positive x-axis* is called East; the direction of the *negative x-axis* is called West; the direction of the *positive y-axis* is called North; the direction of the *negative y-axis* is called South.

Using a compass, we can also specify any direction using the sailor's **bearing** notation, which specifies an *acute angle measure* from either North or South direction *towards* either the East or West direction. Four such directions are depicted in Figure 2.1.6 below.

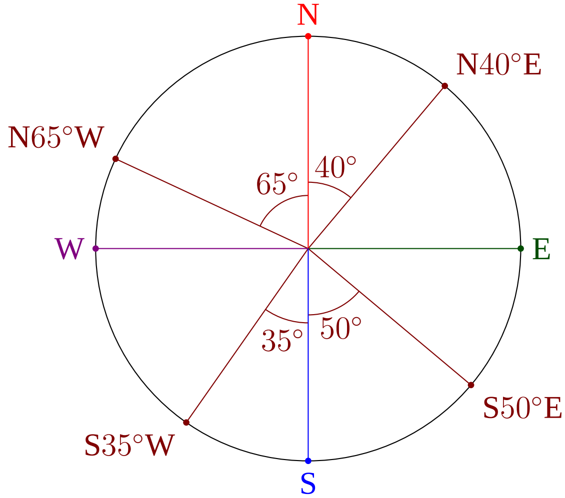


Figure 2.1.6

For example, when we specify N$65°$W, we mean the direction that starts North and turns $65°$ towards West. What we basically doing is simply *using the compass as a protractor*.

▶ Now work Exercises 7-9.

Let us go back now to our story about driving in the desert with the aid of a GPS and a compass. If our GPS device is "smart enough" to tell us not only where we are on the map, but also how to get from our current location to another point, then most likely the *driving direction* will be provided, for example, in the following form: *drive 2 miles in the* N$65°$W *direction*. (Remember, there are no streets/roads in the desert, so you only drive in straight lines between points.) This type of driving directions are what we call **vectors**. The precise mathematical definitions are as follows.

A **vector** in the plane is an *oriented line segment* $\vec{v} = \overrightarrow{AB}$. The endpoints of the vector are named as follows: the (first) point A is called the *source* of the vector; the (second) point is called the *target* of the vector. The length AB is called the *magnitude* of the vector, and is denoted by $\|\vec{v}\|$. A **zero vector** is one that has *zero magnitude*.

CLARIFICATIONS AND ADDITIONAL TERMINOLOGY. For any *non-zero* vector $\vec{v} = \overrightarrow{AB}$, the ray that *emanates from the source A and passes through the target B* is referred to as the **ray supported by $\vec{v}$** (denoted by r in the figure below).

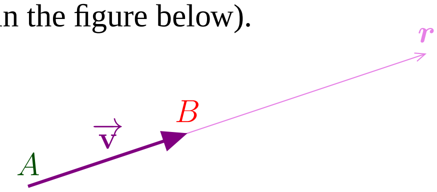


Figure 2.1.7

The **direction of $\vec{v}$** is then the direction of this ray (defined, if we wish, using a compass). Using this terminology we can always decide, for instance, if two vectors point in the *same direction*, or in *opposite directions*, or if their directions form a certain angle, and so on, and so on. We will agree that *zero vectors point in any direction.*

Two vectors are said to be **equivalent**, if *they have the same direction and equal magnitudes*. We agree that *all zero vectors are equivalent*.

CLARIFICATIONS. The figure below depicts four vectors, all pointing in the same direction (they sit on several parallel lines, shown dotted).

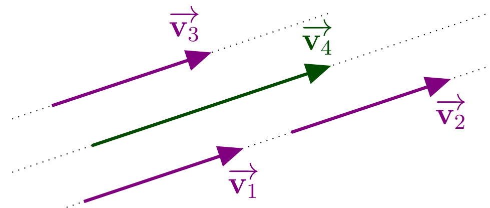


Figure 2.1.8

Among these four vectors, $\overrightarrow{\mathbf{v}}_1$, $\overrightarrow{\mathbf{v}}_2$ and $\overrightarrow{\mathbf{v}}_3$ are equivalent, but $\overrightarrow{\mathbf{v}}_4$ is not equivalent to them: although is has same direction, is is longer (it has greater magnitude).

When interpreting of vectors as "driving directions," we will always think *equivalent* vectors as *identical* objects. This point of view is incorporated in the following list of statements, which summarizes the main features of vectors. (In Rule I we recognize the definition of *vector equivalence*, which is now substituted with *vector coincidence*.)

The Rules of the "Vector Game"

I. *Two vectors coincide, if they have the same direction and equal magnitudes*

II. *"Driving" from a point P along a vector $\overrightarrow{\mathbf{v}}$ is the same as placing a "copy" of $\overrightarrow{\mathbf{v}}$ with P as its source. The target of this "copy" will be the "destination" of the drive.*

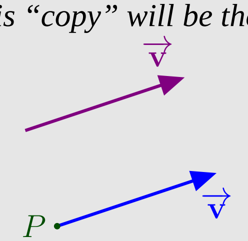


Figure 2.1.9

III. *Given a point P, the relative position of any point Q with respect to P is completely characterized by the vector $\overrightarrow{\mathbf{v}} = \overrightarrow{PQ}$, which we refer to as the position vector of Q relative to P.*

IV. *If we fix a point O in the plane (which we may call the origin, if we like), then by taking position vectors, relative to O, we establish a 1-1 correspondence between the set of all points in the plane and the set of all vectors.* The position vector of the origin O relative to itself is the zero vector, which from now on we denote by $\overrightarrow{\mathbf{0}}$.

Vector Arithmetic

Vectors can be stretched and reversed by devising the following operation.

Scalar Multiplication

Assume $\vec{v}$ is a *vector*, and t is some *real number*. We define the *t-multiple of* $\vec{v}$ to be the unique vector $\vec{w}$, characterized as follows:

(A) The *magnitude* of $\vec{w}$ is:

$$\|\vec{w}\| = |t| \cdot \|\vec{v}\|.$$

(In particular, if $t = 0$, then $\vec{w}$ is the zero vector $\vec{0}$.)

(B) The *direction* of $\vec{w}$ is, defined according to the *sign of* t:

- if $t > 0$, then $\vec{w}$ has *same direction as* $\vec{v}$;
- if $t < 0$, then $\vec{w}$ and $\vec{v}$ have *opposite directions*;
- if $t = 0$, then $\vec{w}$ has *any direction* (because it is the zero vector $\vec{0}$).

The vector $\vec{w}$ will be denoted simply by $t\vec{v}$. As a special case, when we take $t = -1$, the vector $(-1)\vec{v}$ will be denoted simply by $-\vec{v}$, and will be called the *vector opposite of* $\vec{v}$.

If we look for example at the vectors depicted in Figure 2.1.8, we see that $\vec{v_1} = \vec{v_2} = \vec{v_3} = t\vec{v_4}$, with $t > 0$. (In fact, since $\vec{v_1}$ is slightly shorter than $\vec{v_4}$, we can in fact say that we also have the inequality $t < 1$.)

Scalar multiplication is the only operation that keeps vectors "in line," in the following sense.

Given some non-zero vector $\vec{v}$, the vectors that point in either the same, or opposite direction as $\vec{v}$ are exactly the scalar multiples of $\vec{v}$, that is, of the form $\vec{w} = t\vec{v}$.

Vectors can also be *added*, as explained in the following definition.

Vector Addition

Assume two vectors $\vec{v_1}$ and $\vec{v_2}$ are given. We define the *vector sum of* $\vec{v_1}$ *with* $\vec{v_2}$ to be the vector $\vec{w}$, constructed as follows:

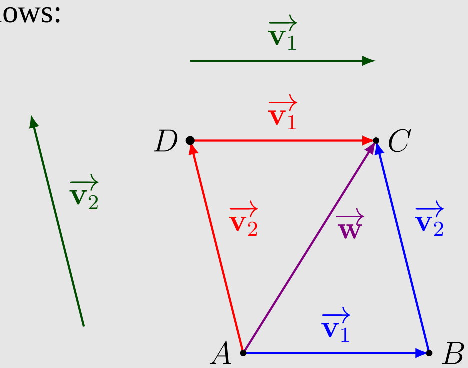


Figure 2.1.10

(i) Start off with a "copy" of $\vec{v_1}$, placed somewhere in the plane, so that is has source at some point A, and target B.

(ii) Place a "copy" of $\vec{v_2}$ to start at B, and let C be the target of this "copy."

(iii) Set $\vec{w} = \overrightarrow{AC}$.

The result $\vec{w}$ of this construction is denoted by $\vec{v_1} + \vec{v_2}$.

CLARIFICATIONS. When we think vectors as "driving directions," the construction of the sum $\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2$ can be understood as follows: start somewhere, drive according to what $\vec{\mathbf{v}}_1$ specifies, then drive according to what $\vec{\mathbf{v}}_2$ specifies. In other words, we travel along a path that can be described as "$\vec{\mathbf{v}}_1$ *followed by* $\vec{\mathbf{v}}_2$." As Figure 2.1.10 suggests (where the two copies mentioned in the definition are $\vec{\mathbf{v}}_1 = \overrightarrow{AB}$ and $\vec{\mathbf{v}}_2 = \overrightarrow{BC}$), we can also use the route "$\vec{\mathbf{v}}_2$ followed by $\vec{\mathbf{v}}_1$" (using two other copies $\vec{\mathbf{v}}_1 = \overrightarrow{AD}$ and $\vec{\mathbf{v}}_1 = \overrightarrow{DC}$). This explains why we always have the equality:

$$\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_1.$$

Using Figure 2.1.10 as a guide, some folks describe the construction of the vector sum as either being given by the so-called *Triangle Rule* (because either way we complete a triangle), or being given by the so-called *Parallelogram Rule* (because we can think of our construction as a way to complete a parallelogram: $ABCD$).

What is not so obvious is the fact that the result of the above construction is *the same, regardless of the point (A), where we decide to start*. This will be clarified shortly in the next topic, where we will make the connection between Vector Geometry and *matrices*.

Vector Coordinates

Throughout this entire topic, we assume we have fixed a length unit, and all coordinate systems we use are square.

Vector Coordinates

Given a vector $\vec{\mathbf{v}}$, the differences between matching coordinates of the endpoints of the vector are named as follows:

- the difference $\big($*x-coordinate of target of* $\vec{\mathbf{v}}\big) - \big($*$x$-coordinate of source of* $\vec{\mathbf{v}}\big)$ is called the *x-coordinate* $\vec{\mathbf{v}}$, or the "**run**" of $\vec{\mathbf{v}}$.
- the difference $\big($*y-coordinate of target of* $\vec{\mathbf{v}}\big) - \big($*$y$-coordinate of source of* $\vec{\mathbf{v}}\big)$ is called the *y-coordinate* $\vec{\mathbf{v}}$, or the "**rise**" of $\vec{\mathbf{v}}$.

Using the Distance Formula in Square Coordinates, one easily obtains:

The Vector Magnitude Formula in Square Coordinates

If the coordinates of a vector $\vec{\mathbf{v}}$ are x and y, then its magnitude is

$$\|\vec{\mathbf{v}}\| = \sqrt{x^2 + y^2} \text{ units.} \tag{2.1.11}$$

This definition of vector coordinates will be improved a little later, when we will discuss the Position Vector Formula, and its coordinate version.

Example 2.1.2. Suppose a vector is presented as $\vec{\mathbf{v}} = \overrightarrow{AB}$, with start point $A(2, -1)$ and target point $B(-1, 1)$, at let us find the coordinates and the magnitude of $\vec{\mathbf{v}}$.

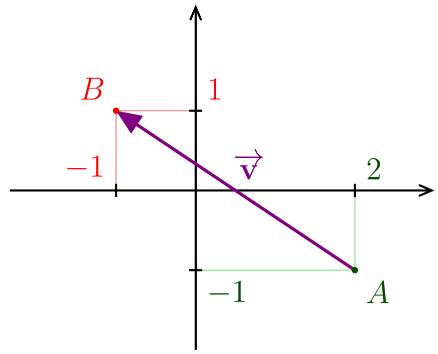


Figure 2.1.11

The x-coordinate of $\vec{\mathbf{v}}$ is: $x = \big(x\text{-coordinate of } B\big) - \big(x\text{-coordinate of } A\big) = (-1) - 2 = -3$. The x-coordinate of $\vec{\mathbf{v}}$ is: $y = \big(y\text{-coordinate of } B\big) - \big(y\text{-coordinate of } A\big) = 1 - (-1) = 2$. Having the coordinates of our vector in hand, its magnitude is:

$$\|\vec{\mathbf{v}}\| = \sqrt{(-3)^2 + 2^2} = \sqrt{13} \text{ units.}$$

CLARIFICATION. An alternative way to computing the coordinates of a vector $\vec{\mathbf{v}}$ sitting somewhere in the plane, is to consider a *copy of the vector, which starts at the origin*. In other words, we seek to present $\vec{\mathbf{v}}$ as the *position vector of a point P with respect to the origin*. Once the point P is found, *the coordinates of $\vec{\mathbf{v}}$ coincide with the coordinates of P*. If we use this approach, then Figure 2.1.11 can be enhanced to look like:

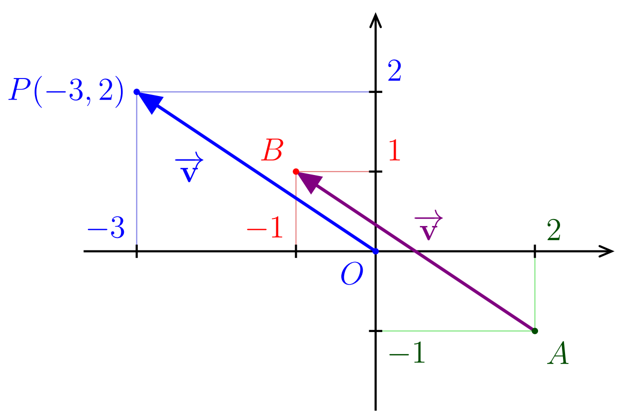


Figure 2.1.12

NOTATION CONVENTION. If the coordinates of a vector $\vec{\mathbf{v}}$ are x and y, then we represent $\vec{\mathbf{v}}$ as a **2×1 matrix:**

$$\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.1.12}$$

Identifying *vectors* (in coordinates) with 2×1 *matrices* is justified by the following fundamental statement.

Vector-Matrix Arithmetic Equivalence Theorem

Vector Arithmetic matches Matrix Arithmetic. More precisely:

I. *If* $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$, *and* t *is some real number, then:*

$$t\vec{\mathbf{v}} = t\begin{bmatrix} x \\ y \end{bmatrix} \underset{\text{(I)}}{=} \begin{bmatrix} tx \\ ty \end{bmatrix}. \tag{2.1.13}$$

II. *If* $\vec{\mathbf{v}}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ *and* $\vec{\mathbf{v}}_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, *then:*

$$\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \underset{\text{(II)}}{=} \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}. \tag{2.1.14}$$

CLARIFICATIONS. The equalities (I) and (II) are exactly the formulas that we know from Algebra, when we learned operations with matrices (see Appendix B.) Since the operations with matrices have very nice properties, we can use them to get their vector counterparts, which are as follows.

Properties of Vector Arithmetic

I. *Associativity of Addition:* $(\vec{\mathbf{u}} + \vec{\mathbf{v}}) + \vec{\mathbf{w}} = \vec{\mathbf{u}} + (\vec{\mathbf{v}} + \vec{\mathbf{w}})$.
II. *Commutativity of Addition:* $\vec{\mathbf{u}} + \vec{\mathbf{v}} = \vec{\mathbf{v}} + \vec{\mathbf{u}}$.
III. *Opposite Property:* $(-\vec{\mathbf{u}}) + \vec{\mathbf{u}} = \vec{\mathbf{u}} + (-\vec{\mathbf{u}}) = \vec{\mathbf{0}}$.
IV. *Easy Scalar Multiplications:* $0\vec{\mathbf{u}} = \vec{\mathbf{0}}$; $1\vec{\mathbf{u}} = \vec{\mathbf{u}}$; $t\vec{\mathbf{0}} = \vec{\mathbf{0}}$.
V. *Associativity of Scalar Multiplication:* $s(t\vec{\mathbf{u}}) = t(s\vec{\mathbf{u}}) = (st)\vec{\mathbf{u}}$.
VI. *Distributivity over Scalar Addition:* $(s+t)\vec{\mathbf{u}} = s\vec{\mathbf{u}} + t\vec{\mathbf{u}}$.
VII. *Distributivity over Vector Addition:* $t(\vec{\mathbf{u}} + \vec{\mathbf{v}}) = t\vec{\mathbf{u}} + t\vec{\mathbf{v}}$.

CLARIFICATIONS AND ADDITIONAL TERMINOLOGY. When dealing with three or more vectors which are added, by the Associativity of Vector Addition (property I above), it is not necessary to use parentheses anymore, so we can simply write long sums like: $\vec{\mathbf{u}} + \vec{\mathbf{v}} + \vec{\mathbf{w}}$.

Concerning the use of Vector Opposites, we can also omit parentheses, by defining the **vector subtraction** operation:

$$\vec{\mathbf{u}} - \vec{\mathbf{v}} = \vec{\mathbf{u}} + (-\vec{\mathbf{v}}).$$

Using 2×1 matrix presentations of vectors, *vector subtraction corresponds, of course, to matrix subtraction*. Using vector subtraction, we can easily compute *position vectors*, as follows.

Position Vector Formula

Assume some point O is fixed in the plane. Suppose P_1 and P_2 are two points, and let $\vec{\mathbf{p_1}} = \overrightarrow{OP_1}$ and $\vec{\mathbf{p_2}} = \overrightarrow{OP_2}$ be their *position vectors relative to the* O. Then the vector $\vec{\mathbf{v}} = \overrightarrow{P_1P_2}$, which represents the *position vector of* P_2 *relative to the* P_1, is given by:

$$\vec{\mathbf{v}} = \overrightarrow{P_1P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1} = \vec{\mathbf{p_2}} - \vec{\mathbf{p_1}}. \tag{2.1.15}$$

CLARIFICATIONS. If we use a square coordinate system, with O as the ***origin***, and we have our two points written in coordinates as $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$, then their position vectors can be easily presented in coordinates as

$$\vec{\mathbf{p_1}} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \quad \text{and} \quad \vec{\mathbf{p_2}} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

and our vector $\vec{\mathbf{v}} = \overrightarrow{P_1P_2}$ will be given in coordinates as:

$$\vec{\mathbf{v}} = \overrightarrow{P_1P_2} = \vec{\mathbf{p_2}} - \vec{\mathbf{p_1}} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{bmatrix}$$

Example 2.1.3. Suppose we have two cars driving in the desert, both starting at some point O. Suppose one car reached point P_1 by driving 3 miles in the N$65°$W direction, and the other car reached point P_2 by driving 2 miles in the S$35°$W direction, with both cars starting from O.

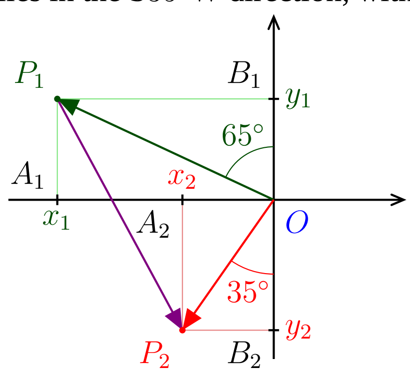


Figure 2.1.13

We are asked to compute the following vectors in coordinates: (i) the vectors $\vec{\mathbf{p_1}} = \overrightarrow{OP_1}$ and $\vec{\mathbf{p_2}} = \overrightarrow{OP_2}$, which are position vectors relative to O; (ii) the vector $\vec{\mathbf{v}} = \overrightarrow{P_1P_2}$, which is the position vector of P_2 relative to P_1. Using these calculations, we are also asked to find the distance between the two cars.

Solution. To find the coordinates of $\vec{\mathbf{p_1}}$ we take a look at the right triangle $\triangle P_1OB_1$, where B_1 is the projection of P_1 on the y-axis (the North-South axis). Setting up the trigonometric functions of the $65°$ angle yields

$$\sin 65° = \frac{B_1P_1}{OP_1} = \frac{B_1P_1}{3} \quad \text{and} \quad \cos 65° = \frac{OB_1}{OP_1} = \frac{OB_1}{3},$$

from which we can compute

$$B_1P_1 = 3\sin 65° \quad \text{and} \quad OB_1 = 3\cos 65°. \tag{2.1.16}$$

If we use the ruler on the y-axis, where B_1 has coordinate y_1, then we know that $OB_1 = |y_1|$, so using (2.1.16), we get

$$|y_1| = 3\cos 65°.$$

Since $\overrightarrow{\mathbf{p_1}}$. points "upwards" (towards North), the number y_1 is *positive*, so the above equality yields:

$$y_1 = 3\cos 65^\circ. \tag{2.1.17}$$

To get the coordinate x_1, we notice first that, when we consider A_1 to be the projection of of P_1 on the x-axis (the East-West axis), we clearly have a triangle congruence $\triangle P_1OB_1 \equiv \triangle P_1OA_1$, which then yields $OA_1 = B_1P_1 = 3\sin 65^\circ$. Secondly, working exactly as above, this will give us

$$|x_1| = 3\sin 65^\circ.$$

Lastly, since $\overrightarrow{\mathbf{p_1}}$. points "to the left" (towards West), the number x_1 is *negative*, so the above equality yields:

$$x_1 = -3\sin 65^\circ. \tag{2.1.18}$$

The coordinates of $\overrightarrow{\mathbf{p_2}}$ are found exactly the same way:

$$x_2 = -2\sin 35^\circ \quad \text{and} \quad y_2 = -2\cos 35^\circ, \tag{2.1.19}$$

so our two vectors are presented in coordinates as

$$\overrightarrow{\mathbf{p_1}} = \begin{bmatrix} -3\sin 65^\circ \\ 3\cos 65^\circ \end{bmatrix} \simeq \begin{bmatrix} -2.719 \\ 1.268 \end{bmatrix} \quad \text{and} \quad \overrightarrow{\mathbf{p_2}} = \begin{bmatrix} -2\sin 35^\circ \\ -2\cos 35^\circ \end{bmatrix} \simeq \begin{bmatrix} -1.147 \\ -1.638 \end{bmatrix}$$

The vector $\overrightarrow{\mathbf{v}} = \overrightarrow{P_1P_2}$ is now computed using the Position Vector Formula

$$\overrightarrow{\mathbf{v}} = \overrightarrow{\mathbf{p_2}} - \overrightarrow{\mathbf{p_1}} = \begin{bmatrix} -2\sin 35^\circ \\ -2\cos 35^\circ \end{bmatrix} - \begin{bmatrix} -3\sin 65^\circ \\ 3\cos 65^\circ \end{bmatrix} = \begin{bmatrix} -2\sin 35^\circ + 3\sin 65^\circ \\ -2\cos 35^\circ - 3\cos 65^\circ \end{bmatrix} \simeq \begin{bmatrix} 1.572 \\ -2.905 \end{bmatrix}$$

The distance between the two cars is the magnitude of this vector, that is,

$$\|\overrightarrow{\mathbf{v}}\| = \sqrt{(-2\sin 35^\circ + 3\sin 65^\circ)^2 + (-2\cos 35^\circ - 3\cos 65^\circ)^2} \simeq 3.304 \text{ miles}.$$

▶ Now work Exercises 10-13.

Using coordinates, one can easily decide *when two vectors are perpendicular*, as explained in the following statement. (The outline of the proof is given in Exercise 30.)

Perpendicular Vectors Theorem

Two vectors given in coordinates as $\overrightarrow{\mathbf{v_1}} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ *and* $\overrightarrow{\mathbf{v_2}} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, *are perpendicular, if and only if:*

$$x_1x_2 + y_1y_2 = 0. \tag{2.1.20}$$

Vectors and Lines

Throughout this entire topic, we assume we have fixed a length unit, together with a square coordinate system. When dealing with points in the plane, we now have two points of view on them, which we will often interchange:

Coordinates	Vector
$P(x,y) \longleftrightarrow$	$\overrightarrow{\mathbf{p}} = \begin{bmatrix} x \\ y \end{bmatrix}$

Unless otherwise specified, the vector $\overrightarrow{\mathbf{p}}$ that *corresponds to* P will always mean the *position vector of* P *relative to the origin*.

Using vectors, we can quite easily "navigate" along lines. To be a bit more specific, suppose we have a line $\mathscr{L}$, which passes through two distinct points $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$. We wish to characterize all points $P(x, y)$ that sit on the line $\mathscr{L}$, either indirectly, or by writing down an equation in x and y. Consider the position vectors $\overrightarrow{\mathbf{p}_0} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$, $\overrightarrow{\mathbf{p}_1} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, $\overrightarrow{\mathbf{p}} = \begin{bmatrix} x \\ y \end{bmatrix}$ of our three points, and the vectors

$$\overrightarrow{\mathbf{u}} = \overrightarrow{P_0P_1} = \overrightarrow{\mathbf{p}_1} - \overrightarrow{\mathbf{p}_0} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_1 - x_0 \\ y_1 - y_0 \end{bmatrix};$$
$$\overrightarrow{\mathbf{v}} = \overrightarrow{P_0P} = \overrightarrow{\mathbf{p}} - \overrightarrow{\mathbf{p}_0} = \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}.$$

By placing both these vectors to start at P_0, we see that the condition that *P sits on $\mathscr{L}$* is equivalent to the condition that *the vectors $\overrightarrow{P_0P_1}$ and $\overrightarrow{P_0P}$ point in either the same direction, or in opposite directions*. This simply means that *we can write $\overrightarrow{\mathbf{v}}$ as a scalar multiple of* $\overrightarrow{\mathbf{u}}$, meaning that $\overrightarrow{\mathbf{v}} = t\overrightarrow{\mathbf{u}}$, for some number t. By replacing these two vectors with their original definitions, the above condition reads:

$$\overrightarrow{\mathbf{p}} - \overrightarrow{\mathbf{p}_0} = t\big(\overrightarrow{\mathbf{p}_1} - \overrightarrow{\mathbf{p}_0}\big). \tag{2.1.21}$$

By adding $\overrightarrow{\mathbf{p}_0}$ to both sides, the above equality can be re-written as $\overrightarrow{\mathbf{p}} = \overrightarrow{\mathbf{p}_0} + t\big(\overrightarrow{\mathbf{p}_1} - \overrightarrow{\mathbf{p}_0}\big) = \overrightarrow{\mathbf{p}_0} + t\overrightarrow{\mathbf{p}_1} - t\overrightarrow{\mathbf{p}_0}$, thus by grouping, we can also re-write (2.1.21) as:

$$\overrightarrow{\mathbf{p}} = t\overrightarrow{\mathbf{p}_1} + (1-t)\overrightarrow{\mathbf{p}_0}. \tag{2.1.22}$$

This way we have proved the following important statement.

Parametric Characterization of Lines

Given a line $\mathscr{L}$ passing through two distinct points $P_0(x_0, y_0)$ and $P_0(x_1, y_1)$, a point $P(x, y)$ sits on $\mathscr{L}$, if and only if there is some real number t, such that

$$\begin{bmatrix} x \\ y \end{bmatrix} = (1-t)\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{2.1.23}$$

Explicitly, the matrix equality (2.1.23) reads:

$$\begin{cases} x = tx_1 + (1-t)x_0 \\ y = ty_1 + (1-t)y_0 \end{cases} \tag{2.1.24}$$

The above parametrization is particularly nice, because it can also tell us something about the location of P on the line, as it relates to P_0 and P_1. The easiest way to understand this, is to think t as a *time variable*, so as t increases from $-\infty$ to ∞, our vehicle drives at constant speed on the line $\mathscr{L}$, passing first through P_0 at time $t = 0$, then passing through P_1 at time $t = 1$. In particular, at $t = \dfrac{1}{2}$, we are at the *midpoint of* $\overline{P_0P_1}$, which will have coordinates

$$\begin{cases} x = \dfrac{1}{2}(x_1 + x_0) \\[2ex] y = \dfrac{1}{2}(y_1 + y_0) \end{cases}$$

In vector notation, the position vector of this midpoint is simply

$$\overrightarrow{\mathbf{m}} = \frac{1}{2}\left(\overrightarrow{\mathbf{p_1}} + \overrightarrow{\mathbf{p_0}}\right).$$

Switching things around a little bit, let us assume for the moment that only one point $P_0(x_0, y_0)$ on $\mathscr{L}$ is given, and instead of $P_1(x_1, y_1)$ we know a *non-zero vector* $\overrightarrow{\mathbf{u}} = \begin{bmatrix} h \\ k \end{bmatrix}$*, which either sits on, or is parallel to* $\mathscr{L}$. Such a vector is called a **direction vector for** $\mathscr{L}$. Then our line can be parametrized either in position vector form

$$\overrightarrow{\mathbf{p}} = \overrightarrow{\mathbf{p_0}} + t\,\overrightarrow{\mathbf{u}}, \tag{2.1.25}$$

or in coordinates:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t \begin{bmatrix} h \\ k \end{bmatrix}; \tag{2.1.26}$$

$$\begin{cases} x = x_0 + th \\[2ex] y = y_0 + tk \end{cases} \tag{2.1.27}$$

(All these formulas are easily obtained from their two-point versions, because a second point P_1 on $\mathscr{L}$ can be easily constructed by means of the position vector $\overrightarrow{\mathbf{p_1}} = \overrightarrow{\mathbf{p_0}} + \overrightarrow{\mathbf{u}}$.)

The phrase *"direction vector"* (for a line) might be a bit deceiving. If for example we start with a direction vector $\overrightarrow{\mathbf{u}}$ for a line $\mathscr{L}$, then its opposite vector $-\overrightarrow{\mathbf{u}}$ is again a direction vector for the same line! Even though a line can have a lot of direction vectors, *any two of them are non-zero scalar multiples of each other*.

Example 2.1.4. Consider the line $\mathscr{L}$ that passes through the points $A(1, -2)$ and $B(3, 8)$, and suppose we are asked to produce three distinct direction vectors for it. One of them is, of course, the vector

$$\overrightarrow{\mathbf{u_1}} = \overrightarrow{AB} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix} = \begin{bmatrix} 3-1 \\ 8-(-2) \end{bmatrix} = \begin{bmatrix} 2 \\ 10 \end{bmatrix}.$$

Since all other direction vectors for $\mathscr{L}$ are non-zero scalar multiples of $\overrightarrow{\mathbf{u_1}}$, we can construct a second direction vector, for instance, by taking

$$\overrightarrow{\mathbf{u_2}} = \tfrac{1}{2}\overrightarrow{\mathbf{u_1}} = \tfrac{1}{2}\begin{bmatrix} 2 \\ 10 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

Another direction vector for $\mathscr{L}$ (this time pointing in the direction opposite to $\vec{\mathbf{u}}_1$) can be taken, for instance, to be:

$$\vec{\mathbf{u}}_3 = -\vec{\mathbf{u}}_2 = \begin{bmatrix} -1 \\ -5 \end{bmatrix}.$$

The main drawback of parametric equations of lines (and many other curves, for that matter) is the fact that they make it difficult to check if a "candidate" point $P(x, y)$ does or does not sit on our line (or curve). For this reason, we seek other types of equation, in which the parameter (t, in our case) is *eliminated*. Fortunately, for lines, this elimination process is easily done: in (2.1.27), which we can also write as $x - x_0 = th$ & $y - y_0 = tk$, we can multiply the first equation by k, the second equation by h, making the right-hand sides in both equations equal, thus obtaining the following.

The Line Equations

I. **Point-Direction Equation.** *A line $\mathscr{L}$, which passes through a point $P_0(x_0, y_0)$ and has $\vec{\mathbf{u}} = \begin{bmatrix} h \\ k \end{bmatrix}$ as a direction vector, can be represented by the equation:*

$$k(x - x_0) = h(y - y_0). \tag{2.1.28}$$

II. **Two-Point Equation.** *A line $\mathscr{L}$, which passes through two distinct points $P_0(x_0, y_0)$ and $P(x_1, y_1)$, can be represented by the equation:*

$$(y_1 - y_0)(x - x_0) = (x_1 - x_0)(y - y_0). \tag{2.1.29}$$

The equation (2.1.29) follows from (2.1.28), by using the vector $\vec{\mathbf{u}} = \overrightarrow{P_0P_1} = \vec{\mathbf{p}}_1 - \vec{\mathbf{p}}_0$.

CLARIFICATIONS. Just about anywhere we have seen line equations before, they were presented in the form of a so-called *general linear equation*

$$ax + by = c. \tag{2.1.30}$$

Of course, either one of the equations (2.1.28) or (2.1.29) can be transformed to match (2.1.30). For example, if we start with $k(x - x_0) = h(y - y_0)$, we can "open up" parentheses, to get

$$kx - kx_0 = hy - hy_0,$$

and adding $kx_0 - hy$ to both sides will yield

$$kx - hy = kx_0 - hy_0, \tag{2.1.31}$$

which can be matched with (2.1.30) as: $a = k$, $b = -h$, $c = kx_0 - hy_0$.

Of course, there is nothing unique about all our coefficients, for instance the equation $2x+3y = 1$ is equivalent to $4x+6y = 2$, so they represent the same line. Concerning general linear equations, the only "safe" statement one can make is:

If two general linear equations $ax + by = c$ and $a'x + b'y = c'$ represent the same line, then the triples (a, b, c) and (a', b', c') are proportional, meaning that there exists some real number t, such that $a' = ta$, $b' = tb$ and $c' = tc$.

The general equation of a line (although not unique) provides us with some useful geometric information, as illustrated by the following statement.

The Frame Theorem

Assume a line $\mathscr{L}$ is presented by a general linear equation

$$ax + by = c. \tag{2.1.32}$$

(A) *The vector* $\vec{\mathbf{u}} = \begin{bmatrix} -b \\ a \end{bmatrix}$ *is a direction vector for $\mathscr{L}$. Furthermore, all other direction vectors for $\mathscr{L}$ are non-zero scalar multiples of $\vec{\mathbf{u}}$.*

(B) *The vector* $\vec{\mathbf{n}} = \begin{bmatrix} a \\ b \end{bmatrix}$ *is perpendicular to $\mathscr{L}$. Furthermore, all other vectors perpendicular to $\mathscr{L}$ are non-zero scalar multiples of $\vec{\mathbf{n}}$.*

Proof. Fix one particular point $P_0(x_0, y_0)$ on the line $\mathscr{L}$, and consider its position vector $\vec{\mathbf{p}}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$. Consider now the point $P_1(x_1, y_1)$, whose position vector is $\vec{\mathbf{p}}_1 = \vec{\mathbf{p}}_0 + \vec{\mathbf{u}}$, thus

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} x_0 - b \\ y_0 + a \end{bmatrix}.$$

Using the fact that x_0 and y_0 satisfy the equation (2.1.32), it follows that

$$ax_1 + by_1 = a(x_0 - b) + b(y_0 + a) = ax_0 - ab + by_0 + ab = ax_0 + by_0 = c,$$

which means that the point P_1 also sits on $\mathscr{L}$, so in particular, the vector $\overrightarrow{P_0P_1} = \vec{\mathbf{p}}_1 - \vec{\mathbf{p}}_0 = \vec{\mathbf{u}}$ is indeed a direction vector for $\mathscr{L}$, thus proving statement (A). As for statement (B), we simply observe that the vectors $\vec{\mathbf{u}} = \begin{bmatrix} -b \\ a \end{bmatrix}$ and $\vec{\mathbf{n}} = \begin{bmatrix} a \\ b \end{bmatrix}$ easily satisfy the condition from the Perpendicular Vectors Theorem, so $\vec{\mathbf{n}}$ is indeed perpendicular to $\vec{\mathbf{u}}$. □

Example 2.1.5. Suppose we have a line $\mathscr{L}$ given by the equation

$$4x - 6y = 5,$$

and we want to find some direction vectors and some perpendicular vectors.

According to the Frame Theorem, one particular direction vector for $\mathscr{L}$ is $\vec{\mathbf{u}} = \begin{bmatrix} -(-6) \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, and any other direction vector $\vec{\mathbf{v}}$ for $\mathscr{L}$ must be a non-zero scalar multiple of $\vec{\mathbf{u}}$, so in our case we are looking at vectors of the form

$$\vec{\mathbf{v}} = s\vec{\mathbf{u}} = s\begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 6s \\ 4s \end{bmatrix},$$

with s any number we like, except 0. For instance, when we let $s = \frac{1}{2}$, we get a new direction vector: $\vec{\mathbf{v}}_1 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$. We can also let $s = -\frac{1}{2}$, in which case we obtain yet another direction vector: $\vec{\mathbf{v}}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

Again by the Frame Theorem, one particular vector that is perpendicular to $\mathscr{L}$ is $\overrightarrow{\mathbf{n}} = \begin{bmatrix} 4 \\ -6 \end{bmatrix}$, and any other vector $\overrightarrow{\mathbf{w}}$ that is perpendicular to $\mathscr{L}$ must be a non-zero scalar multiple of $\overrightarrow{\mathbf{n}}$, so in our case we are looking at vectors of the form

$$\overrightarrow{\mathbf{w}} = s\overrightarrow{\mathbf{n}} = s\begin{bmatrix} 4 \\ -6 \end{bmatrix} = \begin{bmatrix} 4s \\ -6s \end{bmatrix},$$

with s any number we like, except 0. For instance, when we let $s = \frac{1}{2}$, we get a new perpendicular vector: $\overrightarrow{\mathbf{w}_1} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$. We can also let $s = -\frac{1}{2}$, in which case we obtain yet another perpendicular vector: $\overrightarrow{\mathbf{w}_2} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

A quick application of the Frame Theorem, we get the following characterization of parallel and perpendicular lines.

Equations of Parallel/Perpendicular Lines

Given a line $\mathscr{L}$ with equation

$$ax + by = c,$$

the lines, that are *parallel* or *perpendicular* to $\mathscr{L}$, are characterized as follows.

(A) *Every line which is parallel to $\mathscr{L}$ is represented by an equation of the form:*

$$ax + by = c',$$

for some constant $c' \neq c$.

(A) *Every line which is perpendicular to $\mathscr{L}$ is represented by an equation of the form:*

$$-bx + ay = c'',$$

for some constant c''.

Example 2.1.6. Suppose we have the point $P(2, -1)$ and the line $\mathscr{L}$ is given by the equation

$$3x + 5y = 8,$$

and we want to find the (general) equations of: (a) the line $\mathscr{L}_1$ passing through P which is parallel to $\mathscr{L}$, and (b) the line $\mathscr{L}_2$ passing through P which is perpendicular to $\mathscr{L}$.

According to the above characterizations, the line $\mathscr{L}_1$ can be represented by an equation of the form

$$3x + 5y = c_1$$

with c_1 to be determined. By plugging in the coordinates of the given point, which sits on $\mathscr{L}_1$, we get $c_1 = 3(2) + 5(-1) = 1$, so $\mathscr{L}_1$ can be given by the equation

$$3x + 5y = 1. \tag{2.1.33}$$

Likewise, the line $\mathscr{L}_2$ can be represented by an equation of the form

$$-5x + 3y = c_2$$

with c_2 to be determined. By plugging in the coordinates of the given point, which sits on $\mathscr{L}_2$, we get $c_2 = -5(2) + 3(-1) = -13$, so $\mathscr{L}_2$ can be given by the equation

$$-5x + 3y = -13. \tag{2.1.34}$$

CLARIFICATION. Most readers have seen lines described in a slightly different way, by means of the so-called **slope-intercept equations**, which are presented in the form

$$y = mx + p. \tag{2.1.35}$$

In such a presentation, m is the *slope*, and p is the *y-intercept*, that is, the *y-coordinate of the point where the line intersects the y-axis* (whose x-coordinate is, of course, $x = 0$). Although equations of the form (2.1.35) have some advantages, the general equations of lines are much nicer to use, for several reasons:

- The lines that can be presented by (2.1.35) cannot be vertical, whereas their general equation is simply $x = c$.
- The general equation allows us to easily read the geometric information about our line, such as direction or perpendicular vectors, as described in the Frame Theorem.
- Finding the equations of perpendicular lines using (2.1.35) can be a little tricky, whereas using general equations (as in Example 2.1.6) the task is much simplified.
- In most instances, when we try to find the slope-intercept equation of a line, the slope m ends up being a fraction, whereas in general equations fractional coefficients may be avoided.
- Although (2.1.35) allows you to quickly find the y-intercept, the x-intercept requires an algebraic manipulation. If we use general equations in the form $ax + by = c$, these intercepts are straightforward: the x-intercept is $\frac{c}{a}$; the y-intercept is $\frac{c}{b}$.

The only advantages of using slope-intercept equations of lines (if available) are *uniqueness* and the ease of plotting. If a line is presented by an equation of the form (2.1.35), points on it can be easily generated by plugging in various values for x, and using the equation to (easily) compute the matching y-values.

Example 2.1.7. Consider the equations (2.1.33) and (2.1.34) for the lines $\mathscr{L}_1$ and $\mathscr{L}_2$ which we discussed in Example 2.1.7, and let us convert these equations to the slope-intercept form. The main technique is to take each each equation, and solve it for y.

When we consider the equation (2.1.33), we can subtract $3x$ from both sides of the equation, thus getting $5y = -3x + 1$, and then divide by 5, or equivalently multiplying by $\frac{1}{5}$, thus getting

$$y = \frac{1}{5}\big(- 3x + 1\big) = -\frac{3}{5}x + \frac{1}{5}.$$

Similarly, when we consider the equation (2.1.34), we can add $5x$ to both sides of the equation, thus getting $3y = 5x - 13$, and then divide by 3, or equivalently multiplying by $\frac{1}{3}$, thus getting

$$y = \frac{1}{3}\big(5x - 13\big) = \frac{5}{3}x - \frac{13}{3}.$$

▶ Now work Exercises 14-22.

Plane Isometries

The remainder of this Section is quite challenging! We include this topic here only for the sake of completeness. On a first reading, you may want only to familiarize yourselves with formula (2.1.42) – needed in Section 2.2, skip everything else, and go directly to the Exercises (omitting 25 through 32.)

What we will call a *transformation of the plane* will be simply a function $\mathbf{\Theta}$ from the euclidean plane $\mathscr{E}$ into itself, which we will write $\mathbf{\Theta} : P \longmapsto Q$, or $Q = \mathbf{\Theta}(P)$. Same notations will be used with vectors: $\mathbf{\Theta} : \vec{\mathbf{p}} \longmapsto \vec{\mathbf{q}}$, or $\vec{\mathbf{q}} = \mathbf{\Theta}(\vec{\mathbf{p}})$. In coordinates we will write $\mathbf{\Theta} : (x, y) \longmapsto (z, w)$ or $(z, w) = \mathbf{\Theta}(x, y)$, or (if we use vectors written as matrices) $\mathbf{\Theta} : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} z \\ w \end{bmatrix}$, or $\begin{bmatrix} z \\ w \end{bmatrix} = \mathbf{\Theta}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$. All transformations we are going to discuss here will be defined geometrically, and our main challenge will be to write formulas for them, so we will have to write something like:

$$(z, w) = \mathbf{\Theta}(x, y) \text{ means: } \begin{cases} z = & \text{expression in } x \text{ and } y \\ w = & \text{expression in } x \text{ and } y \end{cases}$$

In most cases of interest, we will encounter transformations that *have inverses*. If $\mathbf{\Theta}$ is such a special transformation, its inverse transformation will be denoted by $\mathbf{\Theta}^{-1}$. When computing an inverse transformations we always use the following fact.

Inversion Principle

If $\mathbf{\Theta}$ has $\mathbf{\Theta}^{-1}$ as its inverse, then

$$(z, w) = \mathbf{\Theta}^{-1}(x, y) \text{ means } (x, y) = \mathbf{\Theta}(z, w)$$

The types of transformations we are interested in are those that *preserve distances*, meaning that, for any two points P_1, P_2, we have:

$$\text{dist}(\mathbf{\Theta}(P_1), \mathbf{\Theta}(P_2)) = \text{dist}(P_1, P_2).$$

A plane transformation having this property is called an *isometry*.

The following simple fact follows straight from the above definition.

Composite Isometry Rule

A composition of isometries is again an isometry; in other words, if $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are isometries, then the composed transformation $\mathbf{\Theta}_1 \circ \mathbf{\Theta}_2$, given by

$$(\mathbf{\Theta}_1 \circ \mathbf{\Theta}_2)(P) = \mathbf{\Theta}_1(\mathbf{\Theta}_2(P)),$$

is again an isometry.

The first important class of isometries are those defined as follows.

To *translate a vector* $\vec{p}$ *by* $\vec{v}$ simply means to *add* $\vec{v}$ *to* $\vec{p}$. A transformation of the from

$$\mathfrak{T}_{\vec{v}}(\vec{p}) = \vec{p} + \vec{v}$$

where $\vec{v}$ is some fixed vector, is called a **translation by** $\vec{v}$. In coordinates, if $\vec{v} = \begin{bmatrix} h \\ k \end{bmatrix}$, then

$$\mathfrak{T}_{\vec{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} h \\ k \end{bmatrix} = \begin{bmatrix} x+h \\ y+k \end{bmatrix}.$$

The composition of two translations is again a translation:

$$\mathfrak{T}_{\vec{v}} \circ \mathfrak{T}_{\vec{w}} = \mathfrak{T}_{\vec{v}+\vec{w}} \tag{2.1.36}$$

In particular every translation has an inverse, which is again a translation:

$$\mathfrak{T}_{\vec{v}}^{-1} = \mathfrak{T}_{-\vec{v}}. \tag{2.1.37}$$

Apart from the *trivial* case (the one that has $\vec{v} = \vec{0}$), *translations do not have fixed points*. However, using the Composite Rule above, one can easily see[4] that *every isometry* $\boldsymbol{\Theta}$ *can be uniquely written as a composition* $\boldsymbol{\Theta} = \mathfrak{T}_{\vec{v}} \circ \boldsymbol{\Theta}_0$, with:

- $\mathfrak{T}_{\vec{v}}$ a *translation*, and
- $\boldsymbol{\Theta}_0$ an *isometry that fixes the origin*, that is: $\boldsymbol{\Theta}_0(O) = O$.

Our interest in isometries fixing the origin is justified by the following fundamental result (see Exercises ???-??? for an outline of the proof).

Orthogonal Matrix Theorem

An isometry $\boldsymbol{\Theta}_0$ *fixes the origin, if and only if it can presented in coordinates as multiplication by a* 2×2 *matrix*

$$\boldsymbol{\Theta}_0\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u & z \\ v & w \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ux + zy \\ vx + wy \end{bmatrix} \tag{2.1.38}$$

where the columns $\begin{bmatrix} u \\ v \end{bmatrix}$ *and* $\begin{bmatrix} z \\ w \end{bmatrix}$ *of the matrix are two perpendicular vectors of magnitudes equal to* 1.

CLARIFICATIONS. A 2×2 matrix of the special type described above is called an *orthogonal matrix*. Using the Frame Theorem, orthogonal matrices can be simply characterized by the following conditions: (i) $u^2 + v^2 = 1$; (ii) $\begin{bmatrix} z \\ w \end{bmatrix} = \pm \begin{bmatrix} v \\ -u \end{bmatrix}$. Depending on the sign in (i), the orthogonal matrices are named as follows.

(A) A matrix of the form

$$\mathbf{F} = \begin{bmatrix} u & v \\ v & -u \end{bmatrix}, \text{ with } u^2 + v^2 = 1. \tag{2.1.39}$$

is called a **flip matrix**.

(B) A matrix of the form

[4] When working with vectors in coordinates, we must have $\vec{v} = \boldsymbol{\Theta}(\vec{0})$, so $\boldsymbol{\Theta}_0$ must be given by: $\boldsymbol{\Theta}_0 = \mathfrak{T}_{\vec{v}}^{-1} \circ \boldsymbol{\Theta}_0 = \mathfrak{T}_{-\vec{v}} \circ \boldsymbol{\Theta}$.

$$\mathbf{R} = \begin{bmatrix} u & -v \\ v & u \end{bmatrix}, \text{ with } u^2 + v^2 = 1. \tag{2.1.40}$$

is called a **rotation matrix**.

Why do we use these names? As it turns out, flip matrices are intimately related to the so-called *flip transformations*, which are defined geometrically as follows:

For any given a line $\mathscr{L}$ in the plane, the **flip** (or **reflection**) **about** $\mathscr{L}$ is the transformation $\mathfrak{F}_{\mathscr{L}} : P \longmapsto Q$, defined geometrically by demanding that: *the line $\mathscr{L}$ is the perpendicular bisector of $\overline{PQ}$.*

CLARIFICATION. By construction, flip transformations have many fixed points: *all points on $\mathscr{L}$ are fixed by $\mathfrak{F}_{\mathscr{L}}$.* With this observation in mind, we have the following special version of the Orthogonal Matrix Theorem (see Exercises 23-24 for the proof):

Flip Matrix Theorem

A flip about a line passing through the origin is precisely a transformation that can presented in coordinates as multiplication by a flip matrix, that is:

$$\mathfrak{F}_{\mathscr{L}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u & v \\ v & -u \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ux + vy \\ vx - uy \end{bmatrix} \tag{2.1.41}$$

where u and v are two constants satisfying $u^2 + v^2 = 1$. More precisely, if $\mathscr{L}$ is given by the equation $ax + by = 0$, then $u = \dfrac{b^2 - a^2}{a^2 + b^2}$ and $v = -\dfrac{2ab}{a^2 + b^2}$.

What happens if we start with a flip matrix $\mathbf{F} = \begin{bmatrix} u & v \\ v & -u \end{bmatrix}$ as above, we consider the corresponding flip transformation (2.1.41), and we want to *find the line* $\mathscr{L}$? In other words, given u and v, can we find a and b? One possible approach to this question is given in Exercise 24. Another more direct approach is as follows. Start with a point P *not fixed by* $\mathbf{\Theta}$ (unless $\mathbf{\Theta}$ is the *identity*, there will be plenty of such points), then consider the point $Q = \mathbf{\Theta}(P)$, and then simply use the fact that the vector $\overrightarrow{PQ}$ *must be perpendicular to* $\mathscr{L}$. By the Frame Theorem, one possible choice for the pair of coefficients (a, b) would be obtained simply by taking the coordinates of the vector $\overrightarrow{PQ} = \begin{bmatrix} a \\ b \end{bmatrix}$. A calculation of this sort is shown in Example 2.1.10 below.

Example 2.1.8. As an easy application of the Flip Matrix Theorem, let us compute $\mathfrak{F}_{\mathscr{D}}$, the flip about the "diagonal" line $\mathscr{D}$, given by the equation $y = x$. Of course, we can re-write the equation of $\mathscr{D}$ as $x - y = 0$, so we can use a general equation with $a = 1$, $b = -1$ and $c = 0$. The numbers that appear in (2.1.41) are $m = \dfrac{b^2 - a^2}{a^2 + b^2} = \dfrac{(-1)^2 - 1^2}{1^2 + (-1)^2} = 0$, $n = -\dfrac{2ab}{a^2 + b^2} = -\dfrac{2 \cdot 1 \cdot (-1)}{1^2 + (-1)^2} = 1$, so the formula (2.1.41) yields:

$$\mathfrak{F}_{\mathscr{D}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

Example 2.1.9. Suppose we want to compute $\mathfrak{F}_{x\text{-axis}}$, the flip about the x-axis, which has an even simpler equation $y = 0$, which matches a general equation with $a = 0$, $b = 1$ and $c = 0$. The numbers that appear in (2.1.41) are $m = \frac{b^2 - a^2}{a^2 + b^2} = 1$, $n = -\frac{2ab}{a^2 + b^2} = 0$, so the formula (2.1.41) yields:

$$\mathfrak{F}_{x\text{-axis}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

Example 2.1.10. Someone hints to us that the transformation

$$\boldsymbol{\Theta}(x, y) = \left(\tfrac{3}{5}x - \tfrac{4}{5}y, -\tfrac{4}{5}x - \tfrac{3}{5}y\right),$$

is a flip transformation about a line $\mathscr{L}$ passing through the origin. We are asked to confirm this, and to find the line $\mathscr{L}$.

First of all, we can write $\boldsymbol{\Theta}$ in matrix form:

$$\boldsymbol{\Theta}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \frac{3}{5}x - \frac{4}{5}y \\ -\frac{4}{5}x - \frac{3}{5}y \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ -\frac{4}{5} & -\frac{3}{5} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

so the flip matrix candidate is: $\mathbf{M} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ -\frac{4}{5} & -\frac{3}{5} \end{bmatrix}$. It is pretty easy to see that $\mathbf{F}$ is indeed a flip matrix, so the only thing left to do is to find the line $\mathscr{L}$. Start for instance with the point $P(5, 0)$, so using the given formula for $\boldsymbol{\Theta}$, the point $Q(z, w) = \boldsymbol{\Theta}(P)$ has coordinates $z = 3$ and $w = -4$. In particular, the vector $\overrightarrow{PQ}$ – which is *perpendicular to* $\mathscr{L}$ has coordinates

$$\overrightarrow{PQ} = \begin{bmatrix} 3 \\ -4 \end{bmatrix} - \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}.$$

The fact that the vector $\begin{bmatrix} -2 \\ -4 \end{bmatrix}$ is perpendicular to $\mathscr{L}$, makes it possible to write a general equation for $\mathscr{L}$ as:[5]

$$(-2)x + (-4)y = 0.$$

If we wish, we can "clean up" the above equation, by diving everything by -2, so we can say that $\mathscr{L}$ is also given by the equation:

$$x + 2y = 0.$$

Up to this point we only discussed *flip* matrices. How about *rotation* matrices? Their geometric counterparts are as follows.

[5] General equations of lines *passing through the origin* are always of the form: $ax + by = 0$. (The constant term c is always equal to zero for such lines.)

Given a point C in the plane, **rotation about** C is a *composition* $\mathfrak{R} = \mathfrak{F}_{\mathscr{L}_1} \circ \mathfrak{F}_{\mathscr{L}_1}$ *of two flips*, that is, of the form

$$\mathfrak{R}(\overrightarrow{\mathrm{p}}) = \mathfrak{F}_{\mathscr{L}_1}(\mathfrak{F}_{\mathscr{L}_2}(\overrightarrow{\mathrm{p}})),$$

where $\mathfrak{F}_{\mathscr{L}_1}$ and $\mathfrak{F}_{\mathscr{L}_1}$ are two flipping transformations, about two lines $\mathscr{L}_1$ and $\mathscr{L}_2$ that *intersect at* C. The point C is called the **center of rotation**.

CLARIFICATIONS. When the two lines intersect at the origin, each of the two flip transformations can be presented as matrix multiplication, so using the Flip Matrix Theorem, it follows that we can write rotation transformations about the origin using a *product of two flip matrices*:

$$\mathfrak{R}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u_1 & v_1 \\ v_1 & -u_1 \end{bmatrix} \cdot \begin{bmatrix} u_2 & v_2 \\ v_2 & -u_2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

Using the associativity of matrix multiplication, we can write rotation transformations as a *single* matrix multiplication:

$$\mathfrak{R}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

where

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} u_1 & v_1 \\ v_1 & -u_1 \end{bmatrix} \cdot \begin{bmatrix} u_2 & v_2 \\ v_2 & -u_2 \end{bmatrix}.$$

As it turns out (see Exercise ??), *a rotation matrix is precisely a matrix that can be written as a product of two flip matrices*. Using this fact, we now get our second special version of the Orthogonal Matrix Theorem:

Rotation Matrix Theorem

A rotation about the origin is precisely a transformation that can presented in coordinates as multiplication by a rotation matrix, that is:

$$\mathfrak{R}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u & v \\ -v & u \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ux - vy \\ vx + uy \end{bmatrix} \tag{2.1.42}$$

where u and v are two constants satisfying $u^2 + v^2 = 1$.

Exercises

In Exercises 1-6 we assume a length unit is fixed, together with a square coordinate system. When asked to find various quantities, *use exact values!*

1. Find the equation of a circle of radius 5 (units) centered at $Z(2, -1)$.

2. Find the equation of the circle centered at $A(2, 3)$, which passes through the point $P(-4, 7)$.

3. Find the equation of a circle whose diameter has endpoints $P(2, 3)$ and $Q(1, -5)$.

4. Find the center and the radius of the circle represented by the equation $(x+3)^2+(y-2)^2 = 8$.

5. Find the center and the radius of the circle represented by the equation $4x^2 + (2y - 1)^2 = 9$.

6. Find the center and the radius of the circle represented by the equation $2x^2+4x+2y^2-y=3$.

In Exercises 7-9 we continue to assume a length unit and a square coordinate system are used.

7. Using bearing notation, find the direction that is *opposite* to S$12°34'$E.

8. Using bearing notation, find the two directions that are *perpendicular* to N$76°$W.

9. Start at $P(-3,1)$ and travel on a straight line in the N$45°$E direction until you intersect the y-axis. How long is your travel? What are the coordinates of the destination point? Use exact values.

10. By the Triangle Rule, you know that given points A, B, C, when you consider the vectors $\vec{\mathbf{u}} = \overrightarrow{AB}$, $\vec{\mathbf{v}} = \overrightarrow{BC}$, $\vec{\mathbf{w}} = \overrightarrow{AC}$, you must have the equality

$$\vec{\mathbf{u}} + \vec{\mathbf{v}} = \vec{\mathbf{w}}.$$

Compute all these vectors, in the case when your three points have coordinates $A(1,2)$, $B(3,4)$, $C(5,10)$, and verify the above equality.

11. Given the vectors $\vec{\mathbf{u}} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $\vec{\mathbf{v}} = \begin{bmatrix} 8 \\ -10 \end{bmatrix}$, find the following vectors and their magnitudes:
(i) $2\vec{\mathbf{u}} + 3\vec{\mathbf{v}}$;
(ii) $3\vec{\mathbf{u}} - 2\vec{\mathbf{v}}$.

12. Start at the origin and travel 5 miles in the N$45°$E direction, then travel 5 miles in the S$45°$E direction, then travel 5 miles West.
(i) What are the coordinates of your destination point? Use exact values.
(ii) How far is it from the origin? Use exact values.

13. Repeat the problem above, but with the following driving directions: start at the origin and travel 5 miles West, then 3 miles in the S$30°$E direction, then travel 5 miles in the N$60°$E direction. Use exact values for part (i). For part (ii) round to the nearest 0.001.

14. Find a general equation of a line which passes through the origin, and has $\vec{\mathbf{u}} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$ as a direction vector.

15. Find a non-zero vector that is perpendicular to the line given by

$$7x - 3y = 0.$$

16. Find a direction vector for the line given by

$$3x + 5y = 8.$$

17. Find a general equation of a line which passes through $P(1,2)$ and $Q(2,3)$.

18. Given a line $\mathscr{L}$ with equation
$$7x - 3y = 8,$$
find a general equation of a line $\mathscr{L}'$ which passes through $P(1,1)$ and is parallel to $\mathscr{L}$

19. Given a line $\mathscr{L}$ with equation
$$2x + 3y = 5,$$
find a general equation of a line $\mathscr{L}'$ which passes through $P(-1,1)$ and is perpendicular to $\mathscr{L}$

20. Find the equations of the horizontal and the vertical lines passing through $A(2,-3)$.

21. Find a general equation of a line $\mathscr{L}$ which passes through $B(0,1)$, and has a direction vector that points in the N$45°$E direction.

22. Find a general equation of a line $\mathscr{L}$ which passes through $C(-1,1)$, and has a direction vector that points in the N$45°$W direction.

The remaining Exercises are only for an expert user!

23*. **Proof of the Flip Matrix Theorem.** Let $\mathscr{L}$ be the line given by the equation $ax + by = 0$, and let $\mathfrak{F}_{\mathscr{L}}$ be the corresponding flip transformation. Fix now some point $P(x,y)$, and let $Q(z,w) = \mathfrak{F}_{\mathscr{L}}(P)$ be the image of P under $\mathfrak{F}_{\mathscr{L}}$. The Flip Matrix Theorem simply states that the following formulas hold:
$$\begin{cases} z = \dfrac{(b^2 - a^2)x - 2aby}{a^2 + b^2} \\[2ex] w = \dfrac{-2abx + (a^2 - b^2)y}{a^2 + b^2} \end{cases} \tag{2.1.43}$$
Use the steps below for deriving the above formulas. By the definition of the flip transformation, we know that: (A) $\overline{PQ}$ is *perpendicular to* $\mathscr{L}$, and (B) the *midpoint of* $\overline{PQ}$ *sits on* $\mathscr{L}$), using the steps below.

(i) Consider the position vectors $\overrightarrow{\mathbf{p}} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\overrightarrow{\mathbf{q}} = \begin{bmatrix} z \\ w \end{bmatrix}$ of our two points. Since by condition (A) it follows that the vector
$$\overrightarrow{PQ} = \overrightarrow{\mathbf{q}} - \overrightarrow{\mathbf{p}} = \begin{bmatrix} z - x \\ w - y \end{bmatrix}$$
is perpendicular to the line $\mathscr{L}$, so by the Frame Theorem, this vector is a scalar multiple of $\overrightarrow{\mathbf{n}} = \begin{bmatrix} a \\ b \end{bmatrix}$, so there is some number t, such that
$$\begin{bmatrix} z - x \\ w - y \end{bmatrix} = t \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ta \\ tb \end{bmatrix},$$
thus we have
$$\begin{cases} z = x + ta \\ w = y + tb \end{cases} \tag{2.1.44}$$

(ii) Compute the scalar t, using the midpoint condition (B), which tells us that the coordinates of the midpoint M of the segment $\overline{PQ}$ satisfy $ax_M + by_M = 0$. Using the midpoint formula, we know that the midpoint M has coordinates $x_M = \frac{1}{2}(x+z)$ and $y_M = \frac{1}{2}(y+w)$, so condition (B) will simply say that:

$$a\left(\tfrac{1}{2}(x+z)\right) + b\left(\tfrac{1}{2}(y+w)\right) = 0.$$

Replace z and w using (2.1.44), which will eventually yield an equation in t.

(iii) Replace t with what you found in step (ii), and derive the formulas (2.1.43).

(iv) Verify that the quantities $u = \dfrac{b^2 - a^2}{a^2 + b^2}$ and $v = -\dfrac{2ab}{a^2+b^2}$ do indeed satisfy the equality $u^2 + v^2 = 1$.

24*. Start with two constants satisfying $u^2 + v^2 = 1$, build a flip matrix $\mathbf{F} = \begin{bmatrix} u & v \\ v & -u \end{bmatrix}$, and consider the transformation

$$\boldsymbol{\Theta}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u & v \\ v & -u \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

(i) Show that the points $P(1+u, v)$, $Q(v, 1-u)$ and the origin $O(0,0)$ all sit on a line $\mathscr{L}$. (HINT: Show that the vectors $\overrightarrow{\mathbf{p}} = \begin{bmatrix} 1+u \\ v \end{bmatrix}$ and $\overrightarrow{\mathbf{q}} = \begin{bmatrix} v \\ 1-u \end{bmatrix}$ are multiples of each other.)

(ii) Show that $\boldsymbol{\Theta}$ coincides with the flip transformation $\mathfrak{F}_{\mathscr{L}}$.

25*. Find the formula for the flip $\mathfrak{F}_{\mathscr{L}}$ about a general line, given by the equation $ax + by = c$. HINT: Fix some point $A(s,t)$ on the line, and let $\overrightarrow{\mathbf{a}} = \begin{bmatrix} s \\ t \end{bmatrix}$ be its position vector. For any $\overrightarrow{\mathbf{p}}$ think of finding $\mathfrak{F}_{\mathscr{L}}(\overrightarrow{\mathbf{p}})$ as a three-step process: (i) translate everything by $-\overrightarrow{\mathbf{a}}$ (thus changing the line $\mathscr{L}$ to a new line $\mathscr{L}'$ that passes through the origin: find its equation!); (ii) flip about $\mathscr{L}'$; (iii) translate back by $\overrightarrow{\mathbf{a}}$. In other words, we can write $\mathfrak{F}_{\mathscr{L}}$ as a composition

$$\mathfrak{F}_{\mathscr{L}} = \mathfrak{T}_{\overrightarrow{\mathbf{a}}} \circ \mathfrak{F}_{\mathscr{L}'} \circ \mathfrak{T}_{-\overrightarrow{\mathbf{a}}},$$

so if you work in vector coordinates, your formula for $\mathfrak{F}_{\mathscr{L}}$ will be:

$$\mathfrak{F}_{\mathscr{L}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \mathfrak{F}_{\mathscr{L}'}\left(\begin{bmatrix} x-s \\ y-t \end{bmatrix}\right) + \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} u & v \\ v & -u \end{bmatrix} \cdot \begin{bmatrix} x-s \\ y-t \end{bmatrix} + \begin{bmatrix} s \\ t \end{bmatrix},$$

where $\begin{bmatrix} u & v \\ v & -u \end{bmatrix}$ is some flip matrix. In your final result, s and t should be eliminated, using the line equation $as + bt = c$.

26*. Prove that the product of two flip matrices is a rotation matrix.

27*. Prove that every rotation matrix $\begin{bmatrix} u & v \\ -v & u \end{bmatrix}$ can be written as the product of two matrices. (HINT: One of the two flips can be chosen to be $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.)

28*. Prove that a composition of two flips about parallel lines is a translation.

29*. Suppose you have a rotation $\mathfrak{R}$ with center $C(s,t)$, which might not be the origin! Consider the position vector $\vec{\mathbf{c}} = \begin{bmatrix} s \\ t \end{bmatrix}$ of the center, and the transformation

$$\mathfrak{R}^0 = \mathfrak{T}_{-\vec{\mathbf{c}}} \circ \mathfrak{R} \circ \mathfrak{T}_{\vec{\mathbf{c}}}.$$

Show that $\mathfrak{R}^0$ is a rotation *about the origin*. Conclude that we also have the equality

$$\mathfrak{R} = \mathfrak{T}_{\vec{\mathbf{c}}} \circ \mathfrak{R}^0 \circ \mathfrak{T}_{-\vec{\mathbf{c}}},$$

the rotation transformation $\mathfrak{R}$ can be presented as

$$\mathfrak{R}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \mathfrak{R}^0\left(\begin{bmatrix} x-s \\ y-t \end{bmatrix}\right) + \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} u & -v \\ v & u \end{bmatrix} \cdot \begin{bmatrix} x-s \\ y-t \end{bmatrix} + \begin{bmatrix} s \\ t \end{bmatrix},$$

where $\begin{bmatrix} u & -v \\ v & u \end{bmatrix}$ is some rotation matrix.

30*. **Proof of the Perpendicular Vectors Theorem.** Use the following steps.

(i) Start of by placing both vectors to start at the origin O, so their targets will be the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$. We now are dealing with a triangle $\triangle OP_1P_2$, for which the condition that the given vectors are *perpendicular*, is equivalent to the condition that: $\triangle OP_1P_2$ *is a right triangle, with hypotenuse*. In turn, by Pythagoras' Second Theorem, this condition is equivalent to the equality

$$P_1P_2^{\,2} = OP_1^{\,2} + OP_2^{\,2}. \tag{2.1.45}$$

(ii) Use the Distance Formula to Write each of $P_1P_2^{\,2}$, $OP_1^{\,2}$, $OP_2^{\,2}$ as algebraic expressions in x_1, y_1, x_2, y_2. This will now turn (2.1.45) into an *equation in* x_1, y_1, x_2, y_2.

(iii) Do a bit of Algebra on the equation you got in part (ii) and show that it is equivalent to:

$$x_1x_2 + y_1y_2 = 0.$$

31*. **Parallelogram Law.** Prove that, for any two vectors $\vec{\mathbf{v_1}}$ and $\vec{\mathbf{v_2}}$, one has the equality:

$$\|\vec{\mathbf{v_1}} + \vec{\mathbf{v_2}}\|^2 + \|\vec{\mathbf{v_1}} - \vec{\mathbf{v_2}}\|^2 = 2\|\vec{\mathbf{v_1}}\|^2 + 2\|\vec{\mathbf{v_2}}\|^2.$$

32*. Prove that the condition that two vectors $\vec{\mathbf{v_1}}$ and $\vec{\mathbf{v_2}}$ are perpendicular, is also equivalent to either one of the following two conditions:

(i) $\|\vec{\mathbf{v_1}} + \vec{\mathbf{v_2}}\|^2 = \|\vec{\mathbf{v_1}}\|^2 + \|\vec{\mathbf{v_2}}\|^2$.

(ii) $\left\|\vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1\right\|^2 = \left\|\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2\right\|^2$.

These final Exercises will help you prove the **Orthogonal Matrix Theorem.**

33*. Show that if an isometry $\boldsymbol{\Theta}$ fixes two distinct points P and Q, that is, if $\boldsymbol{\Theta}(P) = P$ and $\boldsymbol{\Theta}(Q) = Q$, then $\boldsymbol{\Theta}$ *fixes all points on the line* PQ. (HINT: Take R on the line PQ, and let $S = \boldsymbol{\Theta}(R)$. Consider the numbers $a = \text{dist}(P, Q)$, $x = \text{dist}(P, R) = \text{dist}(P, S)$ and $y = \text{dist}(Q, R) = \text{dist}(Q, S)$. The fact that R is on PQ means that *one of the three numbers* a, x, y *is equal to the sum of the other two*. In turn, this forces S to sit an PQ too, and finally this forces S to coincide with R.)

34*. Show that if an isometry $\boldsymbol{\Theta}$ fixes the vertices of one triangle, then $\boldsymbol{\Theta}$ *fixes all the points in the plane (in which case* $\boldsymbol{\Theta}$ *is the identity!)*. (HINT: Use the preceding Exercise, by showing first that $\boldsymbol{\Theta}$ fixes all the points on the perimeter of a a triangle, and then it fixes all points that sit on lines passing through two points on the perimeter.)

35*. Show that if an isometry $\boldsymbol{\Theta}$ fixes two points, then either $\boldsymbol{\Theta}$ *fixes all the points in the plane (in which case* $\boldsymbol{\Theta}$ *is the identity!)*, or $\boldsymbol{\Theta}$ *is a flip*.

36*. Show that if an isometry $\boldsymbol{\Theta}$ fixes a point C, then either $\boldsymbol{\Theta}$ *is a flip about a line passing through* C, or $\boldsymbol{\Theta}$ *is a rotation about* C. (HINT: If $\boldsymbol{\Theta}$ has another fixed point distinct from C, then use Exercise 35. If $\boldsymbol{\Theta}(C)$ has no other fixed point, pick some $P \neq C$, and let $Q = \boldsymbol{\Theta}(P)$, of which we know that $Q \neq P$. Using $\text{dist}(P, C) = \text{dist}(Q, C)$, it follows that the perpendicular bisector $\mathscr{L}$ of $\overline{PQ}$ *passes through* C. Consider then the flip $\mathfrak{F}_{\mathscr{L}}$ and the composed isometry $\boldsymbol{\Phi} = \boldsymbol{\Theta} \circ \mathfrak{F}_{\mathscr{L}}$. Show that $\boldsymbol{\Phi}$ fixes both C and Q, so Exercise 35 can be applied to $\boldsymbol{\Phi}$, so either $\boldsymbol{\Phi}$ is a flip, or the identity. Finally, note that, since $\mathfrak{F}_{\mathscr{L}}^{-1} = \mathfrak{F}_{\mathscr{L}}$, we can also write $\boldsymbol{\Theta} = \boldsymbol{\Phi} \circ \mathfrak{F}_{\mathscr{L}}$, so either $\boldsymbol{\Theta}$ is the flip $\mathfrak{F}_{\mathscr{L}}$, or it is a composition of two flips $\boldsymbol{\Theta} = \boldsymbol{\Phi} \circ \mathfrak{F}_{\mathscr{L}}$.)

2.2 The Analytic Construction of the Trigonometric Functions

Up to this point we only learned about the trigonometric functions of *acute angle measures*. It is now time to expand the definitions of the six trigonometric functions to arbitrary numbers.

Rotation Angles and Their Measures

In order to understand rotation angles, it is helpful to tell two short stories.

Imagine you are on a merry-go-round wheel at a playground, and someone (slowly) spins you for a certain period of time. Suppose you have a compass with you, so as you go around on the wheel, the needle on the compass will also spin. What you want to do then is to keep track of the *entire movement of the compass needle during the spin*. Alternatively, you may want to keep track of *how much you have turned on the wheel*.

Another way to look at this problem is to imagine you are a runner competing in a stadium that has a circular track, you run for an hour, and you want to know the distance (on the track) you covered, and also *which way* you ran.

What we will agree to call **radian measures**, or **degree measures**, or **turn measures**, etc. will be simply *real numbers*, without any limitations on how big/small they are, or whether they are positive or negative. When we are taking about a **rotation measure**, we will understand a mathematical quantity that can be expressed at the same time in radians, degrees, turns, etc., according to the usual conversion rules:

$$1 \text{ turn } = 2\pi(radians) = 360^\circ.$$

For instance a rotation measure of -450° can also be represented as $450 \cdot \dfrac{\pi}{180} = \dfrac{5\pi}{2}$ (radians), or as 1.25 turns. Likewise, a rotation measure of $-\dfrac{3\pi}{4}$ (radians) can also be represented as $-\dfrac{3\pi}{4} \cdot \left(\dfrac{180}{\pi}\right)^\circ = -135^\circ$.

Assume, as usual, we fixed a length unit, together with a square coordinate system.

> A **rotation angle** is a pair $\overset{\curvearrowleft}{\mathfrak{a}} = (\boldsymbol{s}, \alpha)$ consisting of
> - a *ray* $\boldsymbol{s}$ in the plane, hereafter referred to as the **initial side of $\overset{\curvearrowleft}{\mathfrak{a}}$**;
> - a *rotation measure* α, hereafter referred to as the **rotation measure of $\overset{\curvearrowleft}{\mathfrak{a}}$**
>
> Given such a rotation angle, we construct its **terminal side** by *rotating the initial side by* α, according to the following conventions:
> - if α is *positive*, the rotation is made *counter-clockwise*;
> - if α is *negative*, the rotation is made *clockwise*;
> - in either case, the net amount of rotation is $|\alpha|$, so when $\alpha = 0$, the terminal side coincides with the initial side.

The meaning of "counter-clockwise" and "clockwise" depends of the orientation of our square coordinate system. We will always agree that the positive x-axis *points to the right* (towards East), and the positive y-axis *points up* (towards North). If we have our coordinate system point the "wrong" way, then we have to understand that everything we draw is on a large glass window, so we can always "correct" our view, if we have to, by looking from the other side of the window.

CLARIFICATION AND ADDITIONAL TERMINOLOGY. Two rotation angles $\overset{\curvearrowleft}{\mathfrak{a}}$ and $\overset{\curvearrowleft}{\mathfrak{b}}$ are said to be **coterminal**, if

(a) their *initial sides coincide*, and also
(b) their *terminal sides also coincide*.

In practice, we do not like to have to deal with condition (b), so we can employ the following easy test instead.

> **Coterminal Angles Test**
>
> *Two rotation angles are coterminal, if and only if:*
> (a) their *initial sides coincide*, and also
> (b') their *rotation measures differ by an integer multiple of one turn*.

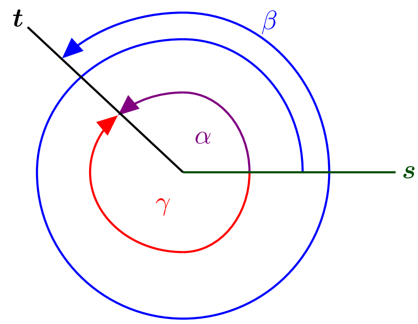


Figure 2.2.1

Example 2.2.1. In Figure 2.2.1 we see three coterminal rotation angles (all having initial side $\boldsymbol{s}$ and terminal side $\boldsymbol{t}$), with rotation measures $\alpha = 135°$, $\beta = 495°$, and $\gamma = -225°$. When we compare, for instance β and γ by taking their difference, we get

$$\beta - \gamma = 495° - (-225°) = 495° + 225° = 720° = 2 \cdot 360°.$$

To manufacture other rotation angles coterminal to these, all we need to do is to use rotation measures of the form

$$135° + n \cdot 360°, \text{ with } n \text{ arbitrary integer.}$$

For instance. we can use $135° + 10 \cdot 360° = 3735°$, or $135° - 7 \cdot 360° = -2385°$.

Of course, when we use radians, $360°$ must be replaced by 2π. For instance, the angles shown in Figure 2.2.1 have rotation measures (in radians) $\alpha = \frac{3\pi}{4}$, $\beta = \frac{11\pi}{4}$, and $\gamma = -\frac{5\pi}{4}$.

Transformations Associated with Rotation Angles

Assume we are given a rotation angle $\widehat{\mathfrak{a}} = (\boldsymbol{s}, \alpha)$. As described above, the process of constructing the *terminal side of* $\widehat{\mathfrak{a}}$ is carried on by means of a *rotation transformation* with center at the *vertex* (that is, at the source of the ray $\boldsymbol{s}$). We denote this rotation transformation by $\mathfrak{R}_{\widehat{\mathfrak{a}}}$, and we call it the **rotation determined by** $\widehat{\mathfrak{a}}$.

When we specialize to the case when the *vertex is at the origin*, we know from the previous Section that this transformation is given in vector coordinates (in matrix form) by:

$$\mathfrak{R}_{\widehat{\mathfrak{a}}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} u & -v \\ v & u \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.2.1}$$

where u and v are two numbers satisfying the equation $u^2+v^2 = 1$. As it can be seen geometrically, a rotation transformation as one given in (2.2.1) *only depends on the rotation measure* α. In other words, *it will act the same way, regardless where the initial side* $\boldsymbol{s}$ *is placed, as long as it source is at the origin*. So the matrix $\begin{bmatrix} u & -v \\ v & u \end{bmatrix}$ employed in (2.2.1), will only depend on α, and for this reason we are going to call it the **α-rotation matrix**, and we are going to denote it by $\mathbf{R}_\alpha$.

The Trigonometric Functions

The main observation that explains the analytic definition of the trigonometric functions is the following.

FACT A. *If α is an acute angle measure, then when we α-rotate the point $P_0(1,0)$ about the origin, we reach the point $P_\alpha(\cos\alpha, \sin\alpha)$.*

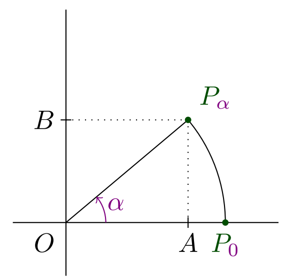


Figure 2.2.2

Everything follows from the fact that, when we take A, B to be the projections of P_α onto the x- and y-axis, we have a right triangle $\triangle OAP_\alpha$, which has hypotenuse $OP_\alpha = OP_0 = 1$, from which we immediately get $OA = \cos\,\alpha$ and $AP_\alpha = \sin\,\alpha$, which means that the coordinates of $P_\alpha(x,y)$ satisfy $|x| = OA = \cos\alpha$ and $|y| = OB = \sin\alpha$. Finally, because x and y are *positive*, we can get rid of absolute values.

What happens when we work with vectors? The position vector of P_0 is $\overrightarrow{\mathbf{p}_0} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, the position vector of P_α is $\overrightarrow{\mathbf{p}_\alpha} = \begin{bmatrix} \cos\,\alpha \\ \sin\,\alpha \end{bmatrix}$, and these two vectors are linked to the rotation matrix $\mathbf{R}_\alpha = \begin{bmatrix} u & -v \\ v & u \end{bmatrix}$ by the identity $\overrightarrow{\mathbf{p}_\alpha} = \mathbf{R}_\alpha \overrightarrow{\mathbf{p}_0}$, which means that we have the identity

$$\begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} = \begin{bmatrix} u & -v \\ v & u \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix},$$

and this way we can rephrase Fact A as:

FACT B. *If α is an acute angle measure, then the α-rotation matrix is:*

$$\mathbf{R}_\alpha = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \tag{2.2.2}$$

Example 2.2.2. Suppose we rotate the point $A(2,1)$ 60° about the origin, and we need to find the coordinates of the point Q that results from this rotation. Write the resulting point in coordinates $Q(x,y)$. If we use vector coordinates, for the position vectors, $\overrightarrow{\mathbf{a}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and the position vector $\overrightarrow{\mathbf{q}} = \begin{bmatrix} x \\ y \end{bmatrix}$, we know that that these vectors are linked using a rotation matrix by the equality $\overrightarrow{\mathbf{q}} = \mathbf{R}_{60^\circ}\,\overrightarrow{\mathbf{a}}$, so we get:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos 60^\circ & -\sin 60^\circ \\ \sin 60^\circ & \cos 60^\circ \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\cdot 2 + \left(-\frac{\sqrt{3}}{2}\right)\cdot 1 \\ \frac{\sqrt{3}}{2}\cdot 2 + \frac{1}{2}\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{\sqrt{3}}{2} \\ \sqrt{3} + \frac{1}{2} \end{bmatrix}$$

Using Facts A and B, as well as the Ratio and Reciprocal Identities (see Section 1.2) as guidelines, we can now construct our general trigonometric functions as follows.

Sine and Cosine of Arbitrary Angles

Given some rotation measure α, the numbers $\cos\alpha$ and $\sin\alpha$ are defined to be the x- and y-coordinates of the point P_α that is obtained by *α-rotating the point $P_0(1,0)$ about the origin*.

Equivalently, if we work in coordinates, *the vector* $\overrightarrow{\mathbf{p}_\alpha} = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$ *is the first column of the α-rotation matrix* $\mathbf{R}_\alpha$, so again the rotation matrix will be given by the formula (2.2.2)

Once sine and cosine are defined, we construct the other four trigonometric functions as follows.

Secant, Cosecant, Tangent and Cotangent for Arbitrary Angles

Given some rotation measure α, the other four trigonometric functions are defined as:

- $\sec\alpha = \dfrac{1}{\cos\alpha}$, *provided* $\cos\alpha \neq 0$;
- $\csc\alpha = \dfrac{1}{\sin\alpha}$, *provided* $\sin\alpha \neq 0$;
- $\tan\alpha = \dfrac{\sin\alpha}{\cos\alpha}$, *provided* $\cos\alpha \neq 0$;
- $\cot\alpha = \dfrac{\cos\alpha}{\sin\alpha}$, *provided* $\sin\alpha \neq 0$.

Outside the given provisions that accompany each one of these functions, that particular function is not defined!

CLARIFICATION. We can compute the values of the six trigonometric functions of a rotation measure α using the preceding definitions, as long as we are able to locate the "special" point P_α described in the definition of sine and cosine. The main characteristic of such a point is that P_α *always sits on the unit circle*, that is on the circle of radius 1 centered at the origin, which has equation

$$x^2 + y^2 = 1. \tag{2.2.3}$$

Equivalently, the position vector $\overrightarrow{\mathbf{p}_\alpha}$ has magnitude $\left\|\overrightarrow{\mathbf{p}_\alpha}\right\| = 1$. The "special" point P_α is completely determined using rotation angles of the following kind.

A rotation angle $\overset{\frown}{\alpha}$ is said to be in **standard position**, if *its initial side coincides with the positive x-axis*.

CLARIFICATIONS. A rotation angle in standard position is completely determined by its rotation measure. Furthermore, if such an angle has measure α, then the "special" point P_α is exactly the *point where the terminal side intersects the unit circle*. In turn, the "special" point P_α completely determines the associated α-rotation transformation and its matrix $\mathbf{R}_\alpha$. So now we can think of three objects which completely determine one another:

$$\boxed{\alpha\text{-rotation transformation about the origin}}$$
$$\updownarrow$$
$$\boxed{\text{rotation matrix } \mathbf{R}_\alpha}$$
$$\updownarrow$$
$$\boxed{P_\alpha \text{ point on the unit circle}}$$

Besides using rotation transformations, the "special" points P_α can also be located intuitively as follows. Assuming α is given in *radians*, in order to reach the point P_α, all we have to do is to "walk" α units on the unit circle, starting at the point $P_0(1, 0)$. (Of course, depending on the sign of α, we "walk" in the counterclockwise direction, if α is positive, and clockwise, otherwise.)

Example 2.2.3. Compute the six trigonometric functions of an angle in standard position, that has the point $P(-\frac{8}{17}, \frac{15}{17})$ on its terminal side.

Solution. We are very lucky here, because the point given to us is on the unit circle! (The equation (2.2.3) is clearly satisfied with $x = -\frac{8}{17}$ and $y = \frac{15}{17}$.) This means that what we are given here is nothing else by the "special" point P_α, and then the six trigonometric functions are computed very easily straight from the definitions:

$$\begin{aligned}
\cos\alpha &= -\frac{8}{17}; & \sin\alpha &= \frac{15}{17};\\
\sec\alpha &= \frac{1}{\cos\alpha} = -\frac{17}{8}; & \csc\alpha &= \frac{1}{\sin\alpha} = \frac{17}{15};\\
\tan\alpha &= \frac{\sin\alpha}{\cos\alpha} = \frac{\frac{15}{17}}{-\frac{8}{17}} = -\frac{15}{8}; & \cot\alpha &= \frac{\cos\alpha}{\sin\alpha} = \frac{-\frac{8}{17}}{\frac{15}{17}} = -\frac{8}{15}.
\end{aligned}$$

Unlike what we saw in the above Example, there many instances when we are not "lucky," so the "special" point P_α is not available, but instead we are only given a point sitting on the terminal side of our angle. In this case, the trigonometric functions of angles in standard position can be computed by the method shown below.

The Coordinate Method

Assume α is a rotation measure, which corresponds to a rotation angle $\overset{\frown}{\mathfrak{a}}$ in *standard position*, and $A(x, y)$ is some point *distinct from the origin and sitting on its terminal side*. Then the six trigonometric functions of α can be computed as follows:

I. Compute the number
$$r = \sqrt{x^2 + y^2}, \tag{2.2.4}$$
that is, the *distance from A to the origin*, which by assumption (that A is distinct for the origin) is *positive*.

II. Once r is known, the six trigonometric functions are given by
$$\begin{aligned} \cos\alpha &= \frac{x}{r}; & \sin\alpha &= \frac{y}{r}; \\ \sec\alpha &= \frac{r}{x}; & \csc\alpha &= \frac{r}{y}; \\ \tan\alpha &= \frac{y}{x}; & \cot\alpha &= \frac{x}{y}. \end{aligned}$$

As usual, the formulas giving secant, cosecant, tangent and cotangent are *only valid when the denominators are not equal to zero*.

CLARIFICATION. One way to see how these formulas come about is to observe that, once the terminal side $\boldsymbol{t}$ of $\overset{\frown}{\mathfrak{a}}$ is known, which is the same as knowing one point A on $\boldsymbol{t}$, we can always find the special point $P_\alpha(\cos\alpha, \sin\alpha)$, by observing that its position vector $\overrightarrow{\mathbf{p}_\alpha} = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$ *points in the same direction* as the position vector $\overrightarrow{\mathbf{a}} = \begin{bmatrix} x \\ y \end{bmatrix}$ of A.

In particular, it follows that $\overrightarrow{\mathbf{a}}$ is a *positive multiple of* $\overrightarrow{\mathbf{p}_\alpha}$, that is, we have
$$\begin{bmatrix} x \\ y \end{bmatrix} = r \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix},$$
for some positive number r, which will the give the equalities
$$\begin{cases} x &= r\cos\alpha \\ y &= r\sin\alpha \end{cases} \tag{2.2.5}$$
All we have observe now is that, since r is positive and $\overrightarrow{\mathbf{p}_\alpha}$ has magnitude 1, we have
$$\sqrt{x^2 + y^2} = \left\|\overrightarrow{\mathbf{a}}\right\| = \left\|r\overrightarrow{\mathbf{p}_\alpha}\right\| = |r| \cdot \left\|\overrightarrow{\mathbf{p}_\alpha}\right\| = r,$$
and then the equalities (2.2.5) yield
$$\begin{cases} \cos\alpha &= \dfrac{x}{r} \\ \sin\alpha &= \dfrac{y}{r} \end{cases}$$

from which everything else follows.

Example 2.2.4. Suppose α is the measure of a rotation angle in standard position, which has the point $A(-3,4)$ on its terminal side. We can compute the six trigonometric functions of α, using the above formulas with $x=-3$, $y=4$, which yield

$$r=\sqrt{(-3)^2+4^2}=\sqrt{9+16}=\sqrt{25}=5,$$

from which we immediately get:

$$\cos\alpha=\frac{x}{r}=-\frac{3}{5};\qquad \sin\alpha=\frac{y}{r}=\frac{4}{5};$$
$$\sec\alpha=\frac{r}{x}=-\frac{5}{3};\qquad \csc\alpha=\frac{r}{y}=\frac{5}{4};$$
$$\tan\alpha=\frac{y}{x}=-\frac{4}{3};\qquad \cot\alpha=\frac{x}{y}=-\frac{3}{4}.$$

The "Holy Grail" of Trigonometry

As it turns out, all the identities we discussed in Section 1.2 also hold for the general trigonometric functions.

The "Holy Grail" of Trigonometry

The Reciprocal Identities:

$$\sec\alpha=\frac{1}{\cos\alpha};\qquad \csc\alpha=\frac{1}{\sin\alpha};$$
$$\sin\alpha=\frac{1}{\csc\alpha};\qquad \cos\alpha=\frac{1}{\sec\alpha};$$
$$\tan\alpha=\frac{1}{\cot\alpha};\qquad \cot\alpha=\frac{1}{\tan\alpha}.$$

The Ratio Identities:

$$\sin\alpha=\frac{\tan\alpha}{\sec\alpha};\qquad \cos\alpha=\frac{\cot\alpha}{\csc\alpha};$$
$$\tan\alpha=\frac{\sin\alpha}{\cos\alpha};\qquad \cot\alpha=\frac{\cos\alpha}{\sin\alpha}.$$

The Product Identities:

$$\sin\alpha=\tan\alpha\cdot\cos\alpha;\qquad \cos\alpha=\cot\alpha\cdot\sin\alpha;$$
$$\tan\alpha=\sin\alpha\cdot\sec\alpha;\qquad \cot\alpha=\cos\alpha\cdot\csc\alpha.$$

The Pythagorean Identities:

$$\sin^2\alpha+\cos^2\alpha=1;$$
$$1+\tan^2\alpha=\sec^2\alpha;$$
$$1+\cot^2\alpha=\csc^2\alpha.$$

CLARIFICATIONS. Unlike what we saw in Section 1.2, proving these identities is fairly easy. On the one hand, the Reciprocal, Ratio and Product Identities now follow from the definitions. On the other hand, the Pythagorean Identities now follow from the main feature of rotation matrices: $u^2+v^2=1$.

The main novelty in the case of general functions is that some of the identities are *provisional*, so they only hold if all terms and operations are defined. For instance the identity $\cot\alpha = \dfrac{1}{\tan\alpha}$ will only work if both $\sin\alpha$ and $\cos\alpha$ are $\neq 0$. Other than that, the Identities are exactly the same as in Section 1.2.

The Quadrant Information

As we have just seen, the computation of the trigonometric functions of some rotation measure α is done most efficiently using the Coordinate Method outlined above, which uses angles in *standard position*, because all we need is a *point on the terminal side*. As it turns out, the "approximate" location of such a point is enough to give us some information on the **sign** of the trigonometric functions of α. To clarify this, all we have to do is to remember the following diagram.

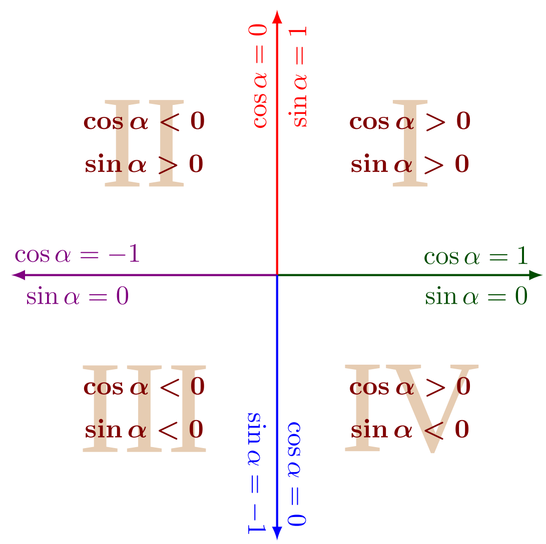


Figure 2.2.3

The figure above depicts four regions in the plane, hereafter referred to as **quadrants**, labeled **I** through **IV** as follows:

I: this quadrant consists of all points with *positive x- and y-coordinates*;
II: this quadrant consists of all points with *negative x-coordinate* and *positive y-coordinate*;
III: this quadrant consists of all points with *negative x- and y-coordinates*;
IV: this quadrant consists of all points with *positive x-coordinate* and *negative y-coordinate*.

So, if a point has both x- and y-coordinates non-zero, then the sign combination of the coordinates completely determines its quadrant location. We will agree that each half of a coordinate axis is shared by two neighboring quadrants. (For instance the poistive y-axis is shared by quadrants I and II).

We will also agree to say that a rotation measure α *sits in a certain quadrant*, if *the terminal side of its corresponding standard position angle sits in that quadrant*.

The quadrant information is absolutely necessary, if we need to solve the following type of a problem.

> **Basic Trigonometry Problem.** *Given one of the values $\sin\alpha$, $\cos\alpha$, $\tan\alpha$, $\cot\alpha$, $\sec\alpha$, α, plus the quadrant α sits in, find the other five values.*

CLARIFICATIONS. The main novelty here, when compared with the case (discussed in Section 1.2) when only acute angles are involved, is the fact that trigonometric functions *can have negative values*. In particular, we have to be careful with the identities derived from the Pythagorean Identities, which now have the following general form.

The Derived Pythagorean Identities

$$\cos\alpha = \pm\sqrt{1-\sin^2\alpha}; \qquad \sin\alpha = \pm\sqrt{1-\cos^2\alpha};$$
$$\sec\alpha = \pm\sqrt{1+\tan^2\alpha}; \qquad \csc\alpha = \pm\sqrt{1+\cot^2\alpha};$$
$$\tan\alpha = \pm\sqrt{\sec^2\alpha - 1}; \qquad \cot\alpha = \pm\sqrt{\csc^2\alpha - 1};$$

Exactly as we learned in Section 1.2, the Basic Trigonometry Problem can be solved by two methods.

Algebraic Method for Solving the Basic Trigonometry Problem

I. The value of one of the unknown five functions is the reciprocal of the value of the given function. Compute it!
II. Using either the given value, or the one computed in the previous step, compute the value of another one of the unknown functions, using one of the Derived Pythagorean Identities. *Use quadrant information to decide what sign you need to choose.*
III. Upon completing steps I and II you would have the values of three (of the six) trigonometric functions. The remaining three values are obtained using either the Reciprocal/Ratio Identities, or the Product Identities.

Example 2.2.5. Suppose α is an angle in the *second quadrant*, and $\tan\,\alpha = -2$. We will find the remaining five values, using the three steps from the Algebraic Method.

I. Using reciprocals, we immediately find $\cot\,\alpha = \dfrac{1}{\tan\,\alpha} = -\dfrac{1}{2}$.
II. Using the derived Pythagorean Identities, we can compute

$$\sec\,\alpha = \pm\sqrt{1+\tan^2\alpha} = \pm\sqrt{1+(-2)^2} = \pm\sqrt{5}.$$

Since α is in the *second quadrant*, we know that $\cos\,\alpha < 0$, so (remember that secant is the reciprocal of cosine), we also know that $\sec\,\alpha < 0$. This means that the correct value for the secant is: $\sec\,\alpha = -\sqrt{5}$.
III. Find the remaining three values using reciprocals and products:

$$\cos\,\alpha = \frac{1}{\sec\,\alpha} = -\frac{1}{\sqrt{5}};$$
$$\sin\,\alpha = \tan\,\alpha\cdot\cos\,\alpha = (-2)\left(-\frac{1}{\sqrt{5}}\right) = \frac{2}{\sqrt{5}};$$
$$\csc\,\alpha = \frac{1}{\sin\,\alpha} = \frac{\sqrt{5}}{2}.$$

The second method for solving our problem is based on the Coordinate Method, which requires that you "cook up" a point on the terminal side. (This is similar to the Geometric Method outlined in Section 1.2.)

Coordinate Method for Solving the Basic Trigonometry Problem

I. Use the given value of one trigonometric function of α, combined with the *quadrant information*, to produce a point on the terminal side, that *matches the value of the given trigonometric function*, which amounts to producing two numbers, out of x, y, or r.

II. Find the missing number, out of x, y, r, using the identity

$$r = \sqrt{x^2 + y^2},$$

combined with the *quadrant information*. (If the missing number is either x or y, the quadrant information is useful for determining its *sign*.)

III. Compute all other five missing values using the formulas given by the Coordinate Method.

Example 2.2.6. Let us redo Example 2.2.5 using the Coordinate Method.

I. We need to produce one point $A(x, y)$ on the terminal side of our angle, based on our given value. Based on the coordinate formulas, we know that the coordinates of A must satisfy the equality

$$\frac{y}{x} = \tan \alpha = -2.$$

Because A is in *quadrant II*, we know that x must be negative, and y must be positive, so a valid point A could have coordinates $x = -1$ and $y = 2$.

II. The missing number is

$$r = \sqrt{x^2 + y^2} = \sqrt{(-1)^2 + 2^2} = \sqrt{5}.$$

III. Using coordinates, the missing values are:

$$\cos \alpha = \frac{x}{r} = -\frac{1}{\sqrt{5}}; \qquad \sin \alpha = \frac{y}{r} = \frac{2}{\sqrt{5}};$$

$$\sec \alpha = \frac{r}{x} = -\sqrt{5}; \qquad \csc \alpha = \frac{r}{y} = \frac{\sqrt{5}}{2};$$

$$\cot \alpha = \frac{x}{y} = -\frac{1}{2}.$$

Example 2.2.7. Suppose α is an angle in the *fourth quadrant*, and $\sec \alpha = \frac{25}{7}$, and let us find the other five trigonometric functions, again using the Coordinate Method.

I. We need to produce one point $A(x, y)$ on the terminal side of our angle, based on our given value. Based on the coordinate formulas, we know that the coordinates of A must satisfy the equality

$$\frac{r}{x} = \sec \alpha = \frac{25}{7}.$$

Because A is in *quadrant IV*, we know that x must be positive, so a valid point A could have $x = 7$ and $r = 25$.

II. The missing number is y, which by the equality $r = \sqrt{x^2+y^2}$ must satisfy $25 = \sqrt{7^2+y^2}$, which yields $7^2+y^2 = 25^2$, or equivalently, $49+y^2 = 625$, which gives $y^2 = 625-49 = 576$, thus:

$$y = \pm\sqrt{576} = \pm 24.$$

Since our angle is in *quadrant IV*, we know that $y < 0$, so the the correct value is: $y = -24$.

III. Using coordinates, the missing values are:

$$\cos\,\alpha = \frac{x}{r} = \frac{7}{25}; \qquad \sin\,\alpha = \frac{y}{r} = -\frac{24}{25};$$

$$\csc\,\alpha = \frac{r}{y} = -\frac{25}{24};$$

$$\tan\,\alpha = \frac{y}{x} = -\frac{24}{7}; \qquad \cot\,\alpha = \frac{x}{y} = -\frac{7}{24}.$$

Values of Trigonometric Functions

So far, we learned how to compute the values of the trigonometric functions of some angle[6] α in one of the following instances:

- we are "lucky," so the "special" point P_α is given to us; or
- we know some point on the terminal side; or
- we know the value of one trigonometric function of α, together with the *quadrant* where α sits in.

What about the situation when α is given to us? For example, based on what we already know about the "familiar" acute angles, as well as some easy cases when P_α can be easily located, we can compile the following table, which contains what we are going to refer to as the **familiar values**:

α in radians	α in degrees	$\sin\,\alpha$	$\cos\,\alpha$	$\tan\,\alpha$	$\cot\,\alpha$	$\sec\,\alpha$	$\csc\,\alpha$
0	$0°$	0	1	0	undefined	1	undefined
$\frac{\pi}{6}$	$30°$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{3}}$	$\sqrt{3}$	$\frac{2}{\sqrt{3}}$	2
$\frac{\pi}{4}$	$45°$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	1	1	$\sqrt{2}$	$\sqrt{2}$
$\frac{\pi}{3}$	$60°$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{1}{\sqrt{3}}$	2	$\frac{2}{\sqrt{3}}$
$\frac{\pi}{2}$	$90°$	1	0	undefined	0	undefined	1

Table 2.2.1

How about other angles? As it turns out, the calculation of the values of the trigonometric functions

[6] When we say *the angle* α, we mean the standard position angle that corresponds to the rotation measure α.

can always be reduced to the calculation of certain values for *first quadrant angles*, as we shall see shortly. Before we clarify this matter, one easy observation is in order here:

Periodicity Property

If two rotation measures α and β *differ by an integer multiple of one turn measure*, then their trigonometric functions *agree*:

$$\begin{aligned} \sin\ \alpha &= \sin\ \beta; & \cos\ \alpha &= \cos\ \beta; \\ \sec\ \alpha &= \sec\ \beta; & \csc\ \alpha &= \csc\ \beta; \\ \tan\ \alpha &= \tan\ \beta; & \cot\ \alpha &= \cot\ \beta. \end{aligned}$$

CLARIFICATION. Saying that α and β *differ by an integer multiple of one turn measure* is the same as saying that their associated standard position angles *have identical terminal sides*, that is, these rotation angles are *coterminal*. For this reason, if we are in such a situation, we will allow ourselves to abuse the language a little bit and say that *α and β are coterminal*.

Example 2.2.8. To compute $\sin 750°$, we simply observe that $750° = 30° + 2 \cdot 360^2$, so $750°$ is coterminal to $30°$ (as they differ by 2 turns: $2 \cdot 360° = 720°$), thus $\sin 750° = \sin 30° = \frac{1}{2}$.

Likewise, when we want to compute $\tan\left(-\frac{23\pi}{4}\right)$, we use the fact that $-\frac{23\pi}{4} - \frac{\pi}{4} = -6\pi = -3(2\pi)$, so $-\frac{23\pi}{4}$ and $\frac{\pi}{4}$ are coterminal (as the differ by 3 turns: $3 \cdot (2\pi) = 6\pi$), thus $\tan\left(-\frac{23\pi}{4}\right) = \tan\frac{\pi}{4} = 1$.

CLARIFICATION. We can write the Periodicity Property is a concise algebraic way, depending on the unit we use (radians or degrees), as:

$$\text{function}(\alpha + 2\,n\,\pi) = \text{function}\,\alpha, \tag{2.2.6}$$
$$\text{function}(\alpha° + n \cdot 360°) = \text{function}\,\alpha°, \tag{2.2.7}$$

for any *integer* n, and function any one of six trigonometric functions cos, sin, tan, cot, sec, or csc.

As we hinted at the beginning of this topic, most calculations of values of trigonometric functions are tied up to those for first quadrant angles, the precise relation being described with the help of the following definition.

The **reference angle** α^{ref} of a standard position angle α is the *measure of the geometric angle formed by the terminal side with the "closest" half of the x-axis*.

CLARIFICATIONS. Reference angles are always *non-negative and no greater than $\frac{1}{4}$ turn*.

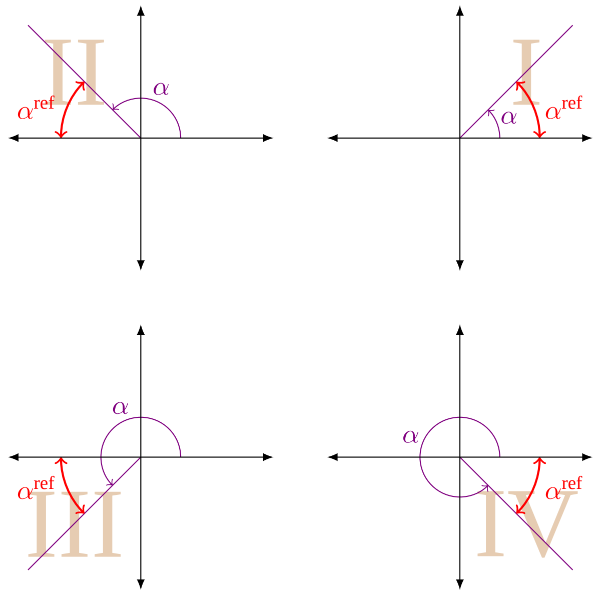


Figure 2.2.4

As suggested by Figure 2.2.4, depending of the quadrant α sits in, its relationship to the reference angle α^{ref}, is characterized as follows.

(I) α is in quadrant I, if and only if α *is coterminal to* α^{ref}.
(II) α is in quadrant II, if and only if α *is coterminal to* $\frac{1}{2}$ *turn* $-\alpha^{\text{ref}}$.
(III) α is in quadrant III, if and only if α *is coterminal to* $\frac{1}{2}$ *turn* $+\alpha^{\text{ref}}$.
(IV) α is in quadrant IV, if and only if α *is coterminal to* $-\alpha^{\text{ref}}$ *(or* 1 *turn* $-\alpha^{\text{ref}}$.

(Of course, when using radians, we replace $\frac{1}{2}$ turn with π, and 1 turn with 2π. Likewise, when using degrees, we replace $\frac{1}{2}$ turn with $180°$, and 1 turn with $360°$.)

The "big deal" about reference angles is contained in the following statement, which will be justified a little later.

Reference Angle Theorem

The values of the trigonometric functions of any rotation measure α coincide up to a possible sign change (depending on the quadrant) with those of the reference angle α^{ref}:

$$\sin\ \alpha = \pm \sin\ \alpha^{\text{ref}}; \qquad \cos\ \alpha = \pm \cos\ \alpha^{\text{ref}};$$
$$\sec\ \alpha = \pm \sec\ \alpha^{\text{ref}}; \qquad \csc\ \alpha = \pm \csc\ \alpha^{\text{ref}};$$
$$\tan\ \alpha = \pm \tan\ \alpha^{\text{ref}}; \qquad \cot\ \alpha = \pm \cot\ \alpha^{\text{ref}};$$

Example 2.2.9. Compute $\sin(-36225°)$ and $\cos(-36225°)$.

Solution. By the Periodicity property, adding an integer multiple of $360°$ will not change the values. This will happen for instance, if we add $36000° = 100 \cdot 360°$. In other words, the angle $\alpha = -36225°$ is coterminal with $-36225° + 36000° = -225°$. Adding another $360°$ will not change the values, so our angle α is also coterminal with $-225° + 360° = 135°$. Now we notice that $135° = 180° - 45°$, which tells us that

- the angle $\alpha = -36225°$ sits in quadrant II, and
- its reference angle is: $\alpha^{\text{ref}} = 45°$,

and consequently:

$$\sin(-36225^\circ) = \pm \sin 45^\circ \quad \text{and} \quad \cos(-36225^\circ) = \pm \cos 45^\circ.$$

Because α sits in the second quadrant, the correct values are:

$$\sin(-36225^\circ) = +\sin 45^\circ = \frac{1}{\sqrt{2}} \quad \text{and} \quad \cos(-36225^\circ) = -\cos 45^\circ = -\frac{1}{\sqrt{2}}.$$

The "Four Point Game"

As it turns out, the Reference Angle Theorem follows from a very simple geometric observation, which is summarized in Table 2.2.2 below, which explains what happens if we change an angle α, using certain transformations for which it is very easy to keep track of coordinates. For simplicity, we assume that we work in radians.

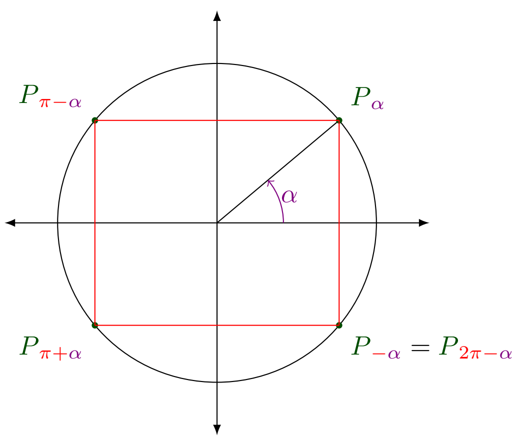


Figure 2.2.5

$\alpha \to$?	$P_\alpha \to$?	coordinates $\to$?
$\pi - \alpha$	reflect about y-axis	$(x, y) \to (-x, y)$
$\pi + \alpha$	rotate $\frac{1}{2}$ turn about origin	$(x, y) \to (-x, -y)$
$2\pi - \alpha$	reflect about x-axis	$(x, y) \to (x, -y)$
$-\alpha$	same as $2\pi - \alpha$	

Table 2.2.2

As a consequence of these features, we obtain the following formula packages, which are very useful when computing trigonometric functions.

Supplement Formulas:

$$\sin(\pi - \alpha) = \sin \alpha; \qquad \cos(\pi - \alpha) = -\cos \alpha;$$
$$\sec(\pi - \alpha) = -\sec \alpha; \qquad \csc(\pi - \alpha) = \csc \alpha;$$
$$\tan(\pi - \alpha) = -\tan \alpha; \qquad \cot(\pi - \alpha) = -\cot \alpha.$$

Add π Formulas:

$$\sin(\pi+\alpha) = -\sin\ \alpha; \qquad \cos(\pi+\alpha) = -\cos\ \alpha;$$
$$\sec(\pi+\alpha) = -\sec\ \alpha; \qquad \csc(\pi+\alpha) = -\csc\ \alpha;$$
$$\tan(\pi+\alpha) = \tan\ \alpha; \qquad \cot(\pi+\alpha) = \cot\ \alpha.$$

Formulas for Negatives:

$$\sin(-\alpha) = \sin(2\pi-\alpha) = -\sin\ \alpha; \qquad \cos(-\alpha) = \cos(2\pi-\alpha) = \cos\ \alpha;$$
$$\sec(-\alpha) = \sec(2\pi-\alpha) = \sec\ \alpha; \qquad \csc(-\alpha) = \csc(2\pi-\alpha) = -\csc\ \alpha;$$
$$\tan(-\alpha) = \tan(2\pi-\alpha) = -\tan\ \alpha; \qquad \cot(-\alpha) = \cot(2\pi-\alpha) = -\cot\ \alpha;$$

Even though Figure 2.2.5 illustrates only the special case when α is in quadrant I, similar figures can be drawn that cover all other possibilities, so no matter which quadrant α sits in, the four points P_α, $P_{\pi-\alpha}$, $P_{\pi+\alpha}$, and $P_{-\alpha}$ (which is same as $P_{2\pi-\alpha}$) will always form a rectangle centered at the origin, with sides parallel to the coordinate axes. Thus *the above identities, as well as the features collected in Table 2.2.2 are still valid, for any value of α, regardless of quadrant!*

Exercises

In Exercises 1–5, find all six trigonometric functions of the angle α, based on the fact that a certain point A sits on its corresponding standard position angle. Use exact values!

1. $A(3, -4)$.
2. $A(-8, 15)$.
3. $A(-12, -16)$.
4. $A(-8, 0)$.
5. $A(0, -15)$.

In Exercises 6–15, find the values of all remaining trigonometric functions of the angle α, based on a given particular value, and the given quadrant information. Use exact values!

6. $\sin\alpha = \frac{1}{3}$, α in quadrant I.
7. $\sin\alpha = \frac{4}{5}$, α in quadrant II.
8. $\sin\alpha = -\frac{3}{5}$, α in quadrant III.
9. $\sin\alpha = -\frac{8}{9}$, α in quadrant IV.
10. $\cos\alpha = -\frac{2}{7}$, α in quadrant II.
11. $\cos\alpha = -\frac{12}{13}$, α in quadrant III.
12. $\tan\alpha = \frac{5}{7}$, α in quadrant III.
13. $\tan\alpha = -3$, α in quadrant IV.

14. $\sec\alpha = 5$, α in quadrant IV.

15. $\csc\alpha = 10$, α in quadrant II.

In Exercises 16–19, find the reference angle α^{ref} for the given angle.

16. $\alpha = 1234°$

17. $\alpha = -2013°$

18. $\alpha = 12$ (radians).

19. $\alpha = -30$ (radians).

In Exercises 20–28, find the exact values of $\sin\alpha$ and $\cos\alpha$, for the given angle.

20. $\alpha = -405°$

21. $\alpha = 300°$

22. $\alpha = 330°$

23. $\alpha = \frac{123\pi}{4}$ (radians)

24. $\alpha = -\frac{13\pi}{6}$ (radians)

25. $\alpha = 1000\pi$ (radians)

26. $\alpha = \frac{52\pi}{3}$ (radians)

27. $\alpha = -\frac{12345\pi}{2}$ (radians)

28. $\alpha = \frac{54321\pi}{2}$ (radians)

29. Given $\sin\alpha = -\frac{1}{4}$, find the exact values of $\sin(-\alpha)$, $\sin(\pi - \alpha)$, and $\sin(\pi + \alpha)$.

30. Given $\tan\alpha = 17$, find the exact values of $\tan(-\alpha)$, $\tan(\pi - \alpha)$, and $\tan(\pi + \alpha)$.

31. Given $\sec\alpha = -8$, find the exact values of $\cos(123\pi - \alpha)$ and $\cos(123\pi + \alpha)$.

In Exercise 32–35, find $\sin\alpha$ and $\cos\alpha$, based on the given information about the terminal side and the quadrant of the corresponding standard position angle. Use exact values.

32. The angle is in quadrant III and the terminal side is on the line $4x - 3y = 0$.

33. The angle is in quadrant II and the terminal side is parallel to the line $4x + 3y = 7$.

34. The angle is in quadrant IV and the terminal side is parallel to the line $2x + 3y = 5$.

35. The angle is in quadrant III and the terminal side is perpendicular to the line $12x + 5y = 100$.

2.3 Trigonometric Identities

As we learned in Algebra, an **identity** is an equality of the form

$$\textit{(Left) Expression} = \textit{(Right) Expression},$$

where each side is an "*expression*," that is some quantity that can be written using algebraic operations and involving certain *functions and variables*. The most common functions used in the identities we learned about in Algebra are *powers*, *radicals*, *exponentials*, and *radicals*. Here is a sample list of identities we are all familiar with from the Algebra course:

$$(x+y)(x-y) = x^2 - y^2; \tag{2.3.1}$$
$$(x+y)^2 = x^2 + 2xy + y^2; \tag{2.3.2}$$
$$(x-y)^2 = x^2 - 2xy + y^2; \tag{2.3.3}$$
$$\left(\sqrt[n]{x}\right)^n = x; \tag{2.3.4}$$
$$\sqrt{x^2} = |x|. \tag{2.3.5}$$

The identities (2.3.1), (2.3.2) and (2.3.3) are example of identities in *two variables*. The identities (2.3.4) and (2.3.5) are examples of identities in *one variable*.

To **verify an identity** means to show that *both expressions are equal, whenever both of them are defined*.

CLARIFICATION. When looking at the above list of examples, the clause "*whenever both (sides) are defined*" will be meaningful only when dealing with (2.3.4), in which the left-hand side is not always defined. (When n is *even*, the left-hand side is only defined, when $x \geq 0$.)

A **trigonometric identity** is one in which both sides involve algebraic operations, common algebraic functions, and trigonometric functions.

The basic "toolkit"

When verifying trigonometric identities, our basic "toolkit" will consist, at a bare minimum, of the packages that make the "Holy Grail" of Trigonometry, that is, the *reciprocal, ratio, product and Pythagorean identities*. As we progress in the Trigonometry course, we will expand our "toolkit" to include more and more formulas. For instance, based on the "Four Point Game" we obtained three additional packages:

Supplement Formulas

$$\sin(\pi-\alpha) = \sin\,\alpha; \qquad \cos(\pi-\alpha) = -\cos\,\alpha;$$
$$\sec(\pi-\alpha) = -\sec\,\alpha; \qquad \csc(\pi-\alpha) = \csc\,\alpha;$$
$$\tan(\pi-\alpha) = -\tan\,\alpha; \qquad \cot(\pi-\alpha) = -\cot\,\alpha.$$

Add π Formulas

$$\sin(\pi+\alpha) = -\sin\alpha; \qquad \cos(\pi+\alpha) = -\cos\alpha;$$
$$\sec(\pi+\alpha) = -\sec\alpha; \qquad \csc(\pi+\alpha) = -\csc\alpha;$$
$$\tan(\pi+\alpha) = \tan\alpha; \qquad \cot(\pi+\alpha) = \cot\alpha.$$

Formulas for Negatives

$$\sin(-\alpha) = \sin(2\pi-\alpha) = -\sin\alpha; \qquad \cos(-\alpha) = \cos(2\pi-\alpha) = \cos\alpha;$$
$$\sec(-\alpha) = \sec(2\pi-\alpha) = \sec\alpha; \qquad \csc(-\alpha) = \csc(2\pi-\alpha) = -\csc\alpha;$$
$$\tan(-\alpha) = \tan(2\pi-\alpha) = -\tan\alpha; \qquad \cot(-\alpha) = \cot(2\pi-\alpha) = -\cot\alpha;$$

Lastly, one more package is available, inspired by what we learned in Section 1.2 about *co-functions*

Complement Formulas

$$\sin\left(\frac{\pi}{2}-\alpha\right) = \cos\alpha; \qquad \cos\left(\frac{\pi}{2}-\alpha\right) = \sin\alpha;$$
$$\sec\left(\frac{\pi}{2}-\alpha\right) = \csc\alpha; \qquad \csc\left(\frac{\pi}{2}-\alpha\right) = \sec\alpha;$$
$$\tan\left(\frac{\pi}{2}-\alpha\right) = \cot\alpha; \qquad \cot\left(\frac{\pi}{2}-\alpha\right) = \tan\alpha.$$

CLARIFICATION. The Complement Formulas hold for *any angle*, not just for acute ones. The easy way to see why they work is to see what transformations are employed, when we want to pass from the "special" point P_α to the "special" point $P_{\frac{\pi}{2}-\alpha}$.

I. First, we move from P_α to $P_{-\alpha}$, which means that we do a *reflection about the x-axis*, so if we write vector coordinates $\vec{\mathbf{p}}_\alpha = \begin{bmatrix} x \\ y \end{bmatrix}$, then $\vec{\mathbf{p}}_{-\alpha} = \begin{bmatrix} x \\ -y \end{bmatrix}$.

II. Second, we rotate by $\pi 2$, so we multiply by the rotation matrix $\vec{\mathbf{p}}_{\frac{\pi}{2}-\alpha} = \mathbf{R}_{\frac{\pi}{2}}\vec{\mathbf{p}}_{-\alpha}$, so the position vector of $P_{\frac{\pi}{2}-\alpha}$ will be

$$\vec{\mathbf{p}}_{\frac{\pi}{2}-\alpha} = \begin{bmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix} \cdot \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix},$$

so the coordinates of $P_{\frac{\pi}{2}-\alpha}$, that is $\cos\left(\frac{\pi}{2}-\alpha\right)$ and $\sin\left(\frac{\pi}{2}-\alpha\right)$, are the same as those of P_α, that is, $\cos\alpha$ and $\sin\alpha$, *except that they are flipped*. In other words, we have the identities

$$\sin\left(\frac{\pi}{2}-\alpha\right) = \cos\alpha,$$
$$\cos\left(\frac{\pi}{2}-\alpha\right) = \sin\alpha,$$

from which the other four immediately follow.

Verifying Identities by the Left-to-Right (or Right-to-Left) Method

The most direct method for verifying a trigonometric identity is to *start with one side, and transform it (in several steps) into the other side*.

Example 2.3.1. Suppose we want to verify the identity

$$\tan t + \cot t = \frac{1}{\sin t \cos t}.$$

We start with the left-hand side (hereafter abbreviated as LHS), and transform it into the right-hand side (hereafter abbreviated as RHS) in four steps:

$$\begin{aligned}
\text{LHS} &= \tan t + \cot t \\
&= \frac{\sin t}{\cos t} + \frac{\cos t}{\sin t} && \leftarrow \boxed{\text{ratio identities}} \\
&= \frac{\sin^2 t}{\sin t \cos t} + \frac{\cos^2 t}{\sin t \cos t} && \leftarrow \boxed{\text{equivalent fractions, with same denominator}} \\
&= \frac{\sin^2 t + \cos^2 t}{\sin t \cos t} && \leftarrow \boxed{\text{combine numerators}} \\
&= \frac{1}{\sin t \cos t} && \leftarrow \boxed{\text{the Pythagorean Identity } \sin^2 t + \cos^2 t = 1} \\
&= \text{RHS}. && \leftarrow \boxed{\text{DONE!}}
\end{aligned}$$

Example 2.3.2. Consider the identity

$$\csc x\big(\sin(-x) + \csc x\big) = \cot^2 x$$

As it is always better if we deal with functions of one angle only, our first step in simplifying LHS will get rid of $\sin(-x)$:

$$\begin{aligned}
\text{LHS} &= \csc x\big(\sin(-x) + \csc x\big) = \csc x(-\sin x + \csc x) && \leftarrow \boxed{\begin{array}{l}\text{formula for negatives:} \\ \qquad \sin(-x) = -\sin x\end{array}} \\
&= -\csc x \sin x + \csc^2 x = -1 + \csc^2 x && \leftarrow \boxed{\begin{array}{l}\text{multiply, then use reciprocal identity } \sin x = \dfrac{1}{\csc x}, \\ \text{which gives } \csc x \sin x = 1\end{array}} \\
&= -1 + (1 + \cot^2 x) = -1 + 1 + \cot^2 x = \cot^2 x && \leftarrow \boxed{\begin{array}{l}\text{use Pythagorean identity} \\ \qquad \csc^2 x = 1 + \cot^2 x, \\ \text{open up, then cancel the 1's}\end{array}} \\
&= \text{RHS}. && \leftarrow \boxed{\text{DONE!}}
\end{aligned}$$

In certain instances, the identity we need to verify is quite complicated, and there is no obvious way to transform one side into the other. One choice is such cases would be to *transform the entire identity* into a new identity, which is *equivalent to the original identity*.

Example 2.3.3. Consider the identity

$$\frac{\sec u + 1}{\tan u} = \frac{\tan u}{\sec u - 1}.$$

Since this identity is a *proportion*, we can write an identity equivalent to it, but which is in fraction-less form:[7]

$$(\sec u + 1)(\sec u - 1) = \tan^2 u.$$

This new identity (which is *equivalent* to the given one) can be verified by transforming the left-hand side into the right-hand side in three easy steps:

$$\begin{aligned}
\text{LHS} &= (\sec u + 1)(\sec u - 1) \\
&= \sec^2 u - 1 \qquad \leftarrow \boxed{\text{difference of squares: } (A+B)(A-B) = A^2 - B^2} \\
&= (1 + \tan^2 u) - 1 \qquad \leftarrow \boxed{\text{the Pythagorean Identity } \sec^2 u = 1 + \tan^2 u} \\
&= 1 + \tan^2 u - 1 = \tan^2 u \qquad \leftarrow \boxed{\text{“open up” parentheses, then cancel the 1's}} \\
&= \text{RHS}. \qquad \leftarrow \boxed{\text{DONE!}}
\end{aligned}$$

The “Three-Way” Method

When handling more complex identities, there are many instances when we do not see any clear way for transforming one side into the other, so our approach will be to “divide and conquer.” So if we look at an identity of the form

$$\text{LHS} = \text{RHS}$$

what we may try is to *transform each side separately into newer/simpler expressions*, with the expectation that *the newer expressions coincide*.

Example 2.3.4. Consider the identity

$$(\cos\theta + 1)(\sec\theta - 1) = \cos\theta\tan^2\theta,$$

which we are going to verify by the “three-way” method, in which we work on each side separately:

$$\begin{aligned}
\text{LHS} &= (\cos\theta + 1)(\sec\theta - 1) \\
&= \cos\theta\sec\theta - \cos\theta + \sec\theta - 1 \qquad \leftarrow \boxed{\text{multiply out (fold)}} \\
&= 1 - \cos\theta + \frac{1}{\cos\theta} - 1 \qquad \leftarrow \boxed{\text{use reciprocals: } \sec\theta = \frac{1}{\cos\theta}, \text{ so } \cos\theta\sec\theta = 1} \\
&= -\cos\theta + \frac{1}{\cos\theta} = -\frac{\cos^2\theta}{\cos\theta} + \frac{1}{\cos\theta} = \frac{1 - \cos^2\theta}{\cos\theta} \qquad \leftarrow \boxed{\text{cancel 1's and subtract fractions}} \\
&= \frac{\sin^2\theta}{\cos\theta} \qquad \leftarrow \boxed{\text{Pythagorean identity: } 1 - \cos^2\theta = \sin^2\theta}
\end{aligned}$$

$\boxed{\text{Stop here: this is your (new) LHS.}}$

---

[7] We know that $\frac{\heartsuit}{\spadesuit} = \frac{\diamondsuit}{\clubsuit}$ can be transformed by cross-multiplication into an equivalent form: $\heartsuit \cdot \clubsuit = \spadesuit \cdot \diamondsuit$.

Now we work on the right-hand side:

$$\begin{aligned}
\text{RHS} &= \cos\theta\tan^2\theta \\
&= \cos\theta\left(\frac{\sin\theta}{\cos\theta}\right)^2 = \cos\theta\cdot\frac{\sin^2\theta}{\cos^2\theta} && \leftarrow \boxed{\text{use ratio identity } \tan\theta = \frac{\sin\theta}{\cos\theta}} \\
&= \frac{\cos\theta\sin^2\theta}{\cos^2\theta} = \frac{\sin^2\theta}{\cos\theta} && \leftarrow \boxed{\text{multiply and simplify}} \\
&= \text{(new) LHS} = \text{(old) LHS} && \leftarrow \boxed{\text{DONE!}}
\end{aligned}$$

Example 2.3.5. Consider the identity

$$\frac{\cos\alpha}{\sin\alpha\sec\alpha} = \csc\alpha - \sin\alpha.$$

We will simplify each side to a form that uses only sine and cosine. For a change, we start this time with the right-hand side

$$\begin{aligned}
\text{RHS} &= \csc\alpha - \sin\alpha \\
&= \frac{1}{\sin\alpha} - \sin\alpha && \leftarrow \boxed{\text{replace } \csc\alpha = \frac{1}{\sin\alpha}} \\
&= \frac{1}{\sin\alpha} - \frac{\sin^2\alpha}{\sin\alpha} = \frac{1-\sin^2\alpha}{\sin\alpha} && \leftarrow \boxed{\text{manipulate fractions}} \\
&= \frac{\cos^2\alpha}{\sin\alpha} && \leftarrow \boxed{\text{use Pythagorean identity } 1-\sin^2\alpha = \cos^2\alpha} \\
& \boxed{\text{Stop here: this is your (new) RHS.}}
\end{aligned}$$

Next we work on the left-hand side

$$\begin{aligned}
\text{LHS} &= \frac{\cos\alpha}{\sin\alpha\sec\alpha} \\
&= \frac{\cos\alpha}{\sin\alpha\cdot\dfrac{1}{\cos\alpha}} && \leftarrow \boxed{\text{replace } \sec\alpha = \frac{1}{\cos\alpha}} \\
&= \frac{\cos\alpha}{\dfrac{\sin\alpha}{\cos\alpha}} = \frac{\cos\alpha}{1}\cdot\frac{\cos\alpha}{\sin\alpha} = \frac{\cos^2\alpha}{\sin\alpha} && \leftarrow \boxed{\text{manipulate fractions}} \\
&= \text{(new) RHS} = \text{(old) RHS}. && \leftarrow \boxed{\text{DONE!}}
\end{aligned}$$

TIPS:
1. The "three-way" method is preferred, unless one side is simple enough, in which case you can use the left-to-right (or right-to-left) method.
2. When simplifying each side, aim at expressions that use as few trigonometric functions as possible.
3. The preferred simplified expressions are those using only sine & cosine. Other "nice" pairs of functions one can use are tangent & secant, or cotangent & cosecant.

False Identities

There are many instances when, although we are tempted to treat certain equalities as identities, what we see in fact are *false identities*, that is, *identities that are not true*. As a simple rule, *in order to show that an equality is a false identity, all we need is one value for the variable(s), for which the equality fails*.

Example 2.3.6. Consider the equality

$$\cos\left(x+\frac{\pi}{2}\right)=\cos x+\cos\frac{\pi}{2}. \tag{2.3.6}$$

In order to show that this is a false identity, all we need is one value of x, that "breaks" the equality. The simplest value is, for instance $x=\frac{\pi}{2}$. Indeed, when we compute both sides of (2.3.6), we get:

$$\begin{aligned}\text{LHS}&=\cos\left(\frac{\pi}{2}+\frac{\pi}{2}\right)=\cos\pi=-1;\\ \text{RHS}&=\cos\frac{\pi}{2}+\cos\frac{\pi}{2}=0+0=0.\end{aligned}$$

CLARIFICATION. As we saw in the above Example, the equality (2.3.6) is not an identity! One can still ask if it is possible to "fix" a "broken" (false) identity. Looking at (2.3.6), we can in fact think of two ways of "fixing" it: either

(A) keep the left-hand side as it is, and try to find a right-hand side expression that matches it, or more interestingly,

(B) find the values of x (if any), for which the equality does work!

When approaching this matter as in (B), what we are in fact doing is treating (2.3.6) as an **equation**, which we then try to solve for x. Actually, both plans (A) and (B) can be carried on at the same time, based on the formulas for complements and for negatives, because we know that

$$\cos\left(x+\frac{\pi}{2}\right)=\cos\left(\frac{\pi}{2}-(-x)\right)=\sin(-x)=-\sin x.$$

So if we want to "fix" (2.3.6) into a true identity, we can write:

$$\cos\left(x+\frac{\pi}{2}\right)=-\sin x, \tag{2.3.7}$$

which is a valid identity. At the same time, if we treat (2.3.6) as an *equation*, then using (2.3.7) as well as the obvious simplification RHS$=\cos x$ (based on $\cos\frac{\pi}{2}=0$), we can rewrite it as

$$-\sin x=\cos x, \tag{2.3.8}$$

which is equivalent to:

$$\tan x = -1. \tag{2.3.9}$$

So now we see, that solving (2.3.6) is equivalent to solving (2.3.9), which is one of the basic trigonometric equations we will learn about in Section 2.7.

Why are the equations (2.3.8) and (2.3.9) really equivalent? It appears that when getting from the first equation to the second one we divide by $\cos x$, so we need to be careful with the possibility that $\cos x = 0$. However, when $\cos x = 0$, it is impossible to have $\sin x = 0$ at the same time (by the Pythagorean identity $\sin^2 x + \cos^2 x = 1$), and this means that *when dividing by $\cos x$ we do not loose any of the solutions of* (2.3.8), because for every solution of (2.3.8) it is impossible to have $\cos x = 0$.

CLARIFICATION. The same way we derived the identity (2.3.7), by writing $x + \frac{\pi}{2} = \frac{\pi}{2} - (-x)$ and using the formulas for complements and for negatives, we can upgrade our "toolbox" with the following formula package.

Anti-Complement Formulas

$$\begin{aligned}
\sin\left(\frac{\pi}{2}+\alpha\right) &= \cos\,\alpha; & \cos\left(\frac{\pi}{2}+\alpha\right) &= -\sin\,\alpha;\\
\sec\left(\frac{\pi}{2}+\alpha\right) &= -\csc\,\alpha; & \csc\left(\frac{\pi}{2}+\alpha\right) &= \sec\,\alpha;\\
\tan\left(\frac{\pi}{2}+\alpha\right) &= -\cot\,\alpha; & \cot\left(\frac{\pi}{2}+\alpha\right) &= -\tan\,\alpha.
\end{aligned}$$

We use the word "anti-complement" as a shortcut for "complement of the negative." One can also consider the *negative complement* of some α, that is, $\alpha - \frac{\pi}{2} = -\left(\frac{\pi}{2} - \alpha\right)$, for which we can easily derive the following identities.

Negative Complement Formulas

$$\begin{aligned}
\sin\left(\alpha-\frac{\pi}{2}\right) &= -\cos\,\alpha; & \cos\left(\alpha-\frac{\pi}{2}\right) &= \sin\,\alpha;\\
\sec\left(\alpha-\frac{\pi}{2}\right) &= \csc\,\alpha; & \csc\left(\alpha-\frac{\pi}{2}\right) &= -\sec\,\alpha;\\
\tan\left(\alpha-\frac{\pi}{2}\right) &= -\cot\,\alpha; & \cot\left(\alpha-\frac{\pi}{2}\right) &= -\tan\,\alpha.
\end{aligned}$$

Example 2.3.7. Consider the equality

$$\sqrt{1-\cos^2 t} = \sin t \tag{2.3.10}$$

Although this equality almost looks right, it is still false, because the value $t = -\frac{\pi}{2}$ "breaks" it. Indeed, using the formulas for negatives we have

$$\cos\left(-\frac{\pi}{2}\right) = \cos\frac{\pi}{2} = 0 \text{ and } \sin\left(-\frac{\pi}{2}\right) = -\sin\frac{\pi}{2} = -1,$$

so when we compute the two sides of (2.3.10) for this particular value of t, we get

$$\text{LHS} = \sqrt{1-\cos^2\left(-\frac{\pi}{2}\right)} = \sqrt{1-0^2} = \sqrt{1} = 1;$$
$$\text{RHS} = \sin^2\left(-\frac{\pi}{2}\right) = -1.$$

If we view (2.3.10) as an equation, then unlike what we saw in the preceding Example, we will have an abundance of solutions, which can be "bunched in intervals," as we shall see shortly. For this reason, treating (2.3.10) as an equation is not interesting, so the better way to "fix" it is by coming up with the correct identity. One easy way to get our "fix" is to use the Pythagorean identity $1-\cos^2 t = \sin^2 t$, combined with the basic algebra identity (2.3.5), so the correct identity in place of (2.3.10) is

$$\sqrt{1-\cos^2 t} = \left|\sin t\right| \tag{2.3.11}$$

which does work for all values of t. When comparing the right-hand side of the correct identity (2.3.11) with the right-hand side of (2.3.10), we see that we would get a good match, precisely when $\sin t = \left|\sin t\right|$, which is the same as saying that $\sin t \geq 0$. Since we know exactly $\sin t \geq 0$, one way to "fix" (2.3.10) is also to write:

$$\sqrt{1-\cos^2 t} = \sin t, \text{ if } t \text{ is in quadrants I or II.} \tag{2.3.12}$$

The above presentation is an example of a valid **conditional identity**. We now see that, when thinking (2.3.10) as an *equation*, it will have as solutions, for example *all t in $[0,\pi]$*, but also, *all t in $[2\pi,3\pi]$*, *all t in $[4\pi,5\pi]$*, etc. and also *all t in $[-2\pi,-\pi]$*, *all t in $[-4\pi,-3\pi]$*, etc. (More on this "huge" solution set will be clarified in Section 2.4.)

Exercises

In Exercises 1-13, you are asked to verify the given identity.

1. $\sec t - \cos t = \tan t \sin t$.

2. $\cos x + \sin x \tan x = \sec x$.

3. $\dfrac{\csc^2\alpha - 1}{\csc^2\alpha} = \cos^2\alpha$.

4. $(\tan s + \cot s)(\sin s + \cos s) = \sec s + \csc s$.

5. $\dfrac{1}{1-\sin w} + \dfrac{1}{1+\sin w} = 2\sec^2 w$.

6. $(\sin\theta + \cos\theta)^2 = 1 + 2\sin\theta\cos\theta$.

7. $\dfrac{1+\cos y}{1-\cos y} = (\csc y + \cot y)^2$.

8. $\sin^4 t - \cos^4 t = \sin^2 t - \cos^2 t$.

9. $\sin\left(\dfrac{3\pi}{2} + x\right) = -\cos x$.

10. $\cos\left(\dfrac{3\pi}{2} + x\right) = \sin x$.

11. $\cos(\pi - x) + \cos x = 0$.

12. $\csc(-x) + \sin x = \cos^2 x \csc(-x)$.

13. $\dfrac{\tan u - \cot u}{\sin u + \cos u} = \sec u - \csc u$.

In each one of Exercises 14-20, you are given a certain equality, which you are asked to show that is a false identity, so you need to find one value for the variable, for which the equality is not true. In each Exercise, you are also asked to find one value of the variable for which the equality does hold.

14. $(\sin x + \cos x)^3 = \sin^3 x + \cos^3 x$.

15. $(\tan t - \cot t)^2 = \tan^2 t - \cot^2 t$.

16. $\sin(-x) = \sin x$.

17. $\cos(-x) = -\cos x$.

18. $\sin(x + \pi) = \sin x$.

19. $\cos(x + \pi) = \cos x$.

20. $\sin(2x) = 2 \sin x$.

2.4 Graphs of Trigonometric Functions

In this section we take a closer look at the trigonometric functions by analyzing their **graphs**. Recall that *graphing a function f* means to *plot all points $P(x, y)$ in the coordinate plane, whose coordinates satisfy:*

$$y = f(x), \quad \text{with } x \text{ in the domain of } f.$$

The set of all points that are plotted according to the above "recipe" is what we call the **graph of the function**.

Graphs are very useful tools because they tell us quite a bit about certain *algebraic features of the function*, which can be characterized *geometrically*, as *features of the graph*. The table below summarizes all the features we are interested in, when analyzing functions and their graphs. (Certain items are detailed following the table.)

FEATURE OF f	ALGEBRAIC DESCRIPTION	GEOMETRIC DESCRIPTION
Domain of f	Set of all x, for which $f(x)$ *is defined*	All *vertical lines* that *intersect the graph*

"Forbidden" x-values	Values of x, for which $f(x)$ *is not defined*	Vertical lines that *do not intersect the graph*; in most cases, these lines are the *vertical asymptotes of the graph*
Range of f	Set of all values of $f(x)$	All *horizontal lines* that *intersect the graph*
Value at $x = 0$	Special value $y = f(0)$, if defined	The *y-intercept of the graph*
Zeros of f	Solutions of the equation $$f(x) = 0$$	The *x-intercepts of the graph*
Interval(s) where f is *increasing*	Interval(s) where $f(x_1) < f(x_2)$, whenever $x_1 < x_2$	Piece(s) of the graph that *climbs*, as we scan them from left to right
Interval(s) where f is *decreasing*	Interval(s) where $f(x_1) > f(x_2)$, whenever $x_1 < x_2$	Piece(s) of the graph that *descend*, as we scan them from left to right
Interval(s) where f is *constant*	Interval(s) where $f(x_1) = f(x_2)$, for any x_1, x_2	Piece(s) of the graph that are *horizontal*
f has a *local maximum value at* a	$f(a) \geq f(x)$, for *all* x *is some open interval containg* a; in this case, $f(a)$ is called a *local maximum value of* f	*"Peaks"* in the graph
f has a *local minimum value at* a	$f(a) \leq f(x)$, for *all* x *is some open interval containg* a; in this case, $f(a)$ is called a *local minimum value of* f	*"Valleys"* in the graph
f is *even*	$f(-x) = f(x)$, for all x in the domain	Graph is *symmetric with respect to the y-axis*
f is *odd*	$f(-x) = -f(x)$, for all x in the domain	Graph is *symmetric with respect to the origin*

Table 2.4.1

CLARIFICATIONS. The x-*intercepts* are, of course, the x-coordinates of the points where the graph intersects the x-axis. Likewise, the y-*intercept* is the y-coordinate of the point where the graph intersects the y-axis. To recover the domain of the function from the graph, we collect all vertical lines that intersect the graph, and then we collect the x-intercepts of these lines. Likewise, to recover the range of the function from the graph, we collect all horizontal lines that intersect the graph, and then we collect the y-intercepts of these lines.

Periodicity

This feature, which is one that is specific to trigonometric functions, is defined as follows.

A function f is said to be **periodic**, if there exists some number $P > 0$, such that

$$f(x+P) = f(x), \text{ for all } x \text{ in the domain of } f.$$

The *smallest positive number* P, that satisfies the above condition, is called the **period of** f.

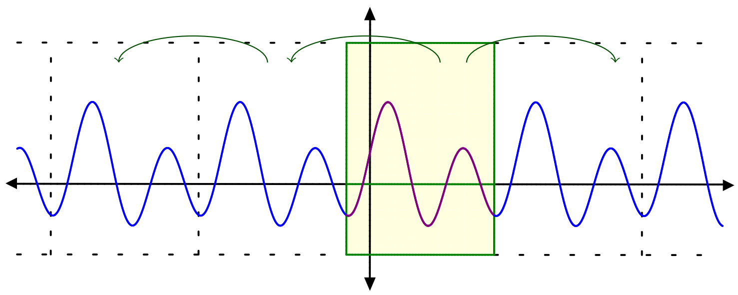


Figure 2.4.1

The graph of a periodic function exhibits a nice "tile" pattern, which is illustrated in the above picture. To graph such a function, all we have to do is to follow these steps:

I. Fix some interval of *length equal to the period*, which we will hereafter call a **fundamental interval**.
II. Graph the function over such an interval, thus creating a "master tile" for the graph.
III. Copy (and paste) the "master tile" repeatedly to the left and to the right, so the picture of the graph is made of "tiles," each looking exactly as the "master tile."

Every time we want to move from one tile to the next one on the right, we plot the same points, except that we replace the x-coordinate using $x \longmapsto x + \text{period}$. (When we move to the left we replace $x \longmapsto x - \text{period}$.)

Periodicity (if present) is very helpful, when we want to solve *elementary equations*, which are those of the form

$$f(x) = \textit{number}. \tag{2.4.1}$$

For such equations we will often apply the following method.

Elementary Equation Solving Method for Periodic Functions

If f is a periodic function, in order to solve an equation of the form (2.4.1), we do the following:

I. We fix a *fundamental interval*, and we first solve the equation in this interval; we call the solution(s) we obtain the *basic solution(s)*.
II. Once the basic solutions are found, *every solution* x of (2.4.1) is of the form:

$$x = \textit{basic solution} + \textit{integer multiple of period}. \tag{2.4.2}$$

CONVENTION: When treating trigonometric functions in this section, *we assume all rotation measures are in radians*.

Periodicity of Trigonometric Functions

(A) *The trigonometric functions* $\sin$, $\cos$, $\sec$ *and* $\csc$ *are all periodic, with period* 2π.
(B) *The trigonometric functions* $\tan$ *and* $\cot$ *are periodic, with period* π.

CLARIFICATION. Statement (A) was discussed in Section 2.2. Statement (B) was treated in the same Section, as part of the "Four Point Game."

The Graph of the Sine Function

Using the method outlined above, in order to graph the sine function, that is, to plot $y = \sin x$, all we need is a *fundamental interval*, and the "master tile." Although any interval of length 2π will do the job, we prefer to choose the fundamental interval to be: $[0, 2\pi)$. The graph of the sine function, together with the fundamental interval of our choice (shown in green) and the "master tile" (shown in purple), is depicted in the figure below.

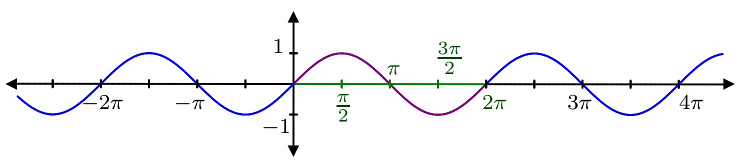


Figure 2.4.2

Features of the Sine Function

(i) Domain $=$ *all real numbers*.
(ii) Range $= [-1, 1]$
(iii) y-intercept $= 0$
(iv) x-intercepts $= n\pi$, n integer
(v) absolute maximum value $= 1$, at $x = \frac{\pi}{2} + 2n\pi$, n integer
(vi) absolute minimum value $= -1$, at $x = -\frac{\pi}{2} + 2n\pi$, n integer
(vii) increasing: on all intervals of the form $\left[-\frac{\pi}{2} + 2n\pi, \frac{\pi}{2} + 2n\pi\right]$, n integer
(viii) decreasing: on all intervals of the form $\left[\frac{\pi}{2} + 2n\pi, \frac{3\pi}{2} + 2n\pi\right]$, n integer
(ix) even/odd: sine function is *odd*
(x) period $= 2\pi$; preferred fundamental interval: $[0, 2\pi)$

CLARIFICATIONS. Most of these features can be *traced back in the "master tile," which is the piece of the graph over the fundamental interval* $[0, 2\pi)$.

For example, when looking for x-intercepts, which mean solutions of the elementary equation

$$\sin x = 0, \tag{2.4.3}$$

then by the method outlined above, all we have to do is to find the basic solutions, which are $x_1 = 0$ and $x_2 = \pi$, because all x-intercepts are presented as either

(a) of the form $0 + 2n\pi = 2n\pi$, with n integer, or
(b) of the form $\pi + 2n\pi = (2n + 1)\pi$, with n integer.

What we see in both (a) and (b) are two *lists of numbers*, which can be easily combined in one list, as indicated above: *integer multiples of* π.

Likewise, if we look for maximum values, which mean solutions of the elementary equation

$$\sin x = 1, \tag{2.4.4}$$

then the basic solution will be $\frac{\pi}{2}$, and then *all* solutions of (2.4.4) will be of the form "$x = \frac{\pi}{2}+$ integer multiple of 2π." Therefore, all peaks will be:

$$\text{peak} = \left(\frac{\pi}{2} + 2n\pi, 1\right), n \text{ integer}.$$

Same goes for valleys, which correspond to the solutions of

$$\sin x = -1. \tag{2.4.5}$$

The basic solution of this equation is $\frac{3\pi}{2}$, so *all* solutions of (2.4.5) will be of the form "$x = \frac{3\pi}{2}+$ integer multiple of 2π." It is convenient here to observe that $\frac{3\pi}{2} = -\frac{\pi}{2} + 2\pi$, and then the above description of all solutions of (2.4.5) will also be[8] "$x = -\frac{\pi}{2}+$ integer multiple of 2π." Therefore, all valleys will be:

$$\text{valley} = \left(-\frac{\pi}{2} + 2n\pi, -1\right), n \text{ integer}.$$

As for intervals, where sine is *increasing*, or *decreasing*, they will be of the form $\left[x_{\text{valley}}, x_{\text{peak}}\right]$, or $\left[x_{\text{peak}}, x_{\text{valley}}\right]$, in an alternating pattern. The peak that follows a valley is the closest on on the right, which is π units away; same goes for valleys that follow peaks, so our pattern is:

sine is increasing on intervals of the form: $\left[x_{\text{valley}}, x_{\text{valley}} + \pi\right]$
sine is decreasing on intervals of the form: $\left[x_{\text{peak}}, x_{\text{peak}} + \pi\right]$

Using what we have learned so far, we can now outline a method of solving the *elementary sine equations*, which are those of the form

$$\sin x = \textit{number}. \tag{2.4.6}$$

As it turns out, the method depends slightly on the value of the right-hand side. Based on the features of the sine function, we already know how to handle four cases, which we summarize as follows.

[8] This pretty clear, because "$\frac{3\pi}{2} +$ multiple of 2π" is same as "$-\frac{\pi}{2} +$ multiple of 2π.".

The "Easy" Sine Equations

(a) *If number is not in the interval* $[-1,1]$, *then the equation* (2.4.6) *has no solution*.

(b) *if number* $=-1$, *then all the solutions of* (2.4.6) *are of the form:* $x=-\dfrac{\pi}{2}+2n\pi$, n *integer*.

(c) *if number* $=1$, *then all the solutions of* (2.4.6) *are of the form:* $x=\dfrac{\pi}{2}+2n\pi$, n *integer*.

(d) *if number* $=0$, *then all the solutions of* (2.4.6) *are of the form:* $x=n\pi$, n *integer*.

The remaining case, which we will call the "hard" sine equation needs to be treated in a special way, in which the first step will be the solving of the *associated reference angle equation*

$$\sin(x^{\text{ref}}) = |number|, \quad x^{\text{ref}} \text{ in } \left[0,\frac{\pi}{2}\right]. \tag{2.4.7}$$

Note that the right-hand side of (2.4.7) uses the *absolute value*, and based on the given restriction, the equation has a *unique solution*. Once the reference angle equation is solved, the "hard" equation can be solved as follows.

The "Hard" Sine Equations

Assume $-1 < number < 1$ with $number \neq 0$, and x^{ref} is the unique solution of the associated reference angle equation (2.4.7).

I. The *original equation* (2.4.6) will always have *exactly two basic solutions*. Depending on the *sign of number*, these basic solutions are as follows.
   (a) *If number is positive, the two basic solutions are:* $x_1 = x^{\text{ref}}$ *and* $x_2=\pi-x^{\text{ref}}$.
   (b) *If number is negative, the two basic solutions are:* $x_1=\pi+x^{\text{ref}}$ *and* $x_2=2\pi-x^{\text{ref}}$.

II. Once the basic solutions are found, *every solution* x of (2.4.6) is of the form:

$$x = x_1 + 2n\pi,\ n \text{ integer, or}$$
$$x = x_2 + 2n\pi,\ n \text{ integer.}$$

CLARIFICATIONS. The fact that the "hard" sine equation has always *two basic solutions* follows from the "Four-Point Game" (see Section 2.2, especially Figure 2.2.5), by which we know that, given any α in $\left[0,\frac{\pi}{2}\right]$, the other rotation angles in $[0,2\pi]$ that have α as their reference angles are: $\pi-\alpha$, $\pi+\alpha$ and $2\pi-\alpha$.

For the "hard" sine equation, the two possibilities discussed above are illustrated in the two figures below.

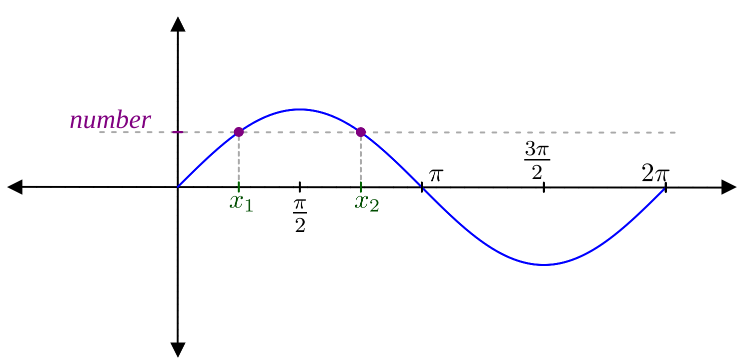


Figure 2.4.3

The figure above depicts case (a), when the right-hand side of (2.4.6) is positive. One of the two basic solutions lies in $\left[0, \frac{\pi}{2}\right]$ – quadrant I; the other one lies in $\left[\frac{\pi}{2}, \pi\right]$ – quadrant II.

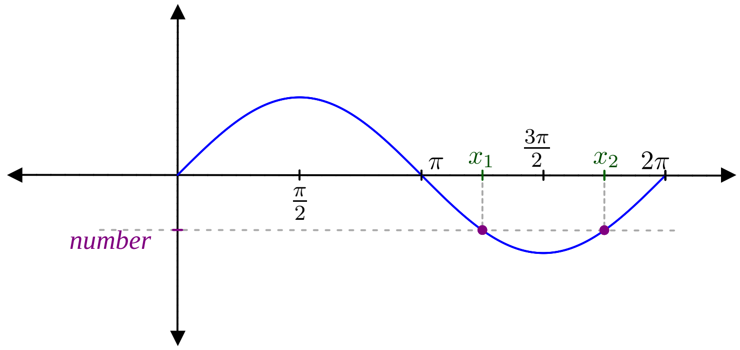


Figure 2.4.4

The figure above depicts case (b), when the right-hand side of (2.4.6) is negative. One of the two basic solutions lies in $\left[\pi, \frac{3\pi}{2}\right]$ – quadrant III; the other one lies in $\left[\frac{3\pi}{2}, 2\pi\right]$ – quadrant IV.

Example 2.4.1. Suppose we want to solve the equation

$$\sin x = \frac{\sqrt{3}}{2}. \tag{2.4.8}$$

The associated reference angle equation is

$$\sin(x^{\text{ref}}) = \frac{\sqrt{3}}{2}, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right].$$

which clearly (based on the "familiar" values of sine for acute angles), has the solution $x^{\text{ref}} = \dfrac{\pi}{3}$. Following the method outlined above the two steps go as follows.

I. Since the original (given) equation has positive right-hand side, the two basic solutions of (2.4.8) are

$$x_1 = x^{\text{ref}} = \frac{\pi}{3};$$
$$x_2 = \pi - x^{\text{ref}} = \pi - \frac{\pi}{3} = \frac{2\pi}{3}.$$

II. Using the above basic solutions, all solutions of (2.4.8) are:

$$x = \frac{\pi}{3} + 2n\pi,\ n \text{ integer};$$
$$x = \frac{2\pi}{3} + 2n\pi,\ n \text{ integer}.$$

Example 2.4.2. Suppose we want to solve the equation

$$\sin x = -\frac{1}{2}. \tag{2.4.9}$$

The associated reference angle equation is

$$\sin(x^{\text{ref}}) = \frac{1}{2}, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right].$$

which clearly (based on the "familiar" values of sine for acute angles), has the solution $x^{\text{ref}} = \frac{\pi}{6}$. Following the method outlined above the two steps go as follows.

I. Since the original (given) equation has negative right-hand side, the two basic solutions of (2.4.9) are

$$x_1 = \pi + x^{\text{ref}} = \pi + \frac{\pi}{6} = \frac{7\pi}{6};$$
$$x_2 = 2\pi - x^{\text{ref}} = 2\pi - \frac{\pi}{6} = \frac{11\pi}{6}.$$

II. Using the above basic solutions, all solutions of (2.4.9) are:

$$x = \frac{7\pi}{6} + 2n\pi,\ n \text{ integer};$$
$$x = \frac{11\pi}{6} + 2n\pi,\ n \text{ integer}.$$

The above technique for solving elementary sine equations is very inefficient. We will revisit it in Section 2.6.

There are cases, when we only want to solve an elementary sine equation in a specified interval. In these instances all we have to do is to *find the values of the integer n, for each list of solutions, which yield the corresponding general solution in the specified interval*.

Example 2.4.3. Suppose we want to solve the equation

$$\sin x = -0.6. \tag{2.4.10}$$

in the interval $[-\pi, 5\pi]$. As we see from the graph below, this problem has six solutions, listed in increasing order: x_1, x_2, x_3, x_4, x_5, x_6, where the basic solutions are x_3 and x_4.

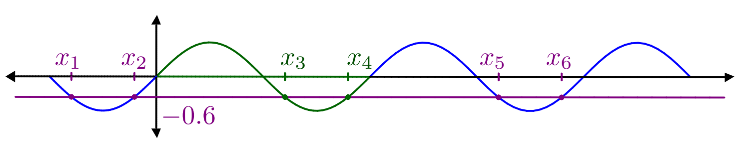


Figure 2.4.5

The associated reference angle equation is

$$\sin(x^{\text{ref}}) = 0.6, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right],$$

which can be solved on a calculator (using the $\boxed{\texttt{sin}^{-1}}$ function; do not forget to set the calculator to work in radians!): $x^{\text{ref}} \simeq 0.643501109$. Following the method outlined above the two steps go as follows.

I. Since the original (given) equation has negative right-hand side, the two basic solutions of (2.4.10) are

$$x_3 = \pi + x^{\text{ref}} \simeq \pi + 0.643501109 \simeq 3.785093762;$$
$$x_4 = 2\pi - x^{\text{ref}} \simeq 2\pi - 0.643501109 \simeq 5.639684198.$$

II. Using the above basic solutions, all solutions of (2.4.10) are:

$$x = x_3 + 2n\pi \simeq 3.785093762 + n \cdot 6.283185307,\ n \text{ integer}; \tag{2.4.11}$$
$$x = x_4 + 2n\pi \simeq 5.639684198 + n \cdot 6.283185307,\ n \text{ integer}. \tag{2.4.12}$$

With these calculations in mind, we need to see
(i) what values of n can be plugged in (2.4.11) to yield numbers in the interval $[-\pi, 5\pi] \simeq [-3.1441592954, 15.70796327]$, and likewise,
(ii) what values of n can be plugged in (2.4.12) to yield numbers in the same interval $[-\pi, 5\pi] \simeq [-3.1441592954, 15.70796327]$.
Equivalently, we need to find how the other four solutions x_1, x_2, x_5, x_6 are linked to the basic solutions x_3, x_4.

For question (i), it is clear that the only values that work in the list (2.4.11) are $n = -1$, $n = 0$ and $n = 1$. Equivalently, among our four additional solutions, the ones that are in the list (2.4.11) are:

$$x_1 = x_3 - 2\pi \simeq 3.785093762 - 6.283185307 \simeq -2.498091545,$$
$$x_5 = x_3 + 2\pi \simeq 3.785093762 + 6.283185307 \simeq 10.06827907.$$

Likewise, in question (ii) the only values that work in the list (2.4.12) are again $n = -1$, $n = 0$ and $n = 1$. Equivalently, among our four additional solutions, the ones that are in the list (2.4.12) are:

$$x_2 = x_4 - 2\pi \simeq 5.639684198 - 6.283185307 \simeq -0.643501109,$$
$$x_6 = x_4 + 2\pi \simeq 5.639684198 + 6.283185307 \simeq 11.92286951.$$

So our conclusion is that equation (2.4.10) has six solutions in the interval $[-\pi, 5\pi]$, which are (approximatively): $x_1 = -2.498091545$, $x_2 = -0.643501109$, $x_3 = 3.785093762$, $x_4 = 5.639684198$, $x_5 = 10.06827907$, and $x_6 = 11.92286951$.

The Graph of the Cosine Function

Graphing the cosine function is executed in the exact same manner as with sine. Our *fundamental interval* will again be: $[0, 2\pi)$. The graph of the cosine function, together with the fundamental interval of our choice (shown in green) and the "master tile" (shown in purple), is depicted in the figure below.

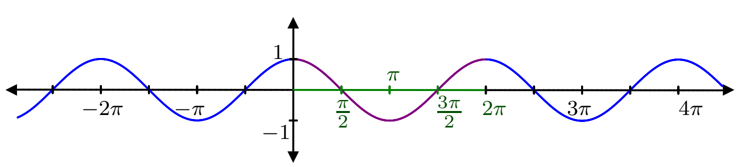


Figure 2.4.6

Features of the Cosine Function

(i) Domain = *all real numbers*.
(ii) Range = $[-1, 1]$
(iii) y-intercept = 1
(iv) x-intercepts = $\frac{\pi}{2} + n\pi$, n integer
(v) absolute maximum value = 1, at $x = 2n\pi$, n integer
(vi) absolute minimum value = -1, at $x = \pi + 2n\pi$, n integer
(vii) increasing: on all intervals of the form $\big[-\pi + 2n\pi, 2n\pi\big]$, n integer
(viii) decreasing: on all intervals of the form $\big[2n\pi, \pi + 2n\pi\big]$, n integer
(ix) even/odd: cosine function is *even*
(x) period = 2π; preferred fundamental interval: $[0, 2\pi)$

CLARIFICATIONS. The easiest way to see why the cosine function has the above features is to use the following well known fact from Algebra.

Horizontal Shift Rule for Graphs

If two functions f and g are related by the identity

$$g(x) = f(x - h),$$

then the graph of g is obtained from the graph of f by a horizontal shift of h units. In particular,

- *if h is positive, the horizontal shift will "push" the graph of f to the right;*
- *if h is negative, the horizontal shift will "push" the graph of f to the left;*

Since, using the Anti-Complement Formulas, we have

$$\cos x = \sin\left(x + \frac{\pi}{2}\right),$$

it follows that *the graph of the cosine function is obtained from the graph of the sine function, by "pushing" it to the left by $\frac{\pi}{2}$.*

Why do we "push" the graph of sine to the left to get the graph of cosine? When we work with $f(x) = \sin x$ and $g(x) = \cos x$, these two functions are related as in the Horizontal Shift Rule, with $h = -\frac{\pi}{2}$, so the shift that gets us from the graph of f to the graph of g will be indeed towards the left.

We can now also outline a method of solving the *elementary cosine equations*, which are those of the form

$$\cos x = \textit{number}. \tag{2.4.13}$$

As was the case with the sine equation, the method again depends slightly on the value of the right-hand side. Based on the features of the cosine function, we already know how to handle four cases, which we summarize as follows.

The "Easy" Cosine Equations

(a) *If number is not in the interval* $[-1, 1]$, *then the equation* (2.4.13) *has no solution*.
(b) *if number* $= -1$, *then all the solutions of* (2.4.13) *are of the form:* $x = \pi + 2n\pi$, n *integer*.
(c) *if number* $= 1$, *then all the solutions of* (2.4.13) *are of the form:* $x = 2n\pi$, n *integer*.
(d) *if number* $= 0$, *then all the solutions of* (2.4.13) *are of the form:* $x = \frac{\pi}{2} + n\pi$, n *integer*.

Exactly as was we did with the sine equation, the remaining case, which we will call the "hard" cosine equation needs to be treated in a special way, in which the first step will be the solving of the *associated reference angle equation*

$$\cos(x^{\text{ref}}) = |\textit{number}|, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right]. \tag{2.4.14}$$

As before, the right-hand side of (2.4.14) uses the *absolute value*, and based on the given restriction, the equation has a *unique solution*. Once the reference angle equation is solved, the "hard" equation can be solved as follows.

The "Hard" Cosine Equations

Assume $-1 < \textit{number} < 1$ with *number* $\neq 0$, and x^{ref} is the unique solution of the associated reference angle equation (2.4.14).

I. The *original equation* (2.4.13) will always have *exactly two basic solutions*. Depending on the *sign of number*, these basic solutions are as follows.
   (a) *If number is positive, the two basic solutions are:* $x_1 = x^{\text{ref}}$ *and* $x_2 = 2\pi - x^{\text{ref}}$.
   (b) *If number is negative, the two basic solutions are:* $x_1 = \pi - x^{\text{ref}}$ *and* $x_2 = \pi + x^{\text{ref}}$.

II. Once the basic solutions are found, *every solution* x of (2.4.13) is of the form:

$$x = x_1 + 2n\pi, \ n \text{ integer, or}$$
$$x = x_2 + 2n\pi, \ n \text{ integer.}$$

CLARIFICATIONS. As was the case with the sine equations, the fact that the "hard" equation has always *two basic solutions* follows from the "Four-Point Game.

For the "hard" cosine equation, the two possibilities discussed above are illustrated in the two figures below.

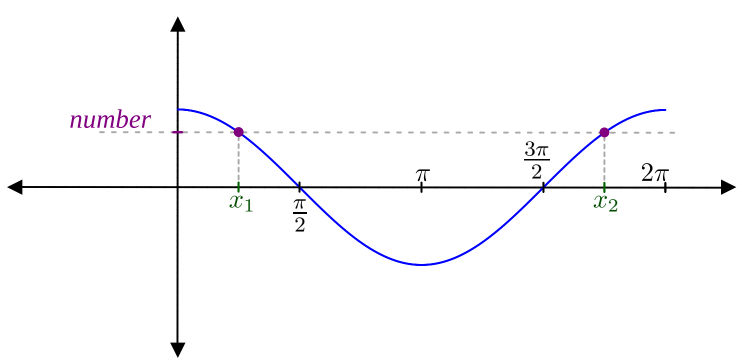


Figure 2.4.7

The figure above depicts case (a), when the right-hand side of (2.4.13) is positive. One of the two basic solutions lies in $\left[0, \frac{\pi}{2}\right]$ – quadrant I; the other one lies in $\left[\frac{3\pi}{2}, 2\pi\right]$ – quadrant IV.

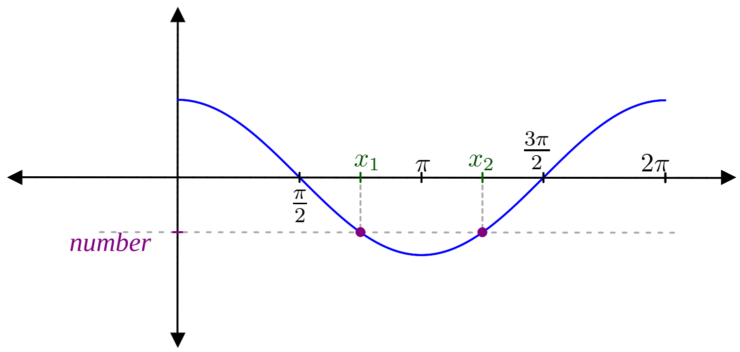


Figure 2.4.8

The figure above depicts case (b), when the right-hand side of (2.4.13) is negative. One of the two basic solutions lies in $\left[\frac{\pi}{2}, \pi\right]$ – quadrant II; the other one lies in $\left[\pi, \frac{3\pi}{2}\right]$ – quadrant III.

Example 2.4.4. Suppose we want to solve the equation

$$\cos x = \frac{\sqrt{2}}{2}. \tag{2.4.15}$$

The associated reference angle equation is

$$\cos(x^{\text{ref}}) = \frac{\sqrt{2}}{2}, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right].$$

which clearly (based on the "familiar" values of sine for acute angles), has the solution $x^{\text{ref}} = \frac{\pi}{4}$. Follows the method outlined above the two steps go as follows.

I. Since the original (given) equation has positive right-hand side, the two basic solutions of (2.4.15) are

$$x_1 = x^{\text{ref}} = \frac{\pi}{4};$$
$$x_2 = 2\pi - x^{\text{ref}} = 2\pi - \frac{\pi}{4} = \frac{7\pi}{4}.$$

II. Using the above basic solutions, all solutions of (2.4.15) are:

$$x = \frac{\pi}{4} + 2n\pi, \ n \text{ integer};$$
$$x = \frac{7\pi}{4} + 2n\pi, \ n \text{ integer}.$$

Example 2.4.5. Suppose we want to solve the equation

$$\cos x = -\frac{1}{2}. \tag{2.4.16}$$

The associated reference angle equation is

$$\cos(x^{\text{ref}}) = \frac{1}{2}, \quad x^{\text{ref}} \text{ in } \left[0, \frac{\pi}{2}\right].$$

which clearly (based on the "familiar" values of sine for acute angles), has the solution $x^{\text{ref}} = \frac{\pi}{3}$. Follows the method outlined above the two steps go as follows.

I. Since the original (given) equation has negative right-hand side, the two basic solutions of (2.4.16) are

$$x_1 = \pi - x^{\text{ref}} = \pi - \frac{\pi}{3} = \frac{2\pi}{3};$$
$$x_2 = \pi + x^{\text{ref}} = \pi + \frac{\pi}{3} = \frac{4\pi}{3}.$$

II. Using the above basic solutions, all solutions of (2.4.16) are:

$$x = \frac{2\pi}{3} + 2n\pi,\ n \text{ integer};$$
$$x = \frac{4\pi}{3} + 2n\pi,\ n \text{ integer}.$$

The above technique for solving elementary cosine equations is very inefficient. We will revisit it in Section 2.6.

The Graphs of the Tangent and Cotangent Functions

For the tangent function, which has period π, our preferred choice for the *fundamental interval* is $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The graph of the tangent, together with the fundamental interval of our choice (shown in green) and the "master tile" (shown in purple), is depicted in the figure below.

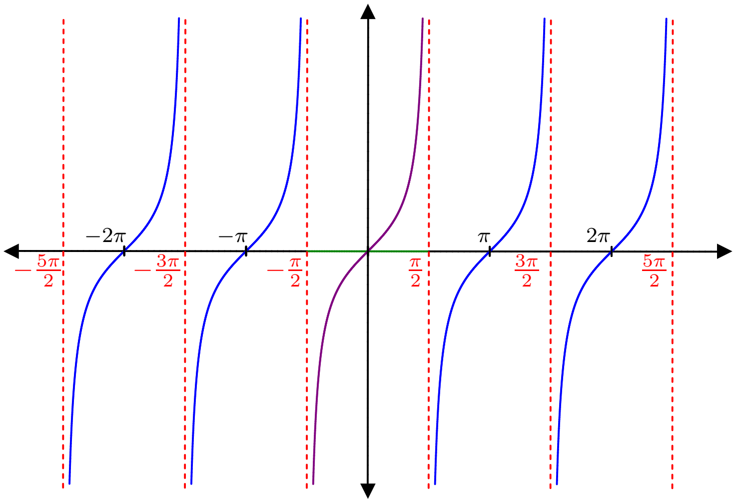


Figure 2.4.9

Features of the Tangent Function

(i) Domain $=$ *all real numbers, except for odd multiples of* $\frac{\pi}{2}$. Thus[9] the graph has *vertical asymptotes:*

$$x = \frac{\pi}{2} + n\pi, \; n \text{ integer}.$$

(ii) Range $=$ *all real numbers*
(iii) y-intercept $= 0$
(iv) x-intercepts $= n\pi$, n integer
(v) local maximum value(s) $=$ *none*
(vi) local minimum value(s) $=$ *none*
(vii) increasing: on all intervals of the form $\big(-\frac{\pi}{2} + n\pi, \frac{\pi}{2} + n\pi\big)$, n integer
(viii) decreasing: on *no interval*
(ix) even/odd: tangent function is *odd*
(x) period $= \pi$; preferred fundamental interval: $\big(-\frac{\pi}{2}, \frac{\pi}{2}\big)$

CLARIFICATIONS. Most of these features can be *traced back in the "master tile," which is the piece of the graph over the fundamental interval* $\big(-\frac{\pi}{2}, \frac{\pi}{2}\big)$. Since $\tan x = \dfrac{\sin x}{\cos x}$, it follows that

- *the zeros (x-intercepts) of tangent are the same as the zeros (x-intercepts) of sine;*
- *the vertical asymptotes for of tangent are at the same points as the zeros (x-intercepts) of cosine*.

The method for solving the *elementary tangent equations*, which are those of the form

$$\tan x = \textit{number}, \tag{2.4.17}$$

is similar to the one for sine or cosine equations, but is significantly simplified. We start, of course, with the *associated reference angle equation*

$$\tan(x^{\text{ref}}) = |\textit{number}|, \quad x^{\text{ref}} \text{ in } \Big[0, \frac{\pi}{2}\Big), \tag{2.4.18}$$

and once we solve (2.4.18), the method goes as follows.

The Tangent Equations

Assume x^{ref} is the unique solution of the associated reference angle equation (2.4.18).

I. The *original equation* (2.4.17) will always have *exactly one basic solution*. Depending on the *sign* of *number*, the unique basic solution is computed as follows.
   (a) *If number is positive, the basic solution is:* $x_0 = x^{\text{ref}}$.
   (b) *If number is negative, the basic solution is:* $x_0 = -x^{\text{ref}}$.

[9] An odd multiple of $\frac{\pi}{2}$ is a number of the form $(2n+1)\frac{\pi}{2} = \frac{\pi}{2} + n\pi$.

II. Once the basic solution is found, *every solution* x of (2.4.17) is of the form:

$$x = x_0 + n\pi,\ n \text{ integer}.$$

CLARIFICATIONS. The fact that the tangent equation has always *only one basic solution* follows from the formula for negatives, and the“Four-Point Game” (see Section 2.2, especially Figure 2.2.5), by which we know that, given any α in $\big[0, \frac{\pi}{2}\big)$, the only other rotation angle in $\big(-\frac{\pi}{2}, \frac{\pi}{2}\big)$ that has α as its reference angle is $-\alpha$.

Example 2.4.6. Suppose we want to solve the equation

$$\tan x = -\frac{1}{\sqrt{3}}. \tag{2.4.19}$$

The associated reference angle equation is

$$\tan(x^{\text{ref}}) = \frac{1}{\sqrt{3}}, \quad x^{\text{ref}} \text{ in } \big[0, \frac{\pi}{2}\big).$$

which clearly (based on the “familiar” values of tangent for acute angles), has the solution $x^{\text{ref}} = \frac{\pi}{6}$. Follows the method outlined above the two steps go as follows.

I. Since the original (given) equation has negative right-hand side, the basic solution of (2.4.19) is

$$x_0 = -x^{\text{ref}} = -\frac{\pi}{6}.$$

II. Using the above basic solution, all solutions of (2.4.19) are:

$$x = -\frac{\pi}{6} + n\pi,\ n \text{ integer}.$$

The above technique for solving elementary tangent equations will be improved in Section 2.6.

As for the graph of the cotangent function, we are going to use a technique similar to the one that allowed us to transform the graph of sine into the graph of cosine. However, transforming tangent into cotangent, in a way that allows us to keep it simple at the graph level, requires two steps, because, using the formulas for negative complements, we can write

$$\cot x = -\tan\big(x - \frac{\pi}{2}\big),$$

which means that, when we define the auxiliary function $f(x) = \tan\big(x - \frac{\pi}{2}\big)$, the following statements hold true.

1. The graph of f is obtained from the graph of tangent by a *right translation by* $\frac{\pi}{2}$.
2. The graph of $\cot x = -f(x)$ is obtained from the graph of f by *flipping about the x-axis*.

Combining these two observations, we see that *the graph of cotangent is obtained from the graph of tangent by pushing first to the right by* $\frac{\pi}{2}$, *then flipping about the x-axis*. For this reason, the cotangent function has its graph as depicted in the figure below.

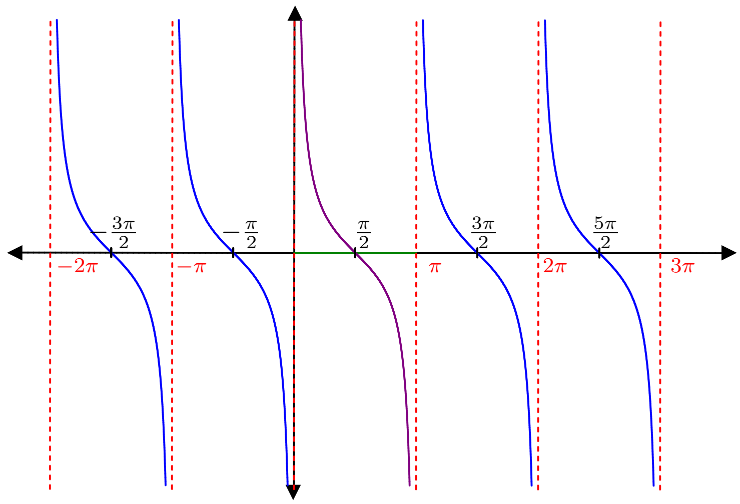


Figure 2.4.10

Our preferred choice for the *fundamental interval* is now $(0, \pi)$. Figure 2.4.10 above shows the graph of the cotangent, together with the fundamental interval of our choice (shown in green) and the “master tile” (shown in purple).

Features of the Cotangent Function

(i) Domain $=$ *all real numbers, except for integer multiples of* π. Thus the graph has *vertical asymptotes:*

$$x = n\pi,\ n \text{ integer}.$$

(ii) Range $=$ *all real numbers*
(iii) y-intercept $=$ none
(iv) x-intercepts $= \dfrac{\pi}{2} + n\pi$, n integer
(v) local maximum value(s) $=$ *none*
(vi) local minimum value(s) $=$ *none*
(vii) increasing: on *no interval*
(viii) decreasing: on all intervals of the form $(n\pi, (n+1)\pi)$, n integer
(ix) even/odd: tangent function is *odd*
(x) period $= \pi$; preferred fundamental interval: $(0, \pi)$

CLARIFICATIONS. The fundamental interval $(0, \pi)$ for cotangent is obtained by pushing the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ – the fundamental interval for tangent – to the right by $\frac{\pi}{2}$. Since $\cot x = \dfrac{\cos x}{\sin x}$, it follows that

- *the zeros (x-intercepts) of cotangent are the same as the zeros (x-intercepts) of cosine;*
- *the vertical asymptotes for of cotangent are at the same points as the zeros (x-intercepts) of sine.*

Concerning the *elementary cotangent equations*

$$\cot x = \textit{number}, \tag{2.4.20}$$

except for the case *number* $= 0$, which corresponds to the x-intercepts, all other cases reduce to the tangent equation

$$\tan x = \frac{1}{\textit{number}},$$

so no additional analysis is needed. In either case, the general solution of (2.4.20) will always be of the form

$$x = x_0 + n\pi, \; n \text{ integer}.$$

The Graphs of the Secant and Cosecant Functions

As the graphs of secant and cosecant are quite complicated, they are seldom used. These graphs, depicted in Figure 2.4.11 and Figure 2.4.12 are only included here for the sake of completeness.

For the secant function, which has period 2π, our preferred choice for the *fundamental "interval"* is $\left(-\frac{\pi}{2}, \frac{3\pi}{2}\right)$, with $\frac{\pi}{2}$ removed. One way to describe such sets is to call them *punctured intervals*. The graph of the secant, with the fundamental punctured interval of our choice (shown in green) and the "master tile" (shown in purple), is depicted in the figure 2.4.11 below.

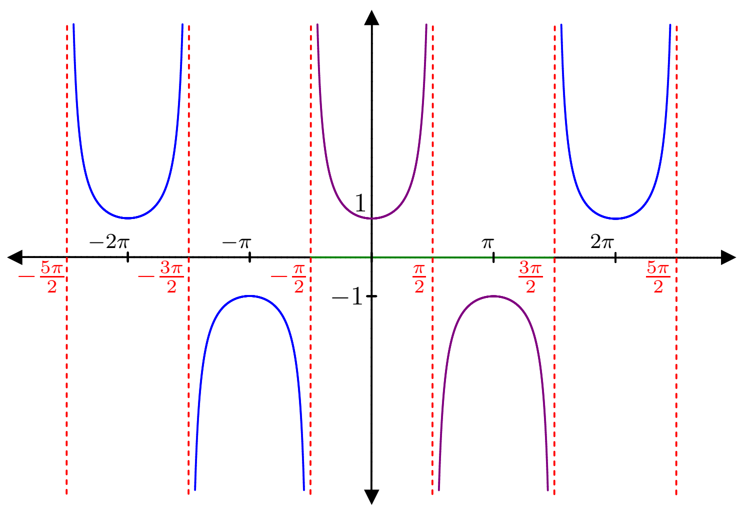


Figure 2.4.11

Features of the Secant Function

(i) Domain = *all real numbers, except for odd multiples of* $\frac{\pi}{2}$. Thus the graph has *vertical asymptotes:*

$$x = \frac{\pi}{2} + n\pi, \; n \text{ integer}.$$

(ii) Range $= (-\infty, -1] \cup [1, \infty)$, *that is, real numbers, either* ≤ -1, *or* ≥ 1

(iii) y-intercept $= 1$
(iv) x-intercepts $=$ none
(v) local maximum value(s) $= -1$, at $x = \pi + 2n\pi$, n integer
(vi) local minimum value(s) $= 1$, at $x = 2n\pi$, n integer
(vii) increasing: on all intervals of the form $\left[2n\pi, \frac{\pi}{2} + 2n\pi\right)$, n integer, and on all intervals of the form $\left(\frac{\pi}{2} + 2n\pi, \pi + 2n\pi\right]$, n integer
(viii) decreasing: on on all intervals of the form $\left[\pi + 2n\pi, \frac{3\pi}{2} + 2n\pi\right)$, n integer, and on all intervals of the form $\left(-\frac{\pi}{2} + 2n\pi, 2n\pi\right]$, n integer
(ix) even/odd: secant function is *even*
(x) period $= 2\pi$; preferred fundamental punctured interval: $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$

CLARIFICATIONS. Since $\sec x = \dfrac{1}{\cos x}$, it follows that: *the vertical asymptotes for of secant are at the same points as the zeros (x-intercepts) of cosine.*

It is not necessary to discuss the elementary secant equations, because (as was the case with cotangent), using reciprocals, they can be reduced to cosine equations.

As for the graph and features of the cosecant function, we can again use graph transformations that allow us to relate cosecant to secant. Indeed, using the negative complement formulas, we know that

$$\sec x = \csc\left(x - \frac{\pi}{2}\right).$$

What this means is that *the graph of cosecant is obtained from the graph of secant by pushing it to the right by* $\frac{\pi}{2}$. Our preferred choice for the *fundamental punctured interval* for cosecant will therefore be $(0, 2\pi)$, with π removed. The graph of the cosecant, with the fundamental punctured interval of our choice (shown in green) and the "master tile" (shown in purple), is depicted in the figure below.

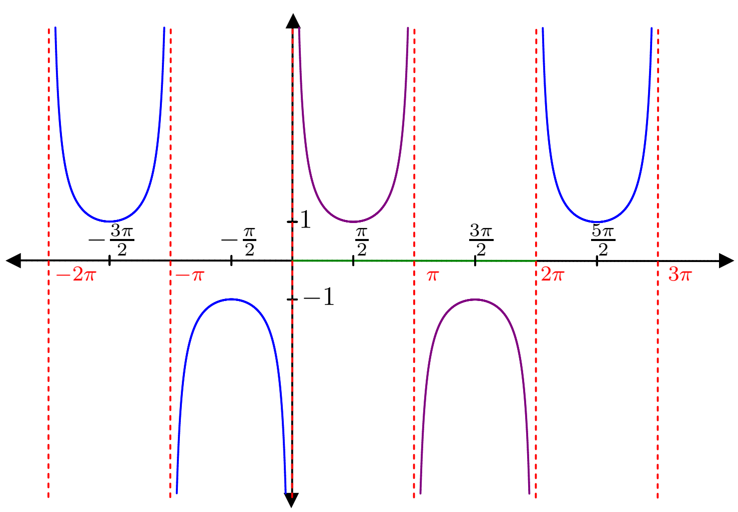


Figure 2.4.12

Features of the Cosecant Function

(i) Domain = *all real numbers, except for multiples of* π. Thus the graph has *vertical asymptotes:*

$$x = n\pi, \; n \text{ integer}.$$

(ii) Range = $(-\infty, -1] \cup [1, \infty)$, *that is, real numbers, either* ≤ -1*, or* ≥ 1

(iii) y-intercept = none

(iv) x-intercepts = none

(v) local maximum value(s) = -1, at $x = -\dfrac{\pi}{2} + 2n\pi$, n integer

(vi) local minimum value(s) = 1, at $x = \dfrac{\pi}{2} + 2n\pi$, n integer

(vii) increasing: on all intervals of the form $\left[\dfrac{\pi}{2} + 2n\pi, \pi + 2n\pi\right)$, n integer, and on all intervals of the form $\left(\pi + 2n\pi, \dfrac{3\pi}{2} + 2n\pi\right]$, n integer

(viii) decreasing: on on all intervals of the form $\left[-\dfrac{\pi}{2} + 2n\pi, 2n\pi\right)$, n integer, and on all intervals of the form $\left(2n\pi, \dfrac{\pi}{2} + 2n\pi\right]$, n integer

(ix) even/odd: cosecant function is *odd*

(x) period = 2π; preferred fundamental punctured interval: $(0, \pi) \cup (\pi, 2\pi)$

CLARIFICATIONS. Since $\csc x = \dfrac{1}{\sin x}$, it follows that: *the vertical asymptotes for of cosecant are at the same points as the zeros (x-intercepts) of sine.*

Exercises

In Exercises 1–9 you are asked to list all features (Period, Even/Odd/Neither, Domain, Range, Intercepts, Maximum and Minimum Values and where they are attained, as well as Intervals where the function is Increasing/Decreasing) for the given function , and then to sketch the graph.

1. $f(x) = 1 + \sin x$

2. $f(x) = 1 - \sin x$

3. $f(x) = -\dfrac{1}{2} + \sin x$

4. $f(x) = -\sin x$

5. $f(x) = 3 + \sin x$

6. $f(x) = 1 + \cos x$

7. $f(x) = \dfrac{\sqrt{2}}{2} + \cos x$

8. $f(x) = \dfrac{\sqrt{3}}{2} - \cos x$

9. $f(x) = 5 + \cos x$

In Exercises 10–14 you are asked to list all features (Domain, Range, Intercepts, Vertical Asymptotes, Intervals where the function is increasing/decreasing, and Periodicity) for the given function , and then to sketch the graph.

10. $f(x) = 1 + \tan x$

11. $f(x) = 1 - \tan x$

12. $f(x) = \tan\left(x + \frac{\pi}{4}\right)$

13. $f(x) = 1 - \tan\left(x + \frac{3\pi}{4}\right)$

14. $f(x) = 1 + \cot x$

In Exercises 15–27 you are asked to solve a trigonometric equation, either by finding all solutions, or only those in a specified interval.

15. Find *all solutions* of: $\sin x = \frac{\sqrt{2}}{2}$. Use exact values.

16. Find *all solutions* of: $\sin x = -\frac{\sqrt{2}}{2}$. Use exact values.

17. Find *all solutions* of: $\sin x = -2$. Use exact values.

18. Given the equation $\sin x = -0.2$, find only the solutions that are in the interval $[-\pi, 5\pi]$. Round to nearest 0.01.

19. Given the equation $\sin x = -\frac{1}{2}$, find only the solutions that are in the interval $[-3\pi, 3\pi]$. Use exact values.

20. Find *all solutions* of: $\cos x = -\frac{\sqrt{2}}{2}$. Use exact values.

21. Given the equation $\cos x = \frac{\sqrt{3}}{2}$, find only the solutions that are in the interval $[3\pi, 9\pi]$. Use exact values.

22. Given the equation $\cos x = -\frac{\sqrt{3}}{2}$, find only the solutions that are in the interval $[-\pi, 4\pi]$. Use exact values.

23. Find *all solutions* of: $\cos x = -3$. Use exact values.

24. Given the equation $\cos x = 0.7$, find only the solutions that are in the interval $[-2\pi, 4\pi]$. Round to nearest 0.01.

25. Find *all solutions* of: $\tan x = -\sqrt{3}$. Use exact values.

26. Given the equation $\tan x = -1$, find only the solutions that are in the interval $[-\pi, 6\pi]$. Use exact values.

27. Given the equation $\tan x = -7$, find only the solutions that are in the interval $[\pi, 6\pi]$. Round to nearest 0.01.

2.5 General Sinusoidal Functions and Their Graphs

In this section we examine an important type of periodic functions, which based on what we learned so far, can be easily studied and graphed.

Shrinking/Stretching Graphs

Besides graph transformations that involve *shifts*, we also have a good idea about how graphs change when we *scale the variables*. In vector coordinates, the transformations we are using are those given by

$$\mathbf{\Sigma}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} px \\ qy \end{bmatrix} = \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

where $p > 0$ and $q > 0$ are two (fixed) real numbers, which we call the x- and y-stretch factors. Such transformations are called **stretch transformations**.

If we think the plane as a big sheet of rubber pinned at the origin, such a transformation *stretches by a factor of p in the horizontal direction, and by a factor of q in the vertical direction*. For example, if we we perform a stretch transformation with stretch factors $p = 1.5$ and $q = 0.8$, then the unit circle will be transformed into a "football-shaped" curve, which is what we call an *ellipse*. (More on these in Chapter 5.)

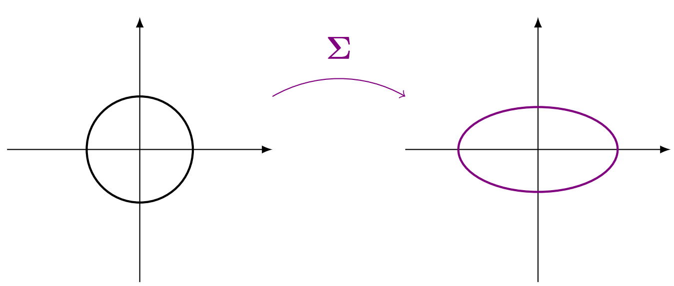


Figure 2.5.1

Of course, a "true stretch" take place when the factor is greater than 1. If a factor is less than 1, a shrinkage will take place. For example, in the case illustrated in Figure 2.5.1 the transformation $\mathbf{\Sigma}$ streches in the horizontal direction and shrinks in the vertical direction.

When graphs of functions are involved, the following easy rule applies.

Graph Scaling Rule

Given two functions f and g, which are linked by:

$$f(x) = ag(bx),$$

the graph of f is obtained from the graph of g by applying a stretch transformation with

(A) *x-stretch factor $p = \dfrac{1}{b}$, and*

(B) *y-stretch factor $q = a$.*

The Standard Sine-Wave Functions

Of special interest to us are the **standard sine-wave functions**, which are those of the form

$$\sigma(x) = a\sin(bx); \qquad a, b \textit{ positive}. \tag{2.5.1}$$

The number a is called the **amplitude of σ**. The number b is called the **frequency of σ**. Using the Graph Scaling Rule, such functions have their graphs looking very much like the graph of sine, except that they are stretched in both directions by the two stretch factors described in (A) and (B) above. In particular, it follows that *σ is periodic, with*

$$\textit{Period} = \frac{2\pi}{b}.$$

The functions σ behave very much like the sine function, except that their features are slightly modified to take the stretchings into account.

Features of Standard Sine-Wave Functions

If the function σ is given by (2.5.1), then it has the following features.

(i) Domain of σ = *all real numbers*.

(ii) Range of $\sigma = [-a, a]$

(iii) y-intercept $= 0$

(iv) x-intercepts $= \dfrac{n\pi}{b}$, n integer

(v) absolute maximum value $= a$, at $x = \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}$, n integer

(vi) absolute minimum value $= -a$, at $x = -\dfrac{\pi}{2b} + \dfrac{2n\pi}{b}$, n integer

(vii) increasing: on all intervals of the form $\left[-\dfrac{\pi}{2b} + \dfrac{2n\pi}{b}, \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}\right]$, n integer

(viii) decreasing: on all intervals of the form $\left[\dfrac{\pi}{2b} + \dfrac{2n\pi}{b}, \dfrac{3\pi}{2b} + \dfrac{2n\pi}{b}\right]$, n integer

(ix) even/odd: the function σ is *odd*

(x) period $= \dfrac{2\pi}{b}$; preferred fundamental interval: $\left[0, \dfrac{2\pi}{b}\right)$

Example 2.5.1. Consider the function $\sigma(x) = 3\sin 2x$.

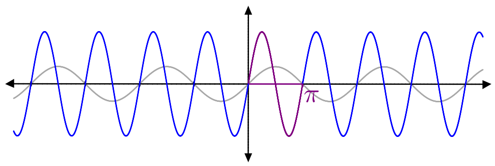


Figure 2.5.2

When compared to the graph of the usual sine function (shown in gray), the graph of σ is *shrunk horizontally in half*, then *stretched vertically by a factor of* 3. The function σ is periodic with *period* $= \dfrac{2\pi}{2} = \pi$, so when we graph it, the "master tile" will be over the fundamental interval: $[0, \pi)$.

Shifted Sine-Wave Functions

Our knowledge of standard sine-wave can also be applied to their *shifted* versions, which have the form

$$f(x) = a\sin\big(b(x-\phi)\big) = \sigma(x-\phi), \tag{2.5.2}$$

where σ is a standard sine-wave function of the form (2.5.1), and ϕ is some other (fixed) number, which is called the **phase shift of** f. The graph of such a function is obtained from the *graph of* σ*, by applying a horizontal shift:*

- *if* ϕ *is positive, the shift is towards the right (by* ϕ *units);*
- *if* ϕ *is negative, the shift is towards the right (by* $-\phi$ *units).*

Of course, the functions we are dealing with here are precisely those that can be presented in the form:

$$f(x) = a\sin(bx+c); \quad a,\, b \text{ positive}. \tag{2.5.3}$$

Features of Shifted Standard Sine-Wave Functions

If a function f is presented as

$$f(x) = a\sin(bx+c), \tag{2.5.4}$$

with a and b *positive*, then it can also be presented as a *shifted standard sine-wave equation* (2.5.2), with phase shift:

$$\phi = -\frac{c}{b}. \tag{2.5.5}$$

Additionally, the function f has:

(i) Domain of f = *all real numbers*.
(ii) Range of $f = [-a, a]$
(iii) y-intercept $= \sin c$
(iv) x-intercepts $= \phi + \dfrac{n\pi}{b}$, n integer
(v) absolute maximum value $= a$, at $x = \phi + \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}$, n integer
(vi) absolute minimum value $= -a$, at $x = \phi - \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}$, n integer
(vii) increasing: on all intervals of the form $[\phi - \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}, \phi + \dfrac{\pi}{2b} + \dfrac{2n\pi}{b}]$, n integer

(viii) decreasing: on all intervals of the form $\left[\phi+\frac{\pi}{2b}+\frac{2n\pi}{b}, \phi+\frac{3\pi}{2b}+\frac{2n\pi}{b}\right]$, n integer

(ix) period $=\frac{2\pi}{b}$; preferred fundamental interval: $\left[\phi, \phi+\text{period}\right)=\left[\phi, \phi+\frac{2\pi}{b}\right)$

Example 2.5.2. Consider the function $f(x)=3\sin\left(2x+\frac{\pi}{3}\right)$.

The amplitude of f is: 3. The period of f is: $\frac{2\pi}{2}=\pi$, and the phase shift will be

$$\phi=-\frac{\frac{\pi}{3}}{2}=-\frac{\pi}{6},$$

so the fundamental interval will have

$$\text{start}=\phi=-\frac{\pi}{6};$$
$$\text{end}=\phi+\text{Period}=-\frac{\pi}{6}+\pi=\frac{5\pi}{6}.$$

The graph of f is then obtained from the graph of the standard sine-wave function $\sigma(x)=3\sin 2x$ – same as the one from Example 2.5.1 – by *shifting it to the left by* $-\frac{\pi}{6}$.

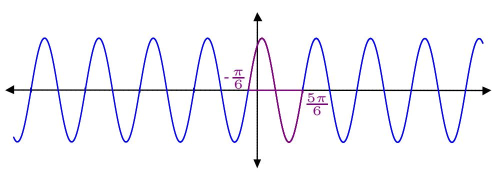


Figure 2.5.3

As it turns out, many functions can be re-written as shifted standard sine-wave functions. Basically, all functions of the form $f(x)=m\sin(nx+p)$ or $f(x)=m\cos(nx+p)$, can be re-written as in (2.5.3), using one of the formulas for negatives, the (anti-)complement formulas, or the add π formulas.

Example 2.5.3. Consider the function $f(x)=-3\cos 2x$.

As it turns out, using the formulas for anti-complements, that is, $\cos\alpha=\sin\left(\alpha+\frac{\pi}{2}\right)$, we can rewrite

$$f(x)=-3\cos 2x=-3\sin\left(2x+\frac{\pi}{2}\right).$$

Since adding/subtracting π changes signs in the $\sin$ function, that is, $\sin(\alpha+\pi)=\sin(\alpha-\pi)=-\sin\alpha$, we can re-write our function as:

$$f(x)=-3\sin\left(2x+\frac{\pi}{2}\right)=3\sin\left(2x+\frac{\pi}{2}-\pi\right)=3\sin\left(2x-\frac{\pi}{2}\right).$$

So when we match f with a *shifted standard sine-wave function*, we get:

- amplitude $=3$;
- period $=\frac{2\pi}{2}=\pi$;

- phase shift:

$$\phi = -\frac{-\frac{\pi}{2}}{2} = \frac{\pi}{4};$$

- fundamental interval with

$$\text{start} = \phi = \frac{\pi}{4};$$
$$\text{start} = \phi + \text{Period} = \frac{\pi}{4} + \pi = \frac{5\pi}{4}.$$

The graph of f is then obtained from the graph of the standard sine-wave function $\sigma(x) = 3\sin 2x$ – same as the one from Example 2.5.1 – by *shifting it to the right by* $\frac{\pi}{4}$.

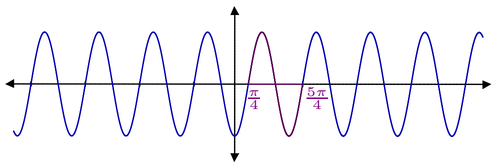


Figure 2.5.4

Finding the Equations of Sinusoidal Graphs

The problem we are concerned with here is the following: *Given a graph (known to be the graph of a shifted sine-wave function), find its equation*. The preferred form of the equation we seek is:

$$y = a\sin(bx + c),$$

with both a and b *positive*, so in effect the right-hand side is a *shifted standard sine-wave function*. Of course, what we are doing is nothing else but tracing back the steps used in graphing such functions, which means that we use the following three-step method.

I. *Find a fundamental interval, that is, an interval on the x-axis that "is responsible" for complete sine-wave; once such and interval is found, with its start and end, we set:*
   - *phase shift: ϕ = start;*
   - *period = end − start, or equivalently, period = length of the fundamental interval.*

II. *Compute the coefficients: $b = \dfrac{2\pi}{\text{period}}$ and $c = -\phi \cdot b$.*

III *Set a = amplitude (by identifying the maximum/minimum y-coordinates on the graph).*

Example 2.5.4. Consider the graph shown in the picture below, in which all tick marks are spaced 1 unit appart.

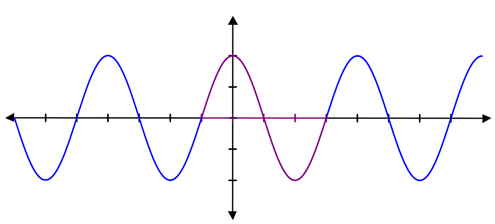


Figure 2.5.5

To find the equation that represents this graph, we follow the three steps outlined above.

I. One complete sine-wave (shown in purple) sits over the interval from -1 to 3, so one choice[10] our fundamental interval is $[-1, 3)$. Based on this choice, our function will have phase shift $\phi = -1$, and period $= 4$.

II. Using the preceding identifications, we have

$$b = \frac{2\pi}{\text{period}} = \frac{2\pi}{4} = \frac{\pi}{2};$$
$$c = -\phi \cdot b = -(-1) \cdot \frac{\pi}{2} = \frac{\pi}{2}.$$

III. As the y-coordinate of the highest points of the graph is 2, we can set $a = \text{amplitude} = 2$.

With all these calculations, we now conclude that the graph from Figure 2.5.5 is simply given by the equation

$$y = 2\sin\left(\frac{\pi}{2}x + \frac{\pi}{2}\right).$$

Exercises In Exercises 1–9 you are asked to do *all* of the following:

(a) If necessary, rewrite the given function in the form $f(x) = a\sin(bx + c)$ with a and b positive.

(b) Find the amplitude, period and phase shift.

(c) Sketch the graph and highlight the "master tile."

1. $f(x) = 4\sin 3x$
2. $f(x) = -4\sin\left(2x + \frac{\pi}{4}\right)$
3. $f(x) = 2\sin\left(3x - \frac{\pi}{3}\right)$
4. $f(x) = -2\sin\left(2\pi x + \frac{\pi}{2}\right)$
5. $f(x) = -4\sin\left(\frac{\pi}{3} - 3x\right)$
6. $f(x) = 5\sin\left(\frac{\pi}{4} - 10x\right)$
7. $f(x) = -4\cos 4x$
8. $f(x) = 2\cos\left(\pi x + \frac{\pi}{3}\right)$
9. $f(x) = -3\cos\left(4x - \frac{\pi}{5}\right)$
10. $f(x) = -2\cos\left(2x + \frac{\pi}{4}\right)$

[10] Picking a fundamental interval is a matter of choice; other valid choices are for example $[-5, -1)$, or $[3, 7)$.

11. $f(x) = -6\cos\left(\frac{\pi}{3} - \frac{2x}{3}\right)$

NOTE: A good supply of problems in which you are asked to find the equation, given the graph (similar to Example 2.5.4), is provided in the K-STATE ONLINE HOMEWORK SYSTEM.

2.6 The Inverse Trigonometric Functions

In this section we introduce certain functions that ultimately will allow us to solve all elementary trigonometric equations, of the form

$$\text{function}(x) = \textit{number}, \tag{2.6.1}$$

where function is one of sin, cos, tan. The main problem with equations of the form (2.6.1) is that they have *many (in fact infinitely many) solutions*, so the main idea in defining our inverse functions will be to build a "recipe" for picking up *one particular solution*.

Example 2.6.1. We have all become familiar with such a construction in Algebra, when we learned about the *square root operation*. The way we constructed the square root was by defining $\sqrt{\textit{number}}$ to be the solution of the equation $?^2 = \textit{number}$, which is ≥ 0. For example, when we want to compute $\sqrt{81}$, we look at the equation $?^2 = 81$, which has *two* solutions ± 9, and we pick the one that is not negative, that is, 9, and this particular solution is what we define $\sqrt{81}$ to be.

Inverse Functions

The best way to understand Example 2.6.1, in a way that is applicable in Trigonometry, is to use *function* terminology.

Given a function f, and some set D contained in the *domain of* f, we say that f is **one-to-one on** D, if:

($\star$) *whenever x_1 and x_2 are two distinct elements in D (that is, $x_1 \neq x_2$), it follows that $f(x_1) \neq f(x_2)$.*

Example 2.6.2. If we consider the function $f(x) = x^2$ (with domain consisting of all real numbers), then f is not 1-1 on its entire domain, because for instance $-1 \neq 1$, but $f(-1) = f(1)$. One way to "fix" this problem is then to consider the set $D = [0, \infty)$. When we restrict to this set, the desired property ($\star$) will hold, so we can safely declare that *f is one-to-one on $[0, \infty)$*.

The Restricted Inverse Function Construction

Given a function f, which *is one-to-one on D*, we can define a function g as follows:

(A) We let the *domain of g* to be *the set of all possible values of $f(x)$, as x runs through D*.

(B) For any *number* in the *domain of g* (as defined above), we define the quantity $g(\textit{number}) = ?$ to be the *unique element ? in D* that satisfies

$$f(?) = \textit{number}. \tag{2.6.2}$$

The function g defined this way, is called the **restricted inverse (function) of f relative to D**. In addition to (A) and (B) above, this function also has:
(C) *range of* $g = D$.
In the case when *f is one-to-one, on its entire domain*, the above function g is called the **inverse (function) of f**, and is denoted by f^{-1}. In this instance,
- *domain of* f^{-1} = *range of* f;
- *range of* f^{-1} = *domain of* f.

Example 2.6.3. If we consider an *exponential function*, that is, one of the form $f(x) = a^x$, with $a \neq 1$ some positive constant, then f is one-to-one on its entire domain (the set of all real numbers). Therefore f has an inverse function associated with it, which is nothing else but the *logarithimc function*: $f^{-1}(x) = \log_a x$, $x > 0$. (The domain of f^{-1} is the range of the exponential, which we know to be equal to $(0, \infty)$.)

Example 2.6.4. If we consider the function $f(x) = x^2$ from Example 2.6.2, then its restricted inverse relative to $[0, \infty)$ is precisely the square root function $g(x) = \sqrt{x}$, $x \geq 0$.

CLARIFICATION. Whenever restricted inverses of functions are constructed, they immediately yield an important pair of formulas, as follows.

Inversion Identities

If the function f is one-to-one on D and g is the restricted inverse of f, relative to D, then:
I. $f\big(g(x)\big) = x$, *for every x in the domain of g;*
II. $g\big(f(x)\big) = x$, *for every x in D.*

The next issue we can consider is *graphing* (restricted) inverse functions. The most effective way to handle this problem is to proceed as follows.

Graph Inversion

Suppose the function f (whose graph is given) is one-to-one on D and g is the *restricted inverse of f, relative to D*. In order to graph g, we follow these steps.

1. We plot/mark only the points on the graph of f that have x-coordinates in D. In other words we only plot the points that satisfy
$$y = f(x), \quad x \text{ in } D. \tag{2.6.3}$$
(If D is the entire domain of f, then (2.6.3) is the entire graph of f.)
2. The graph of g is obtained by *reflecting the points plotted/marked in step 1 about the diagonal line $y = x$*. In effect, this new plot is given by
$$x = f(y), \quad y \text{ in } D. \tag{2.6.4}$$

Example 2.6.5. Consider the set-up from Example 2.6.4, with $f(x) = x^2$ and $D = [0, \infty)$ (shown on the x-axis in green).

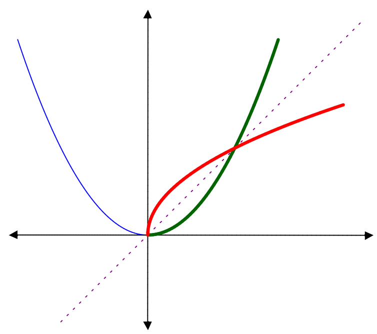


Figure 2.6.1

The entire graph of f is shown in blue, while the plot obtained by step 1 is highlighted (drawn thicker) in green. When we reflect (flip) this set about the diagonal (shown dashed in purple), we obtain exactly the graph of the restricted inverse function $g(x) = \sqrt{x}$, $x \geq 0$, which can also be presented as:

$$x = y^2, \quad y \geq 0.$$

The Arcsine and Arccosine Functions

We wish now to apply the construction of a "reasonable" restricted inverse for the sine function $f(x) = \sin x$. We first need to identify a set D, on which *sine is one-to-one*.

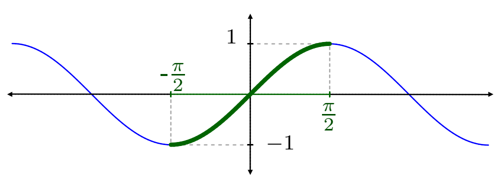


Figure 2.6.2

Based what we see in the above picture, a suitable choice of such a set is: $D = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. After all (see Section 2.4), we also know that *sine is increasing on* $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, and this clearly guarantees one-to-one-ness. With this choice in mind, we can now define the following function.

The **arcsine** function, denoted by $\arcsin$, is the *restricted inverse of the sine function, relative to* $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The *domain of* $\arcsin$ is the interval $[-1, 1]$. Therefore:

$$\arcsin \textit{number} = \, ? \text{ means: } \begin{cases} \sin \, ? = \textit{number} \\ -1 \leq \textit{number} \leq 1 \\ -\dfrac{\pi}{2} \leq \, ? \leq \dfrac{\pi}{2} \end{cases} \tag{2.6.5}$$

Using what we learned about graphing (restricted) inverses, the graph of arcsine is obtained by reflecting about the diagonal the piece of the graph of sine function (shown in Figure 2.6.2 in green) corresponding to the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

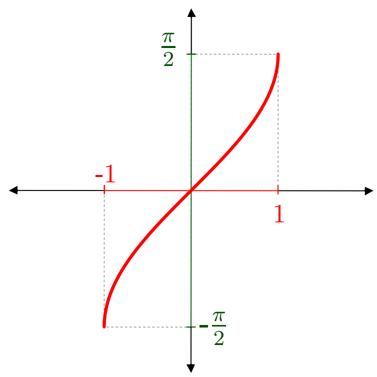


Figure 2.6.3

The above figure depicts the graph of arcsin (shown in thick red), together with its domain, $[-1,1]$ and its range $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Features of the Arcsine Function

(i) Domain $= [-1,1]$.

(ii) Range $= \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

(iii) y-intercept $= 0$

(iv) x-intercept $= 0$

(v) absolute maximum value $= \frac{\pi}{2}$, at $x = 1$

(vi) absolute minimum value $= -\frac{\pi}{2}$, at $x = -1$

(vii) increasing: on the *entire domain* $[-1,1]$.

(viii) decreasing: on no interval

(ix) even/odd: the arcsin function is *odd*

(x) *inversion formulas:*

$$\sin(\arcsin x) = x, \quad \text{for all } x \text{ in } [-1,1]; \tag{2.6.6}$$

$$\arcsin(\sin\alpha) = \alpha, \quad \text{for all } \alpha \text{ in } \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \tag{2.6.7}$$

We can do calculations with values of arcsin, without even caring about what these values really are! After all, in Algebra we freely manipulate numbers like $\sqrt{2}$, $\sqrt{3}$, etc. without needing to approximate their values.

Read the example below very carefully! It contains several important steps needed for an accurate calculation involving arcsin.

Example 2.6.6. Suppose we want to find the **exact** value of

$$\cos\left(\arcsin\left(-\frac{1}{3}\right)\right).$$

For this type of a problem, we do not care of what the value of $\arcsin\left(-\frac{1}{3}\right)$ really is. Instead, we will denote it by some symbol, so using the meaning of arcsine, as in (2.6.5), we can let for instance

$$\arcsin\left(-\frac{1}{3}\right) = \alpha, \text{ meaning that: } \begin{cases} \sin\alpha = -\dfrac{1}{3} \\ -\dfrac{\pi}{2} \leq \alpha \leq \dfrac{\pi}{2} \end{cases} \tag{2.6.8}$$

The above characterization determines α uniquely! Using this notation, our problem can be re-stated as follows: *Given α, characterized by* (2.6.8), *find* $\cos\,\alpha$.

To solve our problem, as restated above, we need to get a better handle on α, particularly to identify the *quadrant*. We do this by *narrowing down the interval*. For this purpose, we simply notice that, since α sits in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, it can only be either

- in Quadrant IV, that is, in the interval $\left[-\frac{\pi}{2}, 0\right]$, or
- in Quadrant I, that is, in the interval $\left[0, \frac{\pi}{2}\right]$.

On the other hand, using (2.6.8), we know that $\sin\alpha$ is *negative*, which leaves us with only one possibility: *α sits in $\left[-\frac{\pi}{2}, 0\right]$, thus in Quadrant IV.*

We can now compute $\cos\,\alpha$ using the Coordinate Method explained in Section 2.2. All we need is to represent α as an angle in standard position, and then find one point on its terminal side. Using the Coordinate Formulas, by which we know that $\sin\,\alpha = \frac{y}{r}$, we can try to look for a point P with $y = 1$ and $r = 3$, so using the formula $r = \sqrt{x^2 + y^2}$, we get (after taking squares): $9 = x^2 + 1$, thus $x^2 = 8$, which yields $x = \pm\sqrt{8}$. Of course, since P is in quadrant IV, its x-coordinate is ≥ 0, so we must have $x = \sqrt{8}$, which finally gives:

$$\cos\left(\arcsin\left(-\frac{1}{3}\right)\right) = \cos\,\alpha = \frac{x}{r} = \frac{\sqrt{8}}{3}.$$

The calculation done in the preceding Example will be improved upon soon! (See the topic entitled "Identities for Inverse Trigonometric Functions" below.)

If we want do a construction similar to the one performed above, but for cosine instead of sine, we need to identify a set D, on which *cosine is one-to-one*.

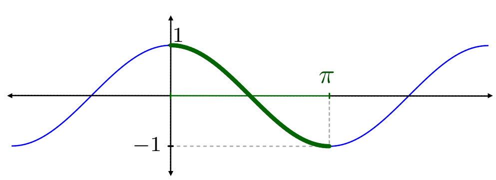


Figure 2.6.4

Based what we see in the above picture, a suitable choice of such a set is now: $D = [0, \pi]$. After all (see Section 2.4), we also know that *cosine is decreasing on* $[0, \pi]$, and this clearly guarantees one-to-one-ness. With this choice in mind, we can now define the following function.

The **arccosine** function, denoted by arccos, is the *restricted inverse of the cosine function, relative to* $[0,\pi]$. The *domain of* arccos is the interval $[-1,1]$. Therefore:

$$\arccos \mathit{number} = ? \text{ means: } \begin{cases} \cos ? = \mathit{number} \\ -1 \leq \mathit{number} \leq 1 \\ 0 \leq ? \leq \pi \end{cases} \tag{2.6.9}$$

Using what we learned about graphing (restricted) inverses, the graph of arccosine is obtained by reflecting about the diagonal the piece of the graph of cosine function (shown in Figure 2.6.4 in green) corresponding to the interval $[0,\pi]$.

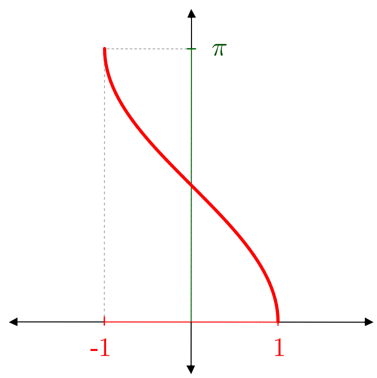


Figure 2.6.5

The above figure depicts the graph of arccos (shown in thick red), together with its domain, $[-1,1]$ and its range $[0,\pi]$.

Features of the Arccosine Function

(i) Domain $=[-1,1]$.
(ii) Range $=[0,\pi]$
(iii) y-intercept $=\dfrac{\pi}{2}$
(iv) x-intercept $=1$
(v) absolute maximum value $=\pi$, at $x=-1$
(vi) absolute minimum value $=0$, at $x=1$
(vii) increasing: on no interval.
(viii) decreasing: on the *entire domain* $[-1,1]$
(ix) even/odd: the arccos function is *neither odd, nor even*
(x) *inversion formulas:*

$$\cos(\arccos x) = x, \quad \text{for all } x \text{ in } [-1,1]; \tag{2.6.10}$$

$$\arccos(\cos\alpha) = \alpha, \quad \text{for all } \alpha \text{ in } [0,\pi]. \tag{2.6.11}$$

The arcsin and arccos functions go "hand in hand," due to the fact that sine and cosine are related by the identity

$$\cos\alpha = \sin\left(\frac{\pi}{2}-\alpha\right).$$

If we inspect carefully the definitions of arcsin and arccos, then the above identity will give us the following important formula.

Complement Formula for Arcsine and Arccosine

$$\arcsin x + \arccos x = \frac{\pi}{2}, \quad \text{for all } x \text{ in } [-1, 1]. \tag{2.6.12}$$

From (2.6.12) one can easily derive the identities

$$\arccos x = \frac{\pi}{2} - \arcsin x, \tag{2.6.13}$$

$$\arcsin x = \frac{\pi}{2} - \arccos x, \tag{2.6.14}$$

which hold for all x in $[-1, 1]$.

In particular, using (2.6.13) we can see that the graph of arccos is obtained from the graph of arcsin by

- flipping about the x-axis (which will produce the graph of $f(x) = -\arcsin x$), then
- shifting upward by $\frac{\pi}{2}$ (which will produce the graph of $\frac{\pi}{2} + f(x) = \arccos x$).

Example 2.6.7. Suppose we are asked to find the **exact** values of $\arcsin\left(-\frac{1}{2}\right)$ and $\arccos\left(-\frac{1}{2}\right)$.

From the list of "familiar" values for sine, we remember that $\sin\frac{\pi}{6} = \frac{1}{2}$. Using the formula for negatives for sine, it follows that

$$\sin\left(-\frac{\pi}{6}\right) = -\frac{1}{2}.$$

If we compare this calculation with the definition of arcsine, we clearly see that we have a good match, that is $? = -\frac{\pi}{6}$ does satisfy all conditions in (2.6.5). So it is safe to conclude that:

$$\arcsin\left(-\frac{1}{2}\right) = -\frac{\pi}{6}.$$

As for $\arccos\left(-\frac{1}{2}\right)$, although we may try to do a similar thing for arccosine, it is easier if we use (2.6.13) instead, which quickly gives us

$$\arccos\left(-\frac{1}{2}\right) = \frac{\pi}{2} - \arcsin\left(-\frac{1}{2}\right) = \frac{\pi}{2} - \left(-\frac{\pi}{6}\right) = \frac{\pi}{2} + \frac{\pi}{6} = \frac{3\pi}{6} + \frac{\pi}{6} = \frac{4\pi}{6}\,\frac{2\pi}{3}.$$

Example 2.6.8. When we are asked to *approximate* certain values of arcsin or arccos, we can always use a calculator, where these functions are nothing else but our "old friends" $\boxed{\texttt{sin}^{-1}}$ and $\boxed{\texttt{cos}^{-1}}$. For instance, using a calculator, we can approximate

$$\arcsin(-0.6) \simeq -0.643501109; \quad \arccos(-0.6) \simeq 2.214297436. \tag{2.6.15}$$

Of course, in order to get this result we need to set the calculator to work with radians. What happens when degrees are used instead? The answer is very simple:

- instead of producing an output in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, the calculator function $\boxed{\texttt{sin}^{-1}}$ will yield a value in the interval $[-90, 90]$;
- instead of producing an output in $[0, \pi]$, the calculator function $\boxed{\texttt{cos}^{-1}}$ will yield a value in the interval $[0, 180]$.

For instance, the values (2.6.15) computed in degrees will now be:

$$\arcsin(-0.6) \simeq -36.86989765°; \quad \arccos(-0.6) \simeq 126.86989765°.$$

Some texts denote arcsin by $\sin^{-1}$, and arccos by $\cos^{-1}$. Such notations are to be avoided, as they are misleading in many ways. First of all, as "full" functions sine and cosine are not one-to-one, so they to not have "full" function inverses. Secondly, the use of exponents has been established as a shortcut for powers of trigonometric functions. (For instance, $\sin^2 x$ is a shortcut for $(\sin x)^2$.) The only reason we see $\boxed{\texttt{sin}^{-1}}$ and $\boxed{\texttt{cos}^{-1}}$ on a calculator is space-saving!

The Arctangent Function

Suppose now we want to find a restricted inverse of the tangent function, for which we need to identify a set D, on which *tangent is one-to-one*.

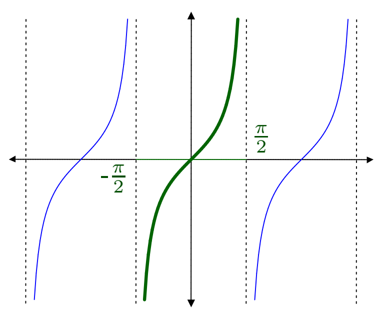


Figure 2.6.6

Based what we see in the above picture, a suitable choice of such a set is now: $D = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. After all (see Section 2.4), we also know that *tangent is increasing on* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and this clearly guarantees one-to-one-ness. With this choice in mind, we can now define the following function.

The **arctangent** function, denoted by arctan, is the *restricted inverse of the tangent function, relative to* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The *domain of* arctan is the *set of all real numbers*. Therefore:

$$\arctan \textit{number} = ? \text{ means: } \begin{cases} \tan ? = \textit{number} \\ \\ -\frac{\pi}{2} < ? < \frac{\pi}{2} \end{cases} \tag{2.6.16}$$

Using what we learned about graphing (restricted) inverses, the graph of arctangent is obtained by reflecting about the diagonal the piece of the graph of tangent function (shown in Figure 2.6.6 in green) corresponding to the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

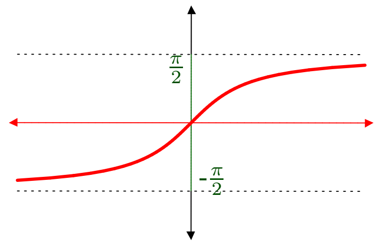


Figure 2.6.7

The above figure depicts the graph of arctan (shown in thick red) and its range $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Features of the Arctangent Function

(i) Domain = *all real numbers*.
(ii) Range $= \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$
(iii) y-intercept $= 0$
(iv) x-intercept $= 0$
(v) local maximum value = *none*
(vi) local minimum value = *none*
(vii) increasing: on the *entire domain*
(viii) decreasing: on *no interval*.
(ix) even/odd: the arctan function is *odd*
(x) *inversion formulas:*

$$\tan(\arctan x) = x, \quad \text{for } all\ x; \tag{2.6.17}$$

$$\arctan(\tan \alpha) = \alpha, \quad \text{for all } \alpha \text{ in } \left(-\frac{\pi}{2}, \frac{\pi}{2}\right). \tag{2.6.18}$$

Example 2.6.9. Suppose we are asked to find the **exact** value of $\arctan\left(-\sqrt{3}\right)$.

From the list of "familiar" values for sine, we remember that $\tan\frac{\pi}{3} = \sqrt{3}$. Using the formula for negatives for tangent, it follows that

$$\tan\left(-\frac{\pi}{3}\right) = -\sqrt{3}.$$

If we compare this calculation with the definition of arctangent, we clearly see that we have a good match, that is $? = -\frac{\pi}{3}$ does satisfy all conditions in (2.6.16). So it is safe to conclude that:

$$\arctan\left(-\sqrt{3}\right) = -\frac{\pi}{3}.$$

As was the case with arcsin and arccos, values of arctan can also be *approximated* using the calculator function $\boxed{\texttt{tan}^{-1}}$, with the calculator set to radians. As was the case with $\boxed{\texttt{sin}^{-1}}$, when degrees are used, the output will be in the interval $(-90, 90)$.

As previously explained, a notation like $\tan^{-1}$ instead of arctan should be avoided.

Identities for Inverse Trigonometric Functions

Exactly as was the case with the usual trigonometric functions, for which we have quite a few identities linking them (see Section 2.3), there are also several identities that involve arcsin, arccos, and arctan. The list of all possible identities is quite large, so we will only limit ourselves to the most important ones. (Other, but not all possible ones, are moved to the Exercise section.)

Among the many identities , six have already been identified above, as the so-called *inversion formulas*: (2.6.6), (2.6.7), (2.6.10), (2.6.11), (2.6.17), and (2.6.18). Two additional identities we have already dsicussed are those for *complements*: (2.6.13) and (2.6.14).

Another set of identities, which concerns *negatives*, is as follows.

Formulas for Negatives

$$\arcsin(-x) = -\arcsin x, \quad \text{for every } x \text{ in } [-1,1]; \tag{2.6.19}$$
$$\arccos(-x) = \pi - \arccos x, \quad \text{for every } x \text{ in } [-1,1]; \tag{2.6.20}$$
$$\arctan(-x) = -\arctan x, \quad \text{for every real number } x. \tag{2.6.21}$$

CLARIFICATIONS. Formulas (2.6.19) and (2.6.21) simply say that arcsin and arcsin are odd. As for the curious formula (2.6.20), it can be easily derived from (2.6.19) and the complement formulas (2.6.13), (2.6.14), because we can write

$$\arccos(-x) = \frac{\pi}{2} - \arcsin(-x) = \frac{\pi}{2} + \arcsin\ x = \frac{\pi}{2} + \frac{\pi}{2} - \arccos\ x = \pi - \arccos x.$$

We conclude with three sets of identities, which expand the inversion formulas (2.6.6), (2.6.10) and (2.6.17). In a certain sense, these identities mirror those derived from the "Holy Grail of Trigonometry."

Trigonometric Functions of Inverses

Trigonometric Functions of arcsine. For every x in $[-1,1]$, we have:

$$\sin\left(\arcsin x\right) = x; \tag{2.6.22}$$
$$\cos\left(\arcsin x\right) = \sqrt{1-x^2}; \tag{2.6.23}$$
$$\tan\left(\arcsin x\right) = \begin{cases} \dfrac{x}{\sqrt{1-x^2}}, \text{ if } x \neq \pm 1; \\ \text{undefined, if } x = \pm 1; \end{cases} \tag{2.6.24}$$
$$\cot\left(\arcsin x\right) = \begin{cases} \dfrac{\sqrt{1-x^2}}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.25}$$
$$\sec\left(\arcsin x\right) = \begin{cases} \dfrac{1}{\sqrt{1-x^2}}, \text{ if } x \neq \pm 1; \\ \text{undefined, if } x = \pm 1; \end{cases} \tag{2.6.26}$$
$$\csc\left(\arcsin x\right) = \begin{cases} \dfrac{1}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.27}$$

Trigonometric Functions of arccosine. For every x in $[-1, 1]$, we have:

$$\sin\left(\arccos x\right) = \sqrt{1-x^2}; \tag{2.6.28}$$

$$\cos\left(\arccos x\right) = x; \tag{2.6.29}$$

$$\tan\left(\arccos x\right) = \begin{cases} \dfrac{\sqrt{1-x^2}}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.30}$$

$$\cot\left(\arccos x\right) = \begin{cases} \dfrac{x}{\sqrt{1-x^2}}, \text{ if } x \neq \pm 1; \\ \text{undefined, if } x = \pm 1; \end{cases} \tag{2.6.31}$$

$$\sec\left(\arccos x\right) = \begin{cases} \dfrac{1}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.32}$$

$$\csc\left(\arccos x\right) = \begin{cases} \dfrac{1}{\sqrt{1-x^2}}, \text{ if } x \neq \pm 1; \\ \text{undefined, if } x = \pm 1; \end{cases} \tag{2.6.33}$$

Trigonometric Functions of arctangent. For every real number x, we have:

$$\sin\left(\arctan x\right) = \frac{x}{\sqrt{1+x^2}}; \tag{2.6.34}$$

$$\cos\left(\arctan x\right) = \frac{1}{\sqrt{1+x^2}}; \tag{2.6.35}$$

$$\tan\left(\arctan x\right) = x. \tag{2.6.36}$$

$$\cot\left(\arctan x\right) = \begin{cases} \dfrac{1}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.37}$$

$$\sec\left(\arctan x\right) = \sqrt{1+x^2}; \tag{2.6.38}$$

$$\csc\left(\arctan x\right) = \begin{cases} \dfrac{\sqrt{1+x^2}}{x}, \text{ if } x \neq 0; \\ \text{undefined, if } x = 0; \end{cases} \tag{2.6.39}$$

CLARIFICATIONS. To obtain (2.6.23) and the rest of the formulas for arcsin, we work exactly as in Example 2.6.6. We set $\arcsin x = \alpha$, so that α is in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, and $\sin\alpha = x$, so that what we need to find is $\cos\alpha$. We know that $\cos\alpha = \pm\sqrt{1-\sin^2\alpha} = \pm\sqrt{1-x^2}$. However, since α is either in Quadrant I or in Quadrant IV, we also know that $\cos\alpha \geq 0$, so regardless what α is, the correct value is $\cos\alpha = +\sqrt{1-x^2}$. With the value of $\cos\alpha$ computed, the value of $\tan\alpha$, and of the others follow immediately.

The identity (2.6.38) and the rest of the formulas for arctan are obtained similarly, by setting up $\arctan x = \alpha$, which is again in Quandrant I or IV, on which $\sec\alpha \geq 0$, and so we get $\sec\alpha = +\sqrt{1+\tan^2\alpha} = +\sqrt{1+x^2}$, from which everything else follows.

To get the identity (2.6.28) and the rest of the formulas for arccos we set up $\arccos x = \alpha$, which is now in $[0, \pi]$ (Quadrant I or II), for which $\sin\alpha \geq 0$. As above, this will give $\cos\alpha = +\sqrt{1-\sin^2\alpha} = +\sqrt{1-x^2}$, from which everything else follows.

Example 2.6.10. Suppose we want to compute the **exact** values all six trigonometric functions of $\arctan(-2)$. We can use identities (2.6.34) and (2.6.36) with $x = -2$, which immediately yield:

$$\sin\big(\arctan(-2)\big) = \frac{-2}{\sqrt{1+(-2)^2}} = -\frac{2}{\sqrt{5}};$$
$$\cos\big(\arctan(-2)\big) = \frac{1}{\sqrt{1+(-2)^2}} = \frac{1}{\sqrt{5}}.$$

Since we also know $\tan\big(\arctan(-2)\big) = -2$, the other functions (secant, cosecant and cotangent) are easily computed using reciprocals:

$$\sec\big(\arctan(-2)\big) = \frac{1}{\cos\big(\arctan(-2)\big)} = \sqrt{5};$$
$$\csc\big(\arctan(-2)\big) = \frac{1}{\sin\big(\arctan(-2)\big)} = -\frac{\sqrt{5}}{2};$$
$$\cot\big(\arctan(-2)\big) = \frac{1}{\tan\big(\arctan(-2)\big)} = -\frac{1}{2}.$$

Exercises

In Exercises 1–12 you are asked to find **exact values** of several inverse trigonometric functions. As you will see, all calculations relie on "familiar" values.

1. $\arcsin\left(\frac{\sqrt{3}}{2}\right)$

2. $\arcsin\left(-\frac{\sqrt{3}}{2}\right)$

3. $\arcsin\left(\frac{\sqrt{2}}{2}\right)$

4. $\arcsin\left(-\frac{\sqrt{2}}{2}\right)$

5. $\arcsin\left(-4\right)$

6. $\arcsin\left(\sin\frac{8\pi}{3}\right)$

7. $\arccos\left(-\frac{\sqrt{3}}{2}\right)$

8. $\arccos\left(-1\right)$

9. $\arccos\big(3\big)$

10. $\arccos\big(\cos\dfrac{5\pi}{3}\big)$

11. $\arctan\big(1\big)$

12. $\arctan\big(-1\big)$

In Exercises 13–17 you are asked to compute **exact values** of several inverse trigonometric functions. Use either Example 2.6.6 or Example 2.6.10 as a guideline

13. $\cos\Big(\arcsin\big(\dfrac{12}{13}\big)\Big)$

14. $\sin\Big(\arccos\big(-\dfrac{8}{17}\big)\Big)$

15. $\tan\Big(\arcsin\big(-\dfrac{3}{5}\big)\Big)$

16. $\cot\Big(-\arcsin\big(\dfrac{1}{7}\big)\Big)$

17. $\cos\Big(\arctan\big(10\big)\Big)$

18*. Find the exact value of $\arccos\big(\cos 10\big)$

19*. Find the exact value of $\arctan\big(\tan(-15)\big)$

2.7 Trigonometric Equations

In this section we discuss *trigonometric equations*, which are those types of equations that are presented as

$$\textit{(Left) Expression} = \textit{(Right) Expression},$$

where each side is an "*expression*" that contains *trigonometric functions involving an unknown number*. When we deal with such equalities, the first question we need to ask ourselves is this: *Is the given equality an identity?* If the answer is "yes," then the equation is very easy to solve. For instance, if we look at the "equation"

$$\sin^2 x + \cos^2 x = 1,$$

this is in fact an identity, so its solutions are *all real numbers*.

The problem becomes interesting, when the given equality is *not* an identity. So what we call "*trigonometric equations*" here are nothing else but what we called "*false identities*" in Section 2.3.

Before we deal with complicated equations, we will revisit the *elementary* trigonometric equations, for which we will improve the solving method using the *inverse trigonometric functions* which we introduced in Section 2.6.

The Elementary Tangent Equation

The equation we are interested in here is of the form

$$\tan x = \textit{number}, \tag{2.7.1}$$

where x (or any other letter) is the *unknown*, and *number* is some constant.

In Section 2.4 we learned that, typically, equations like (2.7.1) have all solutions presented as *one list of numbers*, presented in the form

$$x = x^{\text{basic}} + n\pi, \quad n \text{ integer}, \tag{2.7.2}$$

where x^{basic} stands for what we called the *basic solution*, which is the *solution in the interval* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. In Section 2.4 we also learned that the basic solution was of the form $x^{\text{basic}} = \pm x^{\text{ref}}$, where the sign matches the *sign of number*.

Instead of "playing this sign game" (as we did in Section 2.4; see for instance, Example 2.4.6), we can streamline our old method by observing that, no matter what sign *number* has, the basic solution x^{basic} sits in the *range of* $\arctan$, so it must be equal to $\arctan(\textit{number})$. This way, our new approach to solving (2.7.1) is not to worry about signs, but simply to say the following.

Solutions of Elementary Tangent Equation

For any real number, the solutions of the equation

$$\tan x = \textit{number}$$

are given by:

$$x = \arctan(number) + n\pi, \; n \textit{ integer}.$$

The Elementary Sine and Cosine Equations

The equations we are interested in here are those of one of the forms

$$\sin x = \textit{number}, \tag{2.7.3}$$

$$\cos x = \textit{number}. \tag{2.7.4}$$

In Section 2.4 we learned that, typically, equations like (2.7.3) or (2.7.4) have all solutions presented as *lists of numbers*, presented in the form

$$x = x^{\text{basic}} + 2n\pi, \quad n \text{ integer}, \tag{2.7.5}$$

where x^{basic} stands for what we called the *basic solution(s)*, which were those *solutions in the interval* $[0, 2\pi)$. In Section 2.4 we also learned that the basic solutions are given by certain formulas, which depended on the *sign of number*, and this complicated things a little bit.

What we are about to do will be to "get rid of our headaches" concerning the sign of *number*, and instead of writing our solutions as in (2.7.5), we will try to write them as

$$x = x^{\text{easy}} + 2n\pi, \quad n \text{ integer}, \tag{2.7.6}$$

where x^{easy} will be some other types of solutions, which will be easier to find than x^{basic}, especially when we use arcsin and arccos.

As it turns out, the cause of all our "headaches" related to basic solutions is not the sign of *number*, but the *interval* where the basic solutions sit! If instead of $[0, 2\pi)$ we choose another half-open interval, which still has length 2π, our task might be simplified. Those improved intervals, in which the elementary equations can be solved easier, will be of the form $[\alpha, \alpha + 2\pi)$ or $(\alpha, \alpha + 2\pi]$ (to be identified shortly), and will be called *"easy" intervals*. Whatever our "easy" interval will be, the number(s) denoted by x^{easy} will be the *solution(s) of our equation in the "easy" interval*. As for our choice for such intervals, we will follow a very simple rule:

> *If* function *designates either one of* sin *or* cos, *an "easy" interval is a half-open interval of length* 2π, *such that* function *has its minimum value at the endpoints of the interval*.

In particular, the graph of function over an "easy" interval looks like:

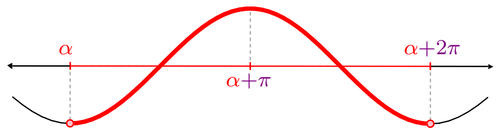


Figure 2.7.1

As seen in Figure 2.7.1, the midpoint of the "easy" interval is at $\alpha+\pi$, and at this point function (which is either sin or cos) attains its *maximum value* function$(\alpha+\pi) = 1$.

CLARIFICATIONS. Before we specialize function to either sin or cos, we can say a few things that work for *both* of them. Of course, when we want to find the "easy" solutions of the elementary equation

$$\text{function}(x) = \textit{number}, \tag{2.7.7}$$

we need to find the x-coordinate(s) of the point(s) where the horizontal line $y = \textit{number}$ intersects the graph, over the "easy" interval.

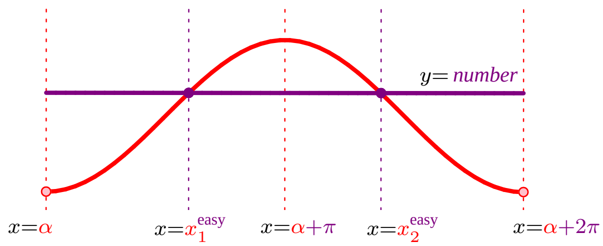


Figure 2.7.2

By inspecting Figure 2.7.2 above, we see that:

- If *number* > 1, or *number* < -1, then the equation (2.7.7) has *no "easy" solutions*.
- If *number* $= -1$, then the equation (2.7.7) has *exactly one "easy" solution*, which is *one of endpoints of the "easy" interval*: $x^{\text{easy}} = \alpha,\ \alpha+2\pi$. (This depends on where we choose the "easy" interval to be closed.)
- If *number* $= 1$, then the equation (2.7.7) again has *exactly one "easy" solution*, which is *the midpoint of the "easy" interval*: $x^{\text{easy}} = \alpha+\pi$.

- equation (2.7.7) has *exactly two "easy" solutions:* x_1^{easy} and x_2^{easy}, which satisfy the equality
$$\frac{x_1^{\text{easy}}+x_2^{\text{easy}}}{2}=\alpha+\pi, \tag{2.7.8}$$
where again $\alpha+\pi$ is the *midpoint of the "easy" interval*.

When we specialize to the sine function, our "easy" interval will be $\left[-\frac{\pi}{2},\frac{3\pi}{2}\right)$ (which has $\frac{\pi}{2}$ as its midpoint), so in this case, the situation depicted in Figure 2.7.2 will become:

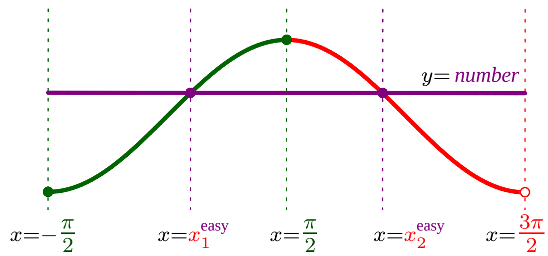


Figure 2.7.3

The picture above (which features case when $-1<$ *number* <1) has the left half of the graph of sine shown in green. As it turns out, this piece corresponds to the interval $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$, which is precisely the *range of* arcsin. This means that one of the "easy" solutions is precisely

$$x_1^{\text{easy}}=\arcsin(\textit{number}).$$

As for the second "easy" solution, using the midpoint formula (2.7.8), which in our case reads $\frac{x_1^{\text{easy}}+x_2^{\text{easy}}}{2}=\frac{\pi}{2}$, we immediately get $x_1^{\text{easy}}+x_2^{\text{easy}}=\pi$, so:

$$x_2^{\text{easy}}=\pi-x_1^{\text{easy}}=\pi-\arcsin(\textit{number}).$$

These findings can then be summarized as shown below.

"Easy" Solutions of the Elementary Sine Equation

When choosing $\left[-\frac{\pi}{2},\frac{3\pi}{2}\right)$ as the 'easy" interval, the elementary equation

$$\sin x=\textit{number},$$

has the following "easy" solutions.

I. If *number* >1, or *number* <-1, there are *no "easy" solutions*.
II. If *number* $=\pm1$, then there is *exactly one "easy" solution*: $x^{\text{easy}}=\arcsin(\textit{number})$.
III. If $-1<$ *number* <1, then there are *exactly two "easy" solutions:* $x_1^{\text{easy}}=\arcsin(\textit{number})$ and $x_2^{\text{easy}}=\pi-\arcsin(\textit{number})$.

Example 2.7.1. Consider the elementary sine equation

$$\sin x=-\frac{1}{2},$$

and let us find its "easy" solutions.

Since (by the "familiar" values of sine) we know $\arcsin\left(\frac{1}{2}\right) = \frac{\pi}{6}$, using the formula for negatives for arcsin we also know $\arcsin\left(-\frac{1}{2}\right) = -\frac{\pi}{6}$. Using the above method, the "easy" solutions of our equation will now be $x_1^{\text{easy}} = -\frac{\pi}{6}$ and $x_2^{\text{easy}} = \pi - \left(-\frac{\pi}{6}\right) = \pi + \frac{\pi}{6} = \frac{7\pi}{6}$.

When we are interested in *all* solutions of an elementary sine equation, we are now going to use formula (2.7.6), which in the case when $-1 < \textit{number} < 1$ will produce *two lists* of numbers

$$x = \arcsin(\textit{number}) + 2n\pi,\ n \text{ integer}; \tag{2.7.9}$$
$$x = \pi - \arcsin(\textit{number}) + 2n\pi,\ n \text{ integer}. \tag{2.7.10}$$

As it turns out, the above two lists can be considered even in the case when $\textit{number} = \pm 1$, but in that case we will have a *duplication*, meaning that the two lists will produce the same numbers.

In any case, it is worth point out that each list produces numbers of the form

$$\pm\arcsin(\textit{number}) + \text{multiple of } \pi.$$

Using this observation the lists (2.7.9) and (2.7.10) can in fact be *combined*, to produce a *single list* of numbers, as shown below.

The "Clean" Solution of the Elementary Sine Equation

The elementary sine equation

$$\sin x = \textit{number},$$

has all its solutions given as follows.

I. If $\textit{number} > 1$, or $\textit{number} < -1$, there are *no solutions*.

II. If $-1 \le \textit{number} \le 1$, then all solutions are of the form:

$$x = (-1)^k \arcsin(\textit{number}) + k\pi,\ k \text{ integer}. \tag{2.7.11}$$

CLARIFICATIONS. The "clean" list given in (2.7.11) can be split into two halves, each one corresponding to one of the two lists (2.7.9) and (2.7.10), based on the *parity of* k.

(a) If k is *even*, thus of the form $k = 2n$, for some integer n, then $(-1)^k = +1$, so the list (2.7.11) produces $x = \arcsin(\textit{number}) + 2n\pi$, a number in (2.7.9).

(b) If k is *odd*, thus of the form $k = 2n+1$, for some integer n, then $(-1)^k = -1$, so the list (2.7.11) produces $x = -\arcsin(\textit{number}) + \pi + 2n\pi$, a number in (2.7.10).

Another nice feature of the "clean" formula (2.7.11) is that it lists the solutions in *increasing order*. (See Example 2.7.2 below.)

As pointed out earlier, in the case when $\textit{number} = \pm 1$, the two halves of (2.7.11) – obtained by choosing k to be either *even* or *odd* – will produce same numbers.

Since $\arcsin(0) = 0$, we see that in the case when $\textit{number} = 0$, the "clean" list (2.7.11) will produce

$$x = k\pi,\ k \text{ integer},$$

which are precisely the x-*intercepts* of sine.

Example 2.7.2. Consider the same elementary sine equation

$$\sin x = -\frac{1}{2},$$

as in Example 2.7.1, and let us now find all of its solutions.

Using the "clean" formula (2.7.11), we can simply write:

$$x = (-1)^k \arcsin\left(-\frac{1}{2}\right) + k\pi = (-1)^k\left(-\frac{\pi}{6}\right) + k\pi = \frac{(-1)^{k+1}\pi}{6} + k\pi,\ k \text{ integer}.$$

(Notice that we simplified $-(-1)^k = (-1)(-1)^k = (-1)^{k+1}$.)

We illustrate how the above formula works by plugging in several values of k

k	-2	-1	0	1	2	3
x	$-\frac{13\pi}{6}$	$-\frac{5\pi}{6}$	$-\frac{\pi}{6}$	$\frac{7\pi}{6}$	$\frac{11\pi}{6}$	$\frac{19\pi}{6}$

Let us see now how the whole story above changes, when we use cosine instead of sine. A good choice for our "easy" interval will be $(-\pi, \pi]$ (which has 0 as its midpoint), so in this case, the situation depicted in Figure 2.7.2 will become:

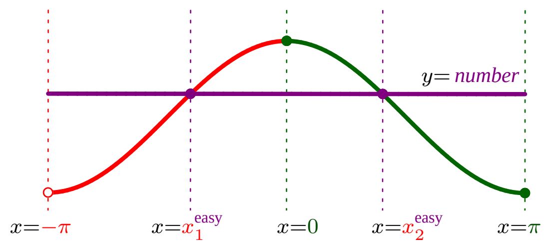


Figure 2.7.4

The picture above (which features case when $-1 <$ *number* < 1) has the right half of the graph of cosine shown in green. As it turns out, this piece corresponds to the interval $[0, \pi]$, which is precisely the *range of* arccos. This means that one of the "easy" solutions is precisely

$$x_2^{\text{easy}} = \arccos(\textit{number}).$$

As for the other "easy" solution, using the midpoint formula (2.7.8), which in our case reads $\dfrac{x_1^{\text{easy}} + x_2^{\text{easy}}}{2} = 0$, we immediately get $x_1^{\text{easy}} + x_2^{\text{easy}} = 0$, so:

$$x_1^{\text{easy}} = -x_2^{\text{easy}} = -\arccos(\textit{number}).$$

These findings can then be summarized as shown below.

"Easy" Solutions of the Elementary Cosine Equation

When choosing $(-\pi, \pi]$ as the 'easy' interval, the elementary equation

$$\cos x = \mathit{number},$$

has the following "easy" solutions.

I. If $\mathit{number} > 1$, or $\mathit{number} < -1$, there are *no "easy" solutions*.

II. If $\mathit{number} = \pm 1$, then there is *exactly one "easy" solution*: $x^{\text{easy}} = \arccos(\mathit{number})$.

III. If $-1 < \mathit{number} < 1$, then there are *exactly two "easy" solutions:* $x_1^{\text{easy}} = -\arccos(\mathit{number})$ and $x_2^{\text{easy}} = \arccos(\mathit{number})$.

Example 2.7.3. Consider the elementary cosine equation

$$\cos x = -\frac{\sqrt{2}}{2},$$

and let us find its "easy" solutions.

Since (by the "familiar" values of sine) we know $\arccos\left(\frac{\sqrt{2}}{2}\right) = \frac{\pi}{4}$, using the formula for negatives for arccos we also know $\arccos\left(-\frac{\sqrt{2}}{2}\right) = \pi - \frac{\pi}{4} = \frac{3\pi}{4}$. Using the above method, the "easy" solutions of our equation will now be $x_1^{\text{easy}} = -\frac{3\pi}{4}$ and $x_2^{\text{easy}} = \frac{3\pi}{4}$.

When we are interested in *all* solutions of an elementary cosine equation, we are now going to use formula (2.7.6), which in the case when $-1 < \mathit{number} < 1$ will produce *two lists* of numbers

$$x = -\arccos(\mathit{number}) + 2n\pi,\ n \text{ integer}; \tag{2.7.12}$$
$$x = \arccos(\mathit{number}) + 2n\pi,\ n \text{ integer}. \tag{2.7.13}$$

As it turns out, the above two lists can be considered even in the case when $\mathit{number} = \pm 1$, but in that case we will have a *duplication*, meaning that the two lists will produce the same numbers.

Of course, the lists (2.7.12) and (2.7.13) can be easily, to produce a *single list* of numbers, as shown below.

The "Clean" Solution of the Elementary Cosine Equation

The elementary cosine equation

$$\cos x = \mathit{number},$$

has all its solutions given as follows.

I. If $\mathit{number} > 1$, or $\mathit{number} < -1$, there are *no solutions*.

II. If $-1 \leq \mathit{number} \leq 1$, then all solutions are of the form:

$$x = \pm\arccos(\mathit{number}) + 2n\pi,\ n \text{ integer}. \tag{2.7.14}$$

CLARIFICATION. As pointed out earlier, in the case when $\mathit{number} = \pm 1$, the two halves of (2.7.14), corresponding to the $+$ or $-$ sign, will produce the same numbers.

Also, using the equality $\arcsin(0) = \frac{\pi}{2}$, we see that in the case when *number* $= 0$, the "clean" list (2.7.11) will produce the numbers $x = \pm\frac{\pi}{2} + 2n\pi$, which can be rewritten in one single list of the form

$$x = \frac{\pi}{2} + k\pi,\ k \text{ integer},$$

which are precisely the x*-intercepts* of cosine.

Example 2.7.4. Consider the same elementary cosine equation

$$\cos x = -\frac{\sqrt{2}}{2},$$

as in Example 2.7.3, and let us now find all of its solutions.

Using the "clean" formula (2.7.14), we can simply write:

$$x = \pm\arccos\left(-\frac{\sqrt{2}}{2}\right) + 2n\pi = \pm\frac{3\pi}{4} + 2n\pi,\ n \text{ integer}.$$

Substitution Methods

When dealing with more complicated equations, we often employ the *substitution* technique, which allow us to *reduce* an equation to a simpler one.

The Substitution Method

Suppose a *(complicated) equation in* x (or any other letter that designates the unknown) is to be solved. We solve the given (complicated) equation by setting up a *substitution* equation, which involves a new variable (say, y, or any other symbol we want), which is of the form

$$\textit{(simple) expression in } x = y, \tag{2.7.15}$$

so that the given equation becomes a

$$\textit{(simpler) equation in } y, \tag{2.7.16}$$

which we refer to as the *reduced* equation. We deem our substitution *successful*, if both the reduced equation (2.7.16) and the substitution equation (2.7.15) are *easy to solve*. If this is the case, the given (complicated) equation can be solved by carrying on the following two steps.

I. Solve the *reduced* equation (2.7.16) for y. (Find all solutions.)

II. With each value of y, found in step I, go back to the *substitution* equation (2.7.15), and solve for x.

Example 2.7.5. Suppose we want to find all solutions of the equation

$$\cos\left(2x - \frac{\pi}{3}\right) = \frac{1}{2}.$$

Using the method outlined above, we start off with the substitution

$$2x - \frac{\pi}{3} = y, \tag{2.7.17}$$

so the given equation reduces to the elementary cosine equation:

$$\cos y = \frac{1}{2}. \tag{2.7.18}$$

I. Using what we learned about elementary cosine equations, all solutions of (2.7.18) are:

$$y = \pm \arccos\left(\frac{1}{2}\right) + 2n\pi = \pm\frac{\pi}{3} + 2n\pi,\ n \text{ integer}. \tag{2.7.19}$$

II. With the above values of y, we now go back to the substitution (2.7.17), which becomes

$$2x - \frac{\pi}{3} = \pm\frac{\pi}{3} + 2n\pi,\ n \text{ integer}.$$

Since the above equation is *linear*, in order to solve it, we first add $\frac{\pi}{3}$, which yields

$$2x = \frac{\pi}{3} \pm \frac{\pi}{3} + 2n\pi,\ n \text{ integer},$$

and then we divide by 2, which is the same as multiplying by $\frac{1}{2}$, which gives (after distributing $\frac{1}{2}$ to all terms):

$$x = \frac{1}{2}\left[\frac{\pi}{3} \pm \frac{\pi}{3} + 2n\pi\right] = \frac{\pi}{6} \pm \frac{\pi}{6} + n\pi,\ n \text{ integer}.$$

If we wish, the above solution(s) can be "cleaned up" a little bit, by splitting the list into two halves (depending on whether $+$ or $-$ is selected). As a result, we find that our solutions can be divided into two nicer looking lists:

$$x = \frac{\pi}{6} + \frac{\pi}{6} + n\pi = \frac{\pi}{3} + n\pi,\ n \text{ integer}; \tag{2.7.20}$$

$$x = \frac{\pi}{6} - \frac{\pi}{6} + n\pi = n\pi,\ n \text{ integer}. \tag{2.7.21}$$

CLARIFICATION. The technique used in the Example above applies to all *"nearly" elementary* equations, which are those of the form

$$\text{function}\big(\textit{expression in } x\big) = \textit{number}, \tag{2.7.22}$$

where function is either one of $\sin$, $\cos$, or $\tan$. The use of the phrase "nearly elementary" is justified by the observation that, once we make an *angle substitution*

$$\textit{expression in } x = y,$$

the equation will be transformed into an elementary equation:

$$\text{function}(y) = \textit{number}.$$

For other (complicated) equations, *function substitutions* are also helpful, as seen in the following example.

Example 2.7.6. Suppose we want to find all solutions of the equation

$$2\sin^2 t - 3\sin t - 2 = 0. \tag{2.7.23}$$

Start off with the function substitution

$$\sin t = z, \tag{2.7.24}$$

so the given equation reduces to an easy quadratic equation.

$$2z^2 - 3z - 2 = 0. \tag{2.7.25}$$

I. Using the Quadratic Formula, the solutions of the reduced equation (2.7.25) are

$$\begin{gathered} z = \frac{-(-3) \pm \sqrt{(-3)^2 - 4 \cdot 2 \cdot (-2)}}{2 \cdot 2} = \frac{3 \pm 5}{4} \\ z_1 = \frac{3+5}{4} = 2; \quad z_2 = \frac{3-5}{4} = -\frac{1}{2}. \end{gathered} \tag{2.7.26}$$

II. Using all solutions (2.7.26) of the reduced equation, we go back to the substitution equation (2.7.24), which we need to solve for t.

(A) When we use $z_1 = 2$, the substitution equation (2.7.24) reads

$$\sin t = 2,$$

which has *no solution*.

(B) When we use $z_2 = -\frac{1}{2}$, the substitution equation (2.7.24) reads

$$\sin t = -\frac{1}{2},$$

which has as solutions (see Example 2.7.2):

$$t = \frac{(-1)^{k+1}\pi}{6} + k\pi, \; k \text{ integer}. \tag{2.7.27}$$

So our conclusion is as follows. The given equation (2.7.23) has as solutions all the numbers given in the list (2.7.27).

Equations Involving More than One Function

So far, we have only treated equations where a single trigonometric function is involved. When two or more functions are involved, it is desirable to reduce the equation to one in which only one function appears. This is illustrated in Example 2.7.7 below. Sometimes, where such a reduction is not feasible, additional techniques, for instance *factoring*, might be helpful, as demonstrated in Example 2.7.8 below.

Example 2.7.7. Consider the equation

$$\sec^2 u = \tan u + 1. \tag{2.7.28}$$

Although two functions contribute to this equation, using the Pythagorean identity

$$\sec^2 u = 1 + \tan^2 u$$

we can eliminate the secant, so our equation becomes

$$1 + \tan^2 u = \tan u + 1,$$

which by subtracting the right-hand side becomes

$$\tan^2 u - \tan u = 0.$$

Using the substitution

$$\tan u = y, \tag{2.7.29}$$

our equation reduces to

$$y^2 - y = 0. \tag{2.7.30}$$

I. Using either the Quadratic Formula, or factoring, the solutions of the reduced equation (2.7.30) are

$$y_1 = 0; \quad y_2 = 1. \tag{2.7.31}$$

II. Using all solutions (2.7.31) of the reduced equation, we go back to the substitution equation (2.7.29), which we need to solve for u.

(A) When we use $y_1 = 0$, the substitution equation (2.7.29) reads

$$\tan u = 0,$$

which has as solutions:

$$u = \arctan 0 + n\pi = n\pi,\ n \text{ integer}. \tag{2.7.32}$$

(B) When we use $y_1 = 1$, the substitution equation (2.7.29) reads

$$\tan u = 1,$$

which has as solutions:

$$u = \arctan 1 + n\pi = \frac{\pi}{4} + n\pi,\ n \text{ integer}. \tag{2.7.33}$$

So our conclusion is as follows. The given equation (2.7.28) has as solutions all the numbers given in the lists (2.7.32) and (2.7.33).

Example 2.7.8. Consider the equation

$$2\sin 2x \cos x = \cos x. \tag{2.7.34}$$

By subtracting the righ-hand side, the equation is equivalent to:

$$2\sin\,2x\,\cos\,x - \cos\,x = 0,$$

which by factoring the left-hand side, reads:

$$(2\sin\,2x\, - 1)\cos\,x = 0. \tag{2.7.35}$$

Since the left-hand side is factored and the right-hand side is zero, the above equation can be split by setting each factor equal zero, thus

$$2\sin\,2x\, - 1 = 0, \text{ or} \tag{2.7.36}$$

$$\cos\,x = 0. \tag{2.7.37}$$

We now solve each equation separately.

(i). The equation (2.7.36) can be easily transformed (by adding 1, then dividing by 2) into

$$\sin\,2x\, = \frac{1}{2}.$$

Using the substitution $2x = y$, this equation reduces to

$$\sin\,y = \frac{1}{2},$$

which has the solutions presented as

$$y = (-1)^k \arcsin\left(\frac{1}{2}\right) + k\pi = (-1)^k\frac{\pi}{6} + k\pi,\ k \text{ integer.}$$

When we go back to our substitution $2x = y$, we get $2y = (-1)^k\dfrac{\pi}{6} + k\pi$, which after dividing by 2 (which is same as multiplying by $\dfrac{1}{2}$), yields

$$x = \frac{1}{2}\left[(-1)^k\frac{\pi}{6} + k\pi\right] = \frac{(-1)^k\pi}{12} + \frac{k\pi}{2},\ k \text{ integer.} \tag{2.7.38}$$

(ii) The equation (2.7.37) has as solutions precisely the x-intercepts of cosine, so its solutions are simply

$$x = \frac{\pi}{2} + k\pi,\ k \text{ integer.} \tag{2.7.39}$$

So our conclusion is as follows. The given equation (2.7.29) has as solutions all the numbers given in the lists (2.7.38) and (2.7.39).

Exercises

In Exercises 1–9 you are asked to find **all solutions** of the given elementary trigonometric equation, using the "clean" solution method. Use **exact values**.

1. $\sin x = \dfrac{\sqrt{3}}{2}$

2. $\sin x = -\dfrac{\sqrt{3}}{2}$

3. $\cos x = \dfrac{\sqrt{2}}{2}$

4. $\cos x = -\dfrac{\sqrt{3}}{2}$

5. $\tan x = -1$

6. $\tan x = \sqrt{3}$

7. $\tan x = -\sqrt{3}$

8. $\tan x = \dfrac{1}{\sqrt{3}}$

9. $\tan x = -\dfrac{1}{\sqrt{3}}$

In Exercises 10–15 use Example 2.4.3 from Section 2.4 as a guideline (with basic solutions replaced by "easy" solutions), to solve the given elementary equation in a specified interval.

10. Given the equation $\sin x = -0.4$, find only the solutions that are in the interval $[-\pi, 3\pi]$. Round to nearest 0.01.

11. Given the equation $\sin x = -\dfrac{1}{2}$, find only the solutions that are in the interval $[-2\pi, 2\pi]$. Use exact values.

12. Given the equation $\cos x = 0.8$, find only the solutions that are in the interval $[-\pi, 3\pi]$. Round to nearest 0.01.

13. Given the equation $\cos x = -\dfrac{\sqrt{2}}{2}$, find only the solutions that are in the interval $[-2\pi, 4\pi]$. Use exact values.

14. Given the equation $\tan x = \sqrt{3}$, find only the solutions that are in the interval $[-\pi, 6\pi]$. Use exact values.

15. Given the equation $\tan x = -3$, find only the solutions that are in the interval $[\pi, 6\pi]$. Round to nearest 0.01.

In Exercises 16–25 you are asked to find **all solution** of the given trigonometric equation. (Use **exact values**.)

16. $\cos\left(\dfrac{\pi}{8} - 4x\right) = 1$

17. $\sin\left(4x + \dfrac{\pi}{10}\right) = \dfrac{1}{2}$

18. $\tan\left(6x+\dfrac{3\pi}{4}\right)=-1$

19. $4\sin^2 z=1$

20. $2\sin^4 t-3\sin^2 t+1=0$

21. $\cos^4 x+\cos^2 x-2=0$

22. $4\sin^2 y=\cos y-1$

23. $\cos^2 w+3\sin w=3$

24. $\sin u\cos u+\sin u-\cos u-1=0$

25. $\tan^4 s=1$.

26*. Rewrite the elementary equation $\cos x=$ *number* as

$$\sin\left(x+\frac{\pi}{2}\right)=\textit{number}.$$

Solve the above equation using the "clean" solution method for the sine equation combined with the substitution $x+\dfrac{\pi}{2}=y$, to come up with an alternate single solution list for the elementary cosine equation, which uses the arcsin function. (Although such an alternate list involves number like $(-1)^k$, it will have the advantage of listing the numbers in *increasing order*.)

27. Solve the equation $3\cos^2 x-10\cos x+3=0$ in the interval $[-\pi,3\pi]$. Round to nearest 0.01.

28. Solve the equation $\tan^2 x=100$ in the interval $[-\pi,\pi]$. Round to nearest 0.01.

29. Solve the equation $4\sin^2 x+7\sin x-2=0$ in the interval $[-3\pi,3\pi]$. Round to nearest 0.01.

Chapter 3

Applications of Trigonometry in Geometry

In this Chapter we discuss several applications of Trigonometry in Geometry, the most important of which are concerned with *triangle solving problems*, as discussed in Section 3.2.

3.1 Applications to Vector Geometry

In this section we explore several aspects of Vector Geometry in which Trigonometry plays a key role.

Vector Direction Revisited

When we first introduced *vectors*, in Section 2.1, we mentioned that, with one exception (the *zero* vector), each vector is completely characterized by its *direction* and *magnitude*. However, we were a bit imprecise about what *direction* really meant. What we chose to do in Section 2.1 was to think *directions* as corresponding to *rays on a compass*. However, if we decide to fix our compass to be the *unit circle*, then we can think of a *direction* as being nothing else but a **unit vector**, that is, a *vector of magnitude equal to* 1. With this interpretation in mind, we have the following statement, which we also use as a definition.

Unit Direction Vectors

Every *non-zero* vector $\vec{\mathbf{v}}$ can be presented *uniquely* as

$$\vec{\mathbf{v}} = r\,\vec{\mathbf{u}}, \tag{3.1.1}$$

with $r > 0$ and $\left\|\vec{\mathbf{u}}\right\| = 1$. The unique unit vector $\vec{\mathbf{u}}$ from this presentation is called the **unit direction vector of** $\vec{\mathbf{v}}$. The main features of the presentation (3.1.1) are as follows.

(A) The number r from (3.1.1) is: $r = \left\|\vec{\mathbf{v}}\right\|$ (the magnitude of $\vec{\mathbf{v}}$). The unit direction vector $\vec{\mathbf{u}}$ can be reconstructed out of $\vec{\mathbf{v}}$ by the identity

$$\vec{\mathbf{u}} = \left(\frac{1}{\left\|\vec{\mathbf{v}}\right\|}\right)\vec{\mathbf{v}}.$$

(B) Two vectors, presented as above, in the form $\vec{\mathbf{v}}_1 = r_1\,\vec{\mathbf{u}}_1$ and $\vec{\mathbf{v}}_2 = r_2\,\vec{\mathbf{u}}_2$, have:

- (i) *same direction*, if and only if *their unit direction vectors coincide:* $\vec{\mathbf{u}}_1 = \vec{\mathbf{u}}_2$;
- (ii) *opposite directions*, if and only if *their unit direction vectors are opposites of each other:* $\vec{\mathbf{u}}_1 + \vec{\mathbf{u}}_2 = \vec{\mathbf{0}}$.

Example 3.1.1. Consider the vectors $\vec{\mathbf{v}}_1 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$, $\vec{\mathbf{v}}_2 = \begin{bmatrix} 6 \\ -8 \end{bmatrix}$ and $\vec{\mathbf{v}}_3 = \begin{bmatrix} -12 \\ 16 \end{bmatrix}$. A quick calculation of magnitudes gives

$$\begin{aligned}
\|\vec{\mathbf{v}}_1\| &= \sqrt{3^2+(-4)^2} = \sqrt{9+16} = \sqrt{25} = 5,\\
\|\vec{\mathbf{v}}_2\| &= \sqrt{6^2+(-8)^2} = \sqrt{36+64} = \sqrt{100} = 10,\\
\|\vec{\mathbf{v}}_2\| &= \sqrt{(-12)^2+16^2} = \sqrt{144+256} = \sqrt{400} = 20,
\end{aligned}$$

so the unit direction vectors are

$$\begin{aligned}
\vec{\mathbf{u}}_1 &= \left(\frac{1}{\|\vec{\mathbf{v}}_1\|}\right)\vec{\mathbf{v}}_1 = \frac{1}{5}\begin{bmatrix}3\\-4\end{bmatrix} = \begin{bmatrix}3/5\\-4/5\end{bmatrix},\\
\vec{\mathbf{u}}_2 &= \left(\frac{1}{\|\vec{\mathbf{v}}_2\|}\right)\vec{\mathbf{v}}_2 = \frac{1}{10}\begin{bmatrix}6\\-8\end{bmatrix} = \begin{bmatrix}6/10\\-8/10\end{bmatrix} = \begin{bmatrix}3/5\\-4/5\end{bmatrix} = \vec{\mathbf{u}}_1,\\
\vec{\mathbf{u}}_3 &= \left(\frac{1}{\|\vec{\mathbf{v}}_3\|}\right)\vec{\mathbf{v}}_3 = \frac{1}{20}\begin{bmatrix}-12\\16\end{bmatrix} = \begin{bmatrix}-12/20\\16/20\end{bmatrix} = \begin{bmatrix}-3/5\\4/5\end{bmatrix} = -\vec{\mathbf{u}}_1,
\end{aligned}$$

so we can say that $\vec{\mathbf{v}}_2$ has *same direction* as $\vec{\mathbf{v}}_1$, while $\vec{\mathbf{v}}_3$ has *direction opposite* to the direction of $\vec{\mathbf{v}}_1$.

The Dot Product of Two Vectors

Given two vectors written in coordinates $\vec{\mathbf{v}}_1 = \begin{bmatrix}x_1\\y_1\end{bmatrix}$, $\vec{\mathbf{v}}_2 = \begin{bmatrix}x_2\\y_2\end{bmatrix}$, we define their **dot product** to be the quantity

$$\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 = x_1x_2 + y_1y_2. \tag{3.1.2}$$

Straight from the definition it is pretty obvious that this operation, which combines *two vectors* to produce a *number*, has the following features.

Properties of Dot Product

I. *Symmetry:* $\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_2 \bullet \vec{\mathbf{v}}_1$.
II. *Distributivity in first variable:* $(\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2)\bullet\vec{\mathbf{v}}_3 = \vec{\mathbf{v}}_1\bullet\vec{\mathbf{v}}_3 + \vec{\mathbf{v}}_2\bullet\vec{\mathbf{v}}_3$.
III. *Distributivity in second variable:* $\vec{\mathbf{v}}_1\bullet(\vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_3) = \vec{\mathbf{v}}_1\bullet\vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_1\bullet\vec{\mathbf{v}}_3$.
IV. *Homogeneity in first variable:* $(t\vec{\mathbf{v}}_1)\bullet\vec{\mathbf{v}}_2 = t(\vec{\mathbf{v}}_1\bullet\vec{\mathbf{v}}_2)$.
IV. *Homogeneity in second variable:* $\vec{\mathbf{v}}_1\bullet(t\vec{\mathbf{v}}_2) = t(\vec{\mathbf{v}}_1\bullet\vec{\mathbf{v}}_2)$.
V. *Magnitude Identity:* $\vec{\mathbf{v}}\bullet\vec{\mathbf{v}} = \|\vec{\mathbf{v}}\|^2$.

As suggested by the identity V, dot products are intimately related to magnitudes. This relationship can be deepened to yield the following important consequence.

Pythagoras' Generalized Theorem

For any two vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$, the magnitudes of their difference and sum satisfy:

$$\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\|^2 = \left\|\vec{\mathbf{v}}_1\right\|^2 + \left\|\vec{\mathbf{v}}_2\right\|^2 - 2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right); \tag{3.1.3}$$

$$\left\|\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2\right\|^2 = \left\|\vec{\mathbf{v}}_1\right\|^2 + \left\|\vec{\mathbf{v}}_2\right\|^2 + 2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right). \tag{3.1.4}$$

Proof. Both identities follow from the properties of the dot product. For example, to prove (3.1.3), we start off with the left-hand side, which we replace using the Magnitude Identity:

$$\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\|^2 = \left(\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right) \bullet \left(\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right),$$

and then using distributivity and symmetry we can write

$$\begin{aligned}\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\|^2 &= \left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_1\right) - \left(\vec{\mathbf{v}}_2 \bullet \vec{\mathbf{v}}_1\right) - \left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right) + \left(\vec{\mathbf{v}}_2 \bullet \vec{\mathbf{v}}_2\right) = \\ &= \left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_1\right) + \left(\vec{\mathbf{v}}_2 \bullet \vec{\mathbf{v}}_2\right) - 2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right).\end{aligned}$$

The identity (3.1.3) now follows immediately by replacing $\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_1\right) = \left\|\vec{\mathbf{v}}_1\right\|^2$ and $\left(\vec{\mathbf{v}}_2 \bullet \vec{\mathbf{v}}_2\right) = \left\|\vec{\mathbf{v}}_2\right\|^2$. The identity (3.1.4) is proved the exact same way. □

Example 3.1.2. Suppose we are given two vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$, of which we know that $\left\|\vec{\mathbf{v}}_1\right\| = 5$, $\left\|\vec{\mathbf{v}}_2\right\| = 7$, and $\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\| = 10$, and we want to compute their dot product, as well as the magnitude of their sum.

After we replace three of the known quantities in (3.1.3), we get

$$10^2 = 5^2 + 7^2 - 2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right),$$

which yields

$$-2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right) = 10^2 - 5^2 - 7^2 = 100 - 25 - 49 = 26,$$

so we immediately get $\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 = -\dfrac{26}{2} = -13.$

Now we can use (3.1.4) and get

$$\left\|\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2\right\|^2 = 5^2 + 7^2 + 2 \cdot (-13) = 25 + 49 - 26 = 48,$$

which gives:[11] $\left\|\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2\right\| = \sqrt{48} = 4\sqrt{3}$.

The Skew Product of Two Vectors

Given two vectors written in coordinates $\vec{\mathbf{v}}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, $\vec{\mathbf{v}}_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, we define their **skew product** to be the quantity

$$\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = x_1 y_2 - x_2 y_1. \tag{3.1.5}$$

Straight from the definition it is pretty obvious that this operation, which again combines *two vectors* to produce a *number*, has the following features.

[11] Of course, when we look at the equation $?^2 = 48$, it will always have two solutions: $? = \pm\sqrt{48}$. However, since magnitudes of vectors are always ≥ 0, we will only retain the positive solution.

Properties of Skew Product

I. *Anti-Symmetry:* $\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = -\vec{\mathbf{v}}_2 \wedge \vec{\mathbf{v}}_1$.
II. *Distributivity in first variable:* $(\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2) \wedge \vec{\mathbf{v}}_3 = \vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_3 + \vec{\mathbf{v}}_2 \wedge \vec{\mathbf{v}}_3$.
III. *Distributivity in second variable:* $\vec{\mathbf{v}}_1 \wedge (\vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_3) = \vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_3$.
IV. *Homogeneity in first variable:* $(t\vec{\mathbf{v}}_1) \wedge \vec{\mathbf{v}}_2 = t(\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2)$.
IV. *Homogeneity in second variable:* $\vec{\mathbf{v}}_1 \wedge (t\vec{\mathbf{v}}_2) = t(\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2)$.

When compared to the dot product, the skew product is a bit unusual, because of anti-symmetry, which yields

$$\vec{\mathbf{v}} \wedge \vec{\mathbf{v}} = 0.$$

Therefore, there is no direct way to relate skew products to magnitudes.

Trigonometric Forms of Dot and Skew Products

Up to this point, Trigonometry has not played any role in our discussion, but now it will become a key player in our story, which begins with the following definition.

The **geometric angle** formed by two *non-zero* vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ is the geometric angle $\angle VOW$ which is constructed as follows:

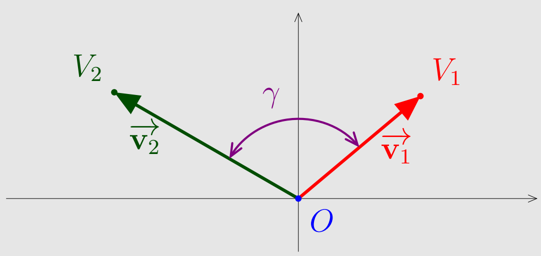


Figure 3.1.1

- O is the *origin* in the coordinate plane;
- V_1 is the (unique) point that has $\vec{\mathbf{v}}_1$ as its position vector with respect to the origin, that is: $\overrightarrow{OV_1} = \vec{\mathbf{v}}_1$;
- V_2 is the (unique) point that has $\vec{\mathbf{v}}_2$ as its position vector with respect to the origin, that is: $\overrightarrow{OV_2} = \vec{\mathbf{v}}_2$.

As usual, we identify this angle with its radian measure, we get a number γ in the interval $[0, \pi]$. (When using degrees, the values will range from $0°$ to $180°$.)

As it turns out, the geometric angle formed by two vectors is that *it only depends on the directions of the two vectors*. In other words, if we replace each vector by its *unit direction vector*, the geometric angle is *the same*.

Besides geometric angles between vectors, one can consider a variant which accounts for *orientation*. To understand how these new angles come about, all we have to do is to consider the *rotation angle*, which has one vector sitting on its initial side, and the other vector sitting on its terminal side.

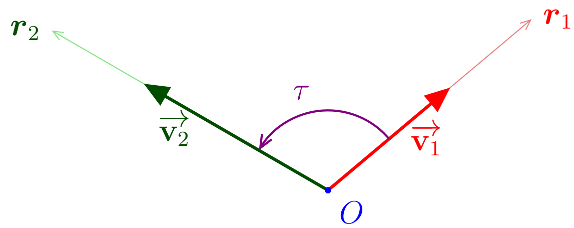


Figure 3.1.2

For example, Figure 3.1.2 depicts one such rotation angle, where $\overrightarrow{\mathbf{v}_1}$ is the *first*, also called the **initial** vector, while $\overrightarrow{\mathbf{v}_2}$ is *second*, also called the **terminal** vector. In order to make this rotation angle *uniquely* determined, we have to specify its measure τ, which by convention we chose to be in the interval $(-\pi, \pi]$. Once this measure is selected, the resulting rotation angle is referred to as **turning angle of $\overrightarrow{\mathbf{v}_1}$ over $\overrightarrow{\mathbf{v}_2}$**. As was the case with geometric angles, we will identify our turning angle with its radian measure τ.

The turning angle constructed *depends on the order* in which we specify the two vectors.

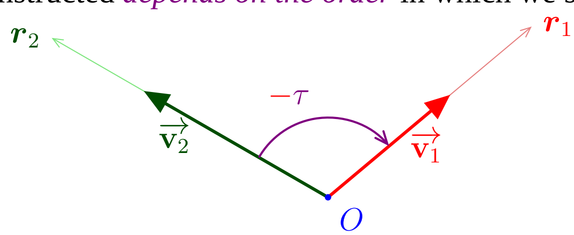


Figure 3.1.3

More precisely, if we switch the order, then as seen by comparing Figures 3.1.2 and 3.1.3 above, the following change takes place.

> Assume two *non-zero* vectors $\overrightarrow{\mathbf{v}_1}$ and $\overrightarrow{\mathbf{v}_2}$ are given. If the turning angle of $\overrightarrow{\mathbf{v}_1}$ over $\overrightarrow{\mathbf{v}_2}$ is τ, then the turning angle of $\overrightarrow{\mathbf{v}_2}$ over $\overrightarrow{\mathbf{v}_1}$ is $-\tau$.

The turning angle is closely related to the geometric angle, as explained below.

Turning Angle Sign Rule

Assume $\overrightarrow{\mathbf{v}_1}$ and $\overrightarrow{\mathbf{v}_2}$ are two *non-zero* vectors, and the geometric angle between them is γ. Then the turning angle τ of $\overrightarrow{\mathbf{v}_1}$ over $\overrightarrow{\mathbf{v}_2}$ is given as follows:

I. If $\overrightarrow{\mathbf{v}_1}$ can be rotated over $\overrightarrow{\mathbf{v}_2}$ by a γ-turn in the *counterclockwise direction*, then $\tau = \gamma$.

II. If $\overrightarrow{\mathbf{v}_1}$ can be rotated over $\overrightarrow{\mathbf{v}_2}$ by a γ-turn in the *clockwise direction*, then $\tau = -\gamma$.

In either case, one has the equality: $\gamma = |\tau|$.

There is yet one more way to look at turning angles, which begins with the observation that (exactly as was the case with *geometric* angles) *turning angles depend on the directions of the initial and terminal vectors*. In other words, if we replace each vector by its *unit direction vector*, the turning angle is *the same*. The second observation is that, if we start with two *unit* vectors $\overrightarrow{\mathbf{u}_1}$ (the initial vector) and $\overrightarrow{\mathbf{u}_2}$, then the turning angle τ of $\overrightarrow{\mathbf{u}_1}$ over $\overrightarrow{\mathbf{u}_2}$ has the property that the associated τ-rotation (about the origin) *transforms* $\overrightarrow{\mathbf{u}_1}$ *into* $\overrightarrow{\mathbf{u}_2}$. Putting those two observations together, we get the following characterization of the turning angle.

Given two *non-zero* vectors $\vec{\mathbf{v}_1}$ and $\vec{\mathbf{v}_2}$, with respective unit direction vectors $\vec{\mathbf{u}_1}$ and $\vec{\mathbf{u}_2}$, the turning angle τ of $\vec{\mathbf{v}_1}$ over $\vec{\mathbf{v}_2}$ is the *the unique number in the interval* $(-\pi, \pi]$, that satisfies the matrix product identity:

$$\vec{\mathbf{u}_2} = \mathbf{R}_\tau \vec{\mathbf{u}_1}, \tag{3.1.6}$$

where $\mathbf{R}_\tau$ is the *τ-rotation matrix*, given by

$$\mathbf{R}_\tau = \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix}.$$

With all these preparations, we can now state the following.

Dot/Skew Product Theorem

Let $\vec{\mathbf{v}_1}$ and $\vec{\mathbf{v}_2}$ be two vectors in the plane.

I. *Assuming both vectors are non-zero, if the turning angle of $\vec{\mathbf{v}_1}$ over $\vec{\mathbf{v}_2}$ is τ, and the geometric angle between $\vec{\mathbf{v}_1}$ and $\vec{\mathbf{v}_2}$ is γ then the dot and the skew products are given by:*

$$\vec{\mathbf{v}_1} \bullet \vec{\mathbf{v}_2} = \left\|\vec{\mathbf{v}_1}\right\| \cdot \left\|\vec{\mathbf{v}_2}\right\| \cdot \cos\tau = \left\|\vec{\mathbf{v}_1}\right\| \cdot \left\|\vec{\mathbf{v}_2}\right\| \cdot \cos\gamma \tag{3.1.7}$$
$$\vec{\mathbf{v}_1} \wedge \vec{\mathbf{v}_2} = \left\|\vec{\mathbf{v}_1}\right\| \cdot \left\|\vec{\mathbf{v}_2}\right\| \cdot \sin\tau \tag{3.1.8}$$

II. *If either one of the vectors is zero, then:*

$$\vec{\mathbf{v}_1} \bullet \vec{\mathbf{v}_2} = \vec{\mathbf{v}_1} \wedge \vec{\mathbf{v}_2} = 0$$

Proof. Case II is pretty clear right from the definition. In Case I, we write $\vec{\mathbf{v}_1} = r_1 \vec{\mathbf{u}_1}$ and $\vec{\mathbf{v}_2} = r_2 \vec{\mathbf{u}_2}$, using the unit direction vectors, so by homgeneity we can write

$$\vec{\mathbf{v}_1} \bullet \vec{\mathbf{v}_2} = r_1 r_2 \left(\vec{\mathbf{u}_1} \bullet \vec{\mathbf{u}_2}\right); \tag{3.1.9}$$
$$\vec{\mathbf{v}_1} \wedge \vec{\mathbf{v}_2} = r_1 r_2 \left(\vec{\mathbf{u}_1} \wedge \vec{\mathbf{u}_2}\right). \tag{3.1.10}$$

If we write our two unit vectors in coordinates as $\vec{\mathbf{u}_1} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{\mathbf{u}_2} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, then using the above characterization of the turning angle τ, we have

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix},$$

which after multiplying the matrices gives the equalities

$$\begin{cases} x_2 &= x_1 \cos\tau - y_1 \sin\tau \\ y_2 &= x_1 \sin\tau + y_1 \cos\tau \end{cases}$$

Using these equalities, we now can express the dot and skew products of our two unit vectors solely in terms of x_1, y_1 and τ:

$$\begin{aligned} \vec{\mathbf{u}_1} \bullet \vec{\mathbf{u}_2} &= x_1 x_2 + y_1 y_2 = x_1 \left(x_1 \cos\tau - y_1 \sin\tau\right) + y_1 \left(x_1 \sin\tau + y_1 \cos\tau\right) = \\ &= x_1^2 \cos\tau - x_1 y_1 \sin\tau + x_1 y_1 \sin\tau + y_1^2 \cos\tau = \left(x_1^2 + y_1^2\right) \cos\tau; \\ \vec{\mathbf{u}_1} \wedge \vec{\mathbf{u}_2} &= x_1 y_2 - y_1 x_2 = x_1 \left(x_1 \sin\tau + y_1 \cos\tau\right) - y_1 \left(x_1 \cos\tau - y_1 \sin\tau\right) = \\ &= x_1^2 \sin\tau + x_1 y_1 \cos\tau - x_1 y_1 \cos\tau + y_1^2 \sin\tau = \left(x_1^2 + y_1^2\right) \sin\tau. \end{aligned}$$

Using the fact that $\overrightarrow{\mathbf{u}_1}$ is a unit vector, we know $x_1^2+y_1^2 = \left\|\overrightarrow{\mathbf{u}_1}\right\|^2 = 1$, so the above two calculations simply give

$$\overrightarrow{\mathbf{u}_1}\bullet\overrightarrow{\mathbf{u}_2} = \cos\tau,$$
$$\overrightarrow{\mathbf{u}_1}\wedge\overrightarrow{\mathbf{u}_2} = \sin\tau,$$

so when we go back to (3.1.9) and (3.1.10), we now have

$$\overrightarrow{\mathbf{v}_1}\bullet\overrightarrow{\mathbf{v}_2} = r_1r_2\left(\overrightarrow{\mathbf{u}_1}\bullet\overrightarrow{\mathbf{u}_2}\right) = r_1r_2\cos\tau = \left\|\overrightarrow{\mathbf{v}_1}\right\|\cdot\left\|\overrightarrow{\mathbf{v}_2}\right\|\cdot\cos\tau, \tag{3.1.11}$$
$$\overrightarrow{\mathbf{v}_1}\wedge\overrightarrow{\mathbf{v}_2} = r_1r_2\left(\overrightarrow{\mathbf{u}_1}\wedge\overrightarrow{\mathbf{u}_2}\right) = r_1r_2\sin\tau = \left\|\overrightarrow{\mathbf{v}_1}\right\|\cdot\left\|\overrightarrow{\mathbf{v}_2}\right\|\cdot\sin\tau \tag{3.1.12}$$

The second equality in (3.1.7) follows immediately from the first one, since we know that $\tau = \pm\gamma$, so regardless of the sign (since cosine is an even function), we have $\cos\tau = \cos\gamma$. □

Example 3.1.3. Suppose two vehicles started their trips in the desert from the same point, and drove on two straight lines as follows: the first vehicle drove 70 miles in the N$55°$E direction, while the second vehicle drove 80 miles in the S$65°$E direction. We wish to compute the distance between the two vehicles at the end of their respective trips.

The trips of both vehicles can be completely described using two vectors as, depicted as in the Figure below.

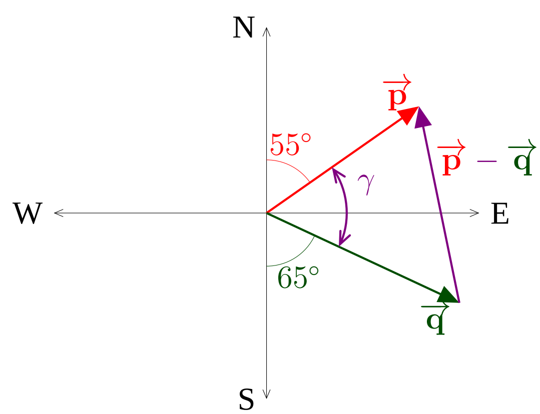


Figure 3.1.4

The vector $\overrightarrow{\mathbf{p}}$ shown above indicates the position of the first vehicle, so (if we agree that our length unit is the mile) it has magnitude $\left\|\overrightarrow{\mathbf{p}}\right\| = 70$. The second vector $\overrightarrow{\mathbf{q}}$ shown above indicates the position of the second vehicle, so it has magnitude $\left\|\overrightarrow{\mathbf{q}}\right\| = 80$. With these preparations, the distance we need to compute is precisely the magnitude $\left\|\overrightarrow{\mathbf{p}} - \overrightarrow{\mathbf{q}}\right\|$.

Although is it possible to compute this distance using coordinates (as seen for instance in Example 2.1.3 from Section 2.1.), it will be a lot simpler if we use the dot product. From Figure 3.1.4 we can also read off the geometric angle γ formed by our two vectors, because we clearly have $55° + \gamma + 65° = 180°$, which yields:

$$\gamma = 180° - 55° - 65° = 60°.$$

Using the Dot/Skew Product Theorem, it follows that

$$\overrightarrow{\mathbf{p}}\bullet\overrightarrow{\mathbf{q}} = \left\|\overrightarrow{\mathbf{p}}\right\|\cdot\left\|\overrightarrow{\mathbf{q}}\right\|\cdot\cos 60° = 70\cdot 80\cdot\frac{1}{2} = 2800. \tag{3.1.13}$$

With this calculation in mind, we can now use Pythagoras' Generalized Theorem to conclude that

$$\left\|\vec{\mathbf{p}} - \vec{\mathbf{q}}\right\|^2 = \left\|\vec{\mathbf{p}}\right\|^2 + \left\|\vec{\mathbf{q}}\right\|^2 - 2\,\vec{\mathbf{p}} \bullet \vec{\mathbf{q}} = 70^2 + 80^2 - 2 \cdot 2800 = 5700,$$

and then (since magnitudes are non-negative), the desired distance is:

$$\left\|\vec{\mathbf{p}} - \vec{\mathbf{q}}\right\| = \sqrt{5700} = 10\sqrt{57} \simeq 75.49834435 \text{ miles}.$$

The Dot/Skew Product Theorem is very useful, because it allows one to compute the angles formed by two vectors, as summarized below.

Vector Angle Formulas

Assume $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are two *non-zero* vectors. The *turning* angle τ of $\vec{\mathbf{v}}_1$ over $\vec{\mathbf{v}}_2$ is the *one and only one angle in the interval* $(-\pi, \pi]$ that satisfies the equalities

$$\begin{cases} \cos\tau & = \dfrac{\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2}{\left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\|} \\ \sin\tau & = \dfrac{\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2}{\left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\|} \end{cases} \tag{3.1.14}$$

In particular, the two angles (geometric and turning) formed by the two vectors can be computed as follows.

I. The *geometric* angle γ between $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$, is given by the formula:

$$\gamma = \arccos\left(\frac{\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2}{\left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\|}\right). \tag{3.1.15}$$

II. In terms of γ, the *turning* angle τ of $\vec{\mathbf{v}}_1$ over $\vec{\mathbf{v}}_2$ is given as

$$\tau = \left(\text{sign of } \vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2\right)\gamma \tag{3.1.16}$$

(We agree that, when $\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = 0$, the above sign is $+$.)

CLARIFICATIONS. The formulas (3.1.14) follow from the Dot/Skew product Theorem, which also gives us

$$\cos\gamma = \frac{\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2}{\left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\|}. \tag{3.1.17}$$

However, since by construction γ belongs to the interval $[0, \pi]$, which is the *range* of arccos, by the inversion formula we have the equality $\gamma = \arccos(\cos\gamma)$, and then formula (3.1.15) follows from (3.1.17).

Example 3.1.4. Suppose we have the vectors $\vec{\mathbf{v}}_1 = \begin{bmatrix} -7 \\ -1 \end{bmatrix}$ and $\vec{\mathbf{v}}_2 = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$, and we want to find the geometric angle γ between them, as well as the turning angle τ of $\vec{\mathbf{v}}_1$ over $\vec{\mathbf{v}}_2$.

We start off by computing magnitudes and the two products:

$$\begin{aligned} \left\|\vec{\mathbf{v}}_1\right\| &= \sqrt{(-7)^2 + (-1)^2} = \sqrt{1 + 49} = \sqrt{50} = 5\sqrt{2}; \\ \left\|\vec{\mathbf{v}}_2\right\| &= \sqrt{6^2 + 8^2} = \sqrt{36 + 64} = \sqrt{100} = 10; \\ \vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 &= (-7) \cdot 6 + (-1) \cdot 8 = -42 - 8 = -50; \\ \vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 &= (-7) \cdot 8 - (-1) \cdot 6 = -56 + 6 = -50. \end{aligned}$$

Using the formula (3.1.15), the geometric angle is:

$$\gamma = \arccos\left(\frac{-50}{5\sqrt{2}\cdot 10}\right) = \arccos\left(-\frac{1}{\sqrt{2}}\right) = \frac{3\pi}{4}.$$

Since the skew product $\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2$ is *negative*, using (3.1.16), the turning angle of $\vec{\mathbf{v}}_1$ over $\vec{\mathbf{v}}_2$ is:

$$\tau = -\frac{3\pi}{4}.$$

Orthogonality

The Vector Angle Formulas can be efficiently applied to questions concerning perpendicularity (a.k.a. *orthogonality*) or parallelism, as illustrated below.

Orthogonality and Parallelism Tests

I. Two non-zero vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are *perpendicular*, if and only if:

$$\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 = 0. \tag{3.1.18}$$

II. Two non-zero vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are *parallel*, meaning that they either have *same direction*, or *opposite directions*, if and only if:

$$\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = 0. \tag{3.1.19}$$

CLARIFICATIONS. Concerning statement II, one can in fact be a bit more precise. If condition (3.1.19) holds, which is the same as saying that the geometric angle γ formed by the two vectors is either 0 or π, then $\cos\gamma = \pm 1$, so the dot product will be

$$\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 = \pm\left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\|,$$

so we can simply say that

II-a. $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ have *same direction*, if and only if:

$$\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = 0 \text{ and } \vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 > 0 \tag{3.1.20}$$

II-b. $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ have *opposite directions*, if and only if:

$$\vec{\mathbf{v}}_1 \wedge \vec{\mathbf{v}}_2 = 0 \text{ and } \vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2 < 0 \tag{3.1.21}$$

An important application of the Orthogonality Test is contained in the following important statement. (The proof is outlined in Exercise 21.)

Vector Component Theorem

If $\vec{w}$ is a non-zero vector, then any vector $\vec{v}$ can be written uniquely as a sum

$$\vec{v} = \vec{p} + \vec{n},$$

with:
(i) $\vec{p}$ either *zero*, or *parallel* to $\vec{w}$;
(ii) $\vec{n}$ either *zero*, or *perpendicular* to $\vec{w}$.
Additionally, the vectors $\vec{p}$ and $\vec{n}$ have the following properties.
(A) The vector $\vec{p}$ is given by:

$$\vec{p} = \left(\frac{\vec{v} \bullet \vec{w}}{\vec{w} \bullet \vec{w}}\right)\vec{w} = \left(\frac{\vec{v} \bullet \vec{w}}{\|\vec{w}\|^2}\right)\vec{w}. \tag{3.1.22}$$

(B) If $\vec{v}$ is non-zero, and γ is the geometric angle between $\vec{v}$ and $\vec{w}$, then the vector $\vec{p}$ can also be presented as:

$$\vec{p} = \left(\frac{\|\vec{v}\| \cdot \cos\gamma}{\|\vec{w}\|}\right)\vec{w}. \tag{3.1.23}$$

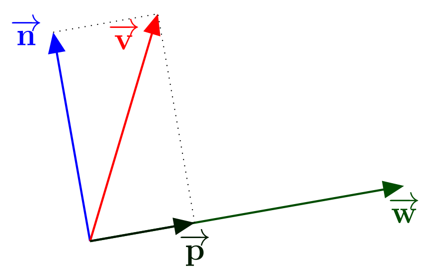


Figure 3.1.5

TERMINOLOGY. The vectors $\vec{p}$ and $\vec{n}$ are referred to as the *components of $\vec{v}$ relative to $\vec{w}$*. More precisely:
(i) the vector $\vec{p}$ is called the *component of $\vec{v}$ along (or parallel to) $\vec{w}$*; this same vector is also referred to as the *projection of $\vec{v}$ on the direction of $\vec{w}$*, and is denoted by $\mathbf{proj}_{\vec{w}}(\vec{v})$;
(ii) the vector $\vec{n} = \vec{v} - \mathbf{proj}_{\vec{w}}(\vec{v})$ is called the *component of $\vec{v}$ which is perpendicular (or normal) to $\vec{w}$*.

Areas of Parallelograms and Triangles

We now put everything we learned to very good use, by deriving a very useful formula for the area of a parallelogram.

Parallelogram Area in Vector Form

If $\vec{v}$ and $\vec{w}$ are two *non-zero, non-parallel* vectors, then the parallelogram $\mathscr{P}$ formed by them has:

$$\text{Area}(\mathscr{P}) = \left|\vec{v} \wedge \vec{w}\right|. \tag{3.1.24}$$

In particular, if γ is the geometric angle formed by these two vectors, we also have the equality:

$$\text{Area}(\mathscr{P}) = \|\overrightarrow{\mathbf{v}}\| \cdot \|\overrightarrow{\mathbf{w}}\| \cdot \sin\gamma. \tag{3.1.25}$$

Proof. When we apply the Vector Component Theorem, we can write

$$\overrightarrow{\mathbf{v}} = \overrightarrow{\mathbf{p}} + \overrightarrow{\mathbf{n}},$$

with $\overrightarrow{\mathbf{p}}$ parallel to $\overrightarrow{\mathbf{w}}$, and $\overrightarrow{\mathbf{n}}$ perpendicular to $\overrightarrow{\mathbf{w}}$.

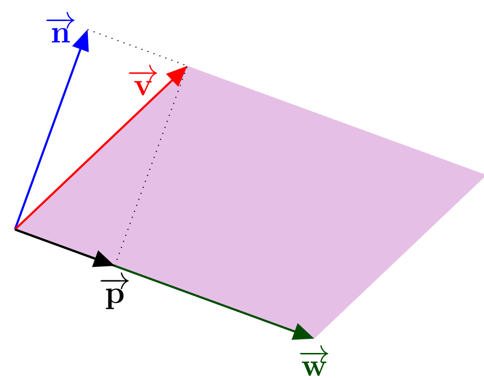


Figure 3.1.6

It is pretty obvious that, when we set up our parallelogram to have $\overrightarrow{\mathbf{w}}$ as its base, its height will be exactly $\overrightarrow{\mathbf{n}}$, so we will have

$$\text{Area} = \|\overrightarrow{\mathbf{n}}\| \cdot \|\overrightarrow{\mathbf{w}}\|. \tag{3.1.26}$$

On the one hand, since $\overrightarrow{\mathbf{n}}$ perpendicular to $\overrightarrow{\mathbf{w}}$. it is clear that

$$\overrightarrow{\mathbf{n}} \wedge \overrightarrow{\mathbf{w}} = \|\overrightarrow{\mathbf{n}}\| \cdot \|\overrightarrow{\mathbf{w}}\| \cdot \sin\left(\pm\frac{\pi}{2}\right) = \pm\|\overrightarrow{\mathbf{n}}\| \cdot \|\overrightarrow{\mathbf{w}}\|,$$

so using (3.1.26) we can now write

$$\text{Area} = \left|\overrightarrow{\mathbf{n}} \wedge \overrightarrow{\mathbf{w}}\right|. \tag{3.1.27}$$

On the other hand, since $\overrightarrow{\mathbf{p}}$ parallel to $\overrightarrow{\mathbf{w}}$, we have $\overrightarrow{\mathbf{p}} \wedge \overrightarrow{\mathbf{w}} = 0$, so by distributivity of the skew product we have

$$\overrightarrow{\mathbf{v}} \wedge \overrightarrow{\mathbf{w}} = (\overrightarrow{\mathbf{p}} + \overrightarrow{\mathbf{n}}) \wedge \overrightarrow{\mathbf{w}} = \overrightarrow{\mathbf{p}} \wedge \overrightarrow{\mathbf{w}} + \overrightarrow{\mathbf{n}} \wedge \overrightarrow{\mathbf{w}} = \overrightarrow{\mathbf{n}} \wedge \overrightarrow{\mathbf{w}},$$

so in (3.1.27) we can substitute the right-hand side with $\overrightarrow{\mathbf{v}} \wedge \overrightarrow{\mathbf{w}}$, which gives us precisely the desired formula (3.1.24).

As for the second formula, all we have to remember is the fact that, by the Dot/Skew Product Theorem, we know that $\overrightarrow{\mathbf{v}} \wedge \overrightarrow{\mathbf{w}} = \|\overrightarrow{\mathbf{v}}\| \cdot \|\overrightarrow{\mathbf{w}}\| \cdot \sin\tau$, where τ is the *turning* angle of $\overrightarrow{\mathbf{v}}$ over $\overrightarrow{\mathbf{w}}$. Since we also know that $\tau = \pm\gamma$, we always have $\sin\tau = \pm\sin\gamma$, thus

$$\overrightarrow{\mathbf{v}} \wedge \overrightarrow{\mathbf{w}} = \pm\|\overrightarrow{\mathbf{v}}\| \cdot \|\overrightarrow{\mathbf{w}}\| \cdot \sin\gamma,$$

and then by taking absolute values,[12] the $\pm$ sign goes away, thus yielding (3.1.25). $\square$

Example 3.1.5. Suppose we have vectors $\vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} -2 \\ -8 \end{bmatrix}$, and we want to find the area of the parallelogram formed by these vectors. All we have to do is to compute the skew product:

$$\vec{v} \wedge \vec{w} = 3 \cdot (-8) - (-2) \cdot 2 = -24 + 4 = -20,$$

and then we conclude that the area is $|-20| = 20$.

Using the parallelogram area formulas, we immediately obtain the area formulas for *triangles*.

Triangle Area in Vector Form

If $\vec{v}$ and $\vec{w}$ are two *non-zero, non-parallel* vectors, then both triangles $\mathscr{T}$ and $\mathscr{T}'$ that can be formed by them have equal areas:

$$\text{Area}(\mathscr{T}) = \text{Area}(\mathscr{T}') = \tfrac{1}{2}\left|\vec{v} \wedge \vec{w}\right|. \tag{3.1.28}$$

In particular, if γ is the geometric angle formed by these two vectors, then we also have the equalities:

$$\text{Area}(\mathscr{T}) = \text{Area}(\mathscr{T}') = \tfrac{1}{2} \cdot \|\vec{v}\| \cdot \|\vec{w}\| \cdot \sin\gamma. \tag{3.1.29}$$

Proof. The two possible triangles that can be formed by the two vectors are depicted below.

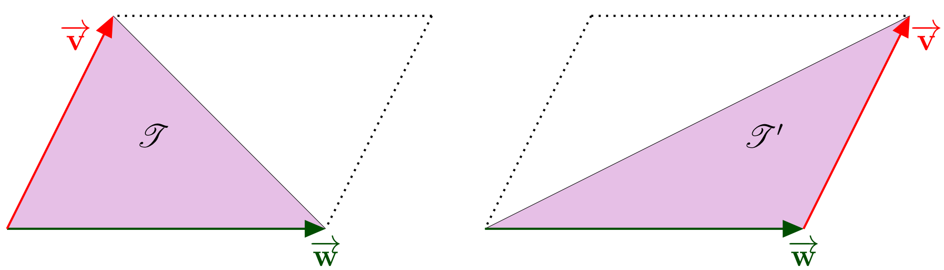


Figure 3.1.7

No matter what triangle we choose ($\mathscr{T}$ or $\mathscr{T}'$), it is pretty clear that the parallelogram $\mathscr{P}$ formed by the two vectors consists of two congruent copies of the chosen triangle, so both triangles will have areas

$$\text{Area}(\mathscr{T}) = \text{Area}(\mathscr{T}') = \tfrac{1}{2} \cdot \text{Area}(\mathscr{P}),$$

and then everything follows from (3.1.24) and (3.1.25). $\square$

CLARIFICATIONS. Additional area formulas are provided in Exercises 23-24

Exercises

In Exercises 1-4 you are asked to compute the dot and skew product of the given vectors.

1. $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

[12] Since γ is in the interval $[0, \pi]$, we always know that $\sin\gamma \geq 0$.

2. $\vec{v} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

3. $\vec{v} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$.

4. $\vec{v} = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

In Exercises 5-7 you have to use the given information to compute the exact value of the dot product. $\vec{v} \bullet \vec{w}$

5. $\|\vec{v}\| = 3$, $\|\vec{w}\| = 5$, $\|\vec{v} + \vec{w}\| = 6$.

6. $\|\vec{v}\| = 10$, $\|\vec{w}\| = 2$, $\|\vec{v} - \vec{w}\| = 9$.

7. $\|\vec{v} + \vec{w}\| = 5$, $\|\vec{v} - \vec{w}\| = 7$.

In Exercises 8-11 you are asked to find the geometric angle and the turning angle of each of the two vectors over the other one.

8. $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Use exact values.

9. $\vec{v} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Use exact values.

10. $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Round to nearest 0.01 of a radian.

11. $\vec{v} = \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} -\sqrt{3} \\ 1 \end{bmatrix}$. Use exact values.

In Exercises 12-15 you are asked to compute the projection of $\vec{v}$ on the direction of $\vec{w}$, where the two vectors are given in each Exercise.

12. $\vec{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

13. $\vec{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

14. $\vec{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

15. $\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

In Exercises 16-19 you are asked to compute the area of the parallelogram formed by the given vectors.

16. $\vec{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$.

17. $\vec{v} = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$.

18. $\vec{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$.

19. $\vec{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

20. **A Pythagorean Identity.** Either using the trigonometric formulas, or the algebraic definitions of the dot and skew products, prove that for any two vectors $\vec{v}_1$ and $\vec{v}_2$, one has the equality

$$\left(\vec{v}_1 \bullet \vec{v}_2\right)^2 + \left(\vec{v}_1 \wedge \vec{v}_2\right)^2 = \left\|\vec{v}_1\right\|^2 \cdot \left\|\vec{v}_2\right\|^2. \tag{3.1.30}$$

21*. **Proof of the Component Theorem.** Suppose $\vec{w}$ and $\vec{v}$ are two vectors, with $\vec{w} \neq \vec{0}$, and assume that we have

$$\vec{v} = \vec{p} + \vec{n},$$

with: (i) $\vec{p}$ either *zero*, or *parallel* to $\vec{w}$; (ii) $\vec{n}$ either *zero*, or *perpendicular* to $\vec{w}$.

Prove the equality $\vec{p} = \left(\dfrac{\vec{v} \bullet \vec{w}}{\left\|\vec{w}\right\|^2}\right)\vec{w}$, following the steps below.

(a) Use condition (ii) to get $\vec{n} \bullet \vec{w} = 0$. Then using the fact that $\vec{n} = \vec{v} - \vec{p}$, obtain the equality: $\vec{v} \bullet \vec{w} = \vec{p} \bullet \vec{w}$.

(b) Argue that, by condition (i), the vector $\vec{p}$ must be of the form

$$\vec{p} = t\vec{w}, \tag{3.1.31}$$

for some number t, and use part (a) to conclude that t satisfies the equalities

$$\vec{v} \bullet \vec{w} = t\left(\vec{w} \bullet \vec{w}\right) = t\left\|\vec{w}\right\|^2.$$

(c) Solve the above equation for t, then replace t in (3.1.31) and obtain the desired formula.

22*. Suppose $\vec{w}$, $\vec{v}$, $\vec{p}$ and $\vec{n}$ are as in the preceding Exercise, and let γ be the geometric angle formed by $\vec{w}$ and $\vec{v}$. Prove that the magnitudes of the vectors $\vec{p}$ and $\vec{n}$ are: $\left\|\vec{p}\right\| = \left\|\vec{v}\right\| \cdot \left|\cos\gamma\right|$ and $\left\|\vec{n}\right\| = \left\|\vec{v}\right\| \cdot \sin\gamma$. Either directly, or using these equalities, prove that:

$$\left\|\vec{p}\right\|^2 + \left\|\vec{n}\right\|^2 = \left\|\vec{v}\right\|^2. \tag{3.1.32}$$

23*. Prove that, given two non-parallel and non-zero vectors $\vec{v}$ and $\vec{w}$, the area of the parallelogram $\mathscr{P}$ formed by them can also be computed as:

$$\text{Area}(\mathscr{P}) = \sqrt{\left\|\vec{v}\right\|^2 \cdot \left\|\vec{w}\right\|^2 - \left(\vec{v} \bullet \vec{w}\right)^2}. \tag{3.1.33}$$

Conclude that the two triangles $\mathscr{T}$ and $\mathscr{T}'$ formed by the two vectors have areas:

$$\text{Area}(\mathscr{T}) = \text{Area}(\mathscr{T}') = \tfrac{1}{2}\sqrt{\left\|\vec{v}\right\|^2 \cdot \left\|\vec{w}\right\|^2 - \left(\vec{v} \bullet \vec{w}\right)^2}. \tag{3.1.34}$$

(HINT: Use the Pythagorean identity (3.1.30) from Exercise 20.)

24*. Use the set-up and notations as in Exercise 23. Prove the equalities:

$$\text{Area}(\mathscr{P}) = \frac{1}{2}\sqrt{4\|\vec{v}\|^2 \cdot \|\vec{w}\|^2 - \left(\|\vec{v} \pm \vec{w}\|^2 - \|\vec{v}\|^2 - \|\vec{w}\|^2\right)^2}. \tag{3.1.35}$$

$$\text{Area}(\mathscr{T}) = \text{Area}(\mathscr{T}') = \frac{1}{4}\sqrt{4\|\vec{v}\|^2 \cdot \|\vec{w}\|^2 - \left(\|\vec{v} \pm \vec{w}\|^2 - \|\vec{v}\|^2 - \|\vec{w}\|^2\right)^2}. \tag{3.1.36}$$

(HINT: Use the preceding Exercise, and the Pythagorean formulas (3.1.3) and (3.1.4).)

3.2 Applications to Triangle Geometry

When were first introduced to the word TRIGONOMETRY, we learned that it has something to do with *triangles*. This was explained very early in the course, when we used the trigonometric functions to solve *right* triangles. With the help of of two fundamental results we are going to learn about in this section (the LAW OF COSINE and the LAW OF SINES) we will be in position to solve *arbitrary* triangles.

As every triangle has six elements (three sides and three angles), the problem of *solving* them will be divided into several cases, for which the following labeling convention is used.

CASE	GIVEN DATA	*Elements to find*
SSS	Three sides	Three angles
SAS	Two sides, and the angle *formed by the given sides*	One side and two angles
SSA	Two sides, and an angle *facing one of the given sides*	One side and two angles
ASA	One side, and two angles *neither of which faces the given side*	Two sides and one angle
AAS	One side, and two angles *one of which faces the given side*	Two sides and one angle

Table 3.2.1

When dealing with problems that require triangle solving, we need to be aware of the following:

A. Most of our problems are solved with a calculator. In most cases the problem demands the angles to be computed in degrees, so the calculator must be set to degree mode.

B. The Labeling Convention from Section 1.1 is in effect. So when we deal, for instance, with a triangle $\triangle ABC$,

- the letter a denotes the side BC, which *faces the angle* $\widehat{A}$;
- the letter a denotes the side AC, which *faces the angle* $\widehat{B}$;
- the letter c denotes the side AB, which *faces the angle* $\widehat{C}$.

C. Identifying the correct case is essential in a successful solution!

The Law of Cosine (a.k.a Pythagoras' Generalized Theorem)

We have already seen Pythagoras' Generalized Theorem presented in Section 3.1 in the form that involves the *dot product*, which gave us two identities (3.1.3) and (3.1.4). The one identity we are interested in is (3.1.3), which reads:

$$\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\|^2 = \left\|\vec{\mathbf{v}}_1\right\|^2 + \left\|\vec{\mathbf{v}}_2\right\|^2 - 2\left(\vec{\mathbf{v}}_1 \bullet \vec{\mathbf{v}}_2\right). \tag{3.2.1}$$

If we replace the dot product using its trigonometric form, we can also write the above identity as:

$$\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\|^2 = \left\|\vec{\mathbf{v}}_1\right\|^2 + \left\|\vec{\mathbf{v}}_2\right\|^2 - 2 \cdot \left\|\vec{\mathbf{v}}_1\right\| \cdot \left\|\vec{\mathbf{v}}_2\right\| \cdot \cos\gamma, \tag{3.2.2}$$

where γ is the *geometric angle* formed by the vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$.

Suppose now we have a triangle, and we label its sides as side_1, side_2, and side_3.

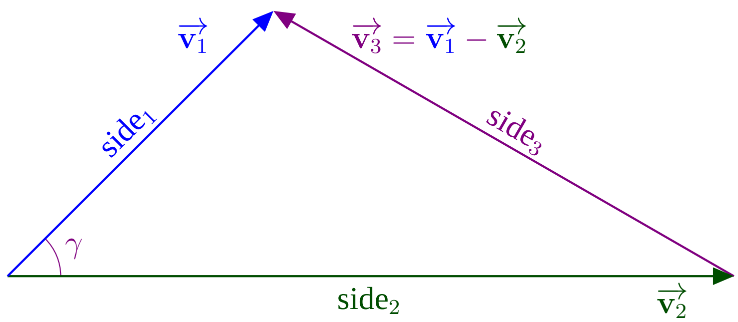


Figure 3.2.1

Suppose we place two vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ sitting on the two sides as shown above, so we can identify $\left\|\vec{\mathbf{v}}_1\right\| = \text{side}_1$ and $\left\|\vec{\mathbf{v}}_2\right\| = \text{side}_2$. Using the triangle rule, it follows that the vector $\vec{\mathbf{v}}_3 = \vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2$ has magnitude $\left\|\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2\right\| = \text{side}_3$. With all these identifications, we can re-write (3.2.2) in a form that does not refer to any vectors whatsoever, and thus obtain the following fundamental statement.

The Law Of Cosine

Any triangle, with sides labeled side_1, side_2, and side_3, satisfies the identity:

$$[\text{side}_3]^2 = [\text{side}_1]^2 + [\text{side}_2]^2 - 2 \cdot \text{side}_1 \cdot \text{side}_2 \cdot \cos\left(\text{angle facing side}_3\right). \tag{3.2.3}$$

The reason we refer to the Law of Cosine as Pythagoras' Generalized Theorem is the fact that, in the special case when the angle facing side_3 is a *right angle* (in which case its cosine is zero), we recover Pythagoras' usual Theorem, since the triangle formed by side_1, side_2, and side_3 will be a *right triangle*, with side_3 as its *hypotenuse*.

CLARIFICATION. With the set-up from the statement of the Law of Cosine, we can of course switch the sides around (by relabeling), so besides (3.2.3), we will also get two additional identities:

$$[\text{side}_1]^2 = [\text{side}_2]^2 + [\text{side}_3]^2 - 2 \cdot \text{side}_2 \cdot \text{side}_3 \cdot \cos\left(\text{angle facing side}_1\right); \tag{3.2.4}$$

$$[\text{side}_2]^2 = [\text{side}_1]^2 + [\text{side}_3]^2 - 2 \cdot \text{side}_1 \cdot \text{side}_3 \cdot \cos\left(\text{angle facing side}_2\right). \tag{3.2.5}$$

Also, by easy algebraic manipulations, we can re-write all of (3.2.3), (3.2.4) and (3.2.5) as:

$$\cos\left(\text{angle facing side}_1\right) = \frac{[\text{side}_2]^2 + [\text{side}_3]^2 - [\text{side}_1]^2}{2 \cdot \text{side}_2 \cdot \text{side}_3}; \tag{3.2.6}$$

$$\cos\left(\text{angle facing side}_2\right) = \frac{[\text{side}_1]^2 + [\text{side}_3]^2 - [\text{side}_2]^2}{2 \cdot \text{side}_1 \cdot \text{side}_3}; \tag{3.2.7}$$

$$\cos\left(\text{angle facing side}_3\right) = \frac{[\text{side}_1]^2 + [\text{side}_2]^2 - [\text{side}_3]^2}{2 \cdot \text{side}_1 \cdot \text{side}_2}. \tag{3.2.8}$$

Solving the SSS Problem

If we are given all three sides of a triangle, the three angles can be easily computed using the formulas (3.2.6), (3.2.7), (3.2.8). Since each one of these formulas leads to an equation of the form

$$\cos ? = \textit{number},$$

and the unknown angle is in the interval $(0, \pi)$ (or $(0^\circ, 180^\circ)$, if we use degrees), the above equation will only have one solution, given by the arccosine function. So we can safely rewrite each one of the equalities (3.2.6), (3.2.7), (3.2.8) in the following form.

Derived Law of Cosine for Angles

The angles of any triangle, with sides labeled side_1, side_2, and side_3, are given by

$$\text{angle facing side}_1 = \arccos\left(\frac{[\text{side}_2]^2 + [\text{side}_3]^2 - [\text{side}_1]^2}{2 \cdot \text{side}_2 \cdot \text{side}_3}\right); \tag{3.2.9}$$

$$\text{angle facing side}_2 = \arccos\left(\frac{[\text{side}_1]^2 + [\text{side}_3]^2 - [\text{side}_2]^2}{2 \cdot \text{side}_1 \cdot \text{side}_3}\right); \tag{3.2.10}$$

$$\text{angle facing side}_3 = \arccos\left(\frac{[\text{side}_1]^2 + [\text{side}_2]^2 - [\text{side}_3]^2}{2 \cdot \text{side}_1 \cdot \text{side}_2}\right). \tag{3.2.11}$$

Based on these formulas, the SSS problem can be solved as follows.

Solution of the SSS Problem based on Law of Cosine

Given three sides in a triangle, we find the three angles as follows.

I. Find two missing angles using two of the formulas from the Derived Law of Cosine for Angles.
II. Once two angles are found, the third one is found by subtracting the two angles found in step I, from 180°.

In the case of an SSS problem, it is quite possible to have no solution! This would happen precisely when one of the triangle inequalities $\text{side}_1 + \text{side}_2 > \text{side}_3$, $\text{side}_1 + \text{side}_3 > \text{side}_2$, $\text{side}_2 + \text{side}_3 > \text{side}_1$, fails. For example, if we want to solve a triangle with sides $a = 5$, $b = 10$ and $c = 17$, we quickly see that the inequality $a + b > c$ *does not work*, so we have no such triangle!

Example 3.2.1. Suppose we want to solve the triangle $\triangle ABC$, in which we are given $a = 7$, $b = 5$ and $c = 3$ (all measured in inches).

The procedure outlined above is carried on as follows.

I. We begin by finding the angles $\widehat{A}$ and $\widehat{B}$:

$$\widehat{A} = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right) = \arccos\left(\frac{5^2 + 3^2 - 7^2}{2 \cdot 5 \cdot 3}\right) = \arccos\left(-\frac{15}{30}\right) = 120^\circ;$$
$$\widehat{B} = \arccos\left(\frac{a^2 + c^2 - b^2}{2ac}\right) = \arccos\left(\frac{7^2 + 3^2 - 5^2}{2 \cdot 7 \cdot 3}\right) = \arccos\left(\frac{33}{42}\right) \simeq 38.2132107^\circ.$$

II. With the two angles we found, we can find the third angle:

$$\widehat{C} = 180^\circ - \widehat{A} - \widehat{B} \simeq 180^\circ - 120^\circ - 38.2132107^\circ \simeq 21.7867893^\circ.$$

Solving the SAS Problem

If we are given two sides of a triangle, and the angle formed by them (which faces the missing side), then we can find the third side using the Law of Cosine, in any one of its presentations: (3.2.3), (3.2.4), or (3.2.5). Since each one of these formulas leads to an equation of the form

$$[\text{missing side}]^2 = \textit{number},$$

and the unknown is *positive*, we can take the square root, so we can safely rewrite each one of the equalities (3.2.3), (3.2.4), (3.2.5). in the following form.

Derived Law of Cosine for Sides

The sides side_1, side_2, side_3, in any triangle, satisfy the equalities

$$\text{side}_1 = \sqrt{[\text{side}_2]^2 + [\text{side}_3]^2 - 2 \cdot \text{side}_2 \cdot \text{side}_3 \cdot \cos\left(\text{angle facing side}_1\right)}; \tag{3.2.12}$$

$$\text{side}_2 = \sqrt{[\text{side}_1]^2 + [\text{side}_3]^2 - 2 \cdot \text{side}_1 \cdot \text{side}_3 \cdot \cos\left(\text{angle facing side}_2\right)}. \tag{3.2.13}$$

$$\text{side}_3 = \sqrt{[\text{side}_1]^2 + [\text{side}_2]^2 - 2 \cdot \text{side}_1 \cdot \text{side}_2 \cdot \cos\left(\text{angle facing side}_3\right)}. \tag{3.2.14}$$

Based on these formulas, the SAS problem can be solved as follows.

Solution of the SAS Problem based on Law of Cosine

Given two sides of a triangle, and the angle formed by them, we find the third side and the other two angles as follows.

I. Find the missing side using one of the formulas from the Derived Law of Cosine for Sides.

II. Find one of the missing angles using one of the formulas from the Derived Law of

Cosine for Angles.

III. Once one missing angle are found, the third one is found by subtracting the two angles (the given one and the one found in step II) from $180°$.

Example 3.2.2. Suppose we want to solve the triangle $\triangle ABC$, in which we are given $a = 10$, $b = 5$ (both measured in inches), and $\widehat{C} = 35°$.

The procedure outlined above is carried on as follows.

I. We begin by finding the third side

$$c = \sqrt{a^2 + b^2 - 2ab\cos\widehat{C}} = \sqrt{10^2 + 5^2 - 2\cdot 10\cdot 5\cdot \cos 35°} \simeq 6.563900942 \text{ in.}$$

II. Next we find one of the two missing angles

$$\widehat{A} = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right) = \arccos\left(\frac{5^2 + 6.563900942^2 - 10^2}{2\cdot 5\cdot 6.563900942}\right) \simeq 119.0926395°.$$

III. With the two angles $\widehat{A} =$ and $\widehat{C}$ in hand, we can find the third angle:

$$\widehat{B} = 180° - \widehat{A} - \widehat{C} \simeq 180° - 119.0926395° - 35° \simeq 25.9073605°.$$

Area Formulas

When we obtained the Law of Cosine using the dot product, we "played" with two vectors, which we placed on two sides of the triangle. Using Figure 3.2.1 as a guideline, we can also compute the area of the triangle, with the help of formula (3.1.29) from Section 3.1, which reads:

$$\text{Area}(\mathscr{T}) = \frac{1}{2}\cdot \|\vec{\mathbf{v}}_1\| \cdot \|\vec{\mathbf{v}}_2\| \cdot \sin\gamma.$$

So when we replace again $\|\vec{\mathbf{v}}_1\| = \text{side}_1$ and $\|\vec{\mathbf{v}}_2\| = \text{side}_2$, we obtain the following useful result.

Side-Angle-Side Area Formula

Any triangle $\mathscr{T}$, with sides labeled side_1, side_2, and side_3, has area given by:

$$\text{Area}(\mathscr{T}) = \frac{1}{2}\cdot \text{side}_1 \cdot \text{side}_2 \cdot \sin\left(\text{angle facing side}_3\right). \tag{3.2.15}$$

CLARIFICATION. With the set-up as above, we can of course switch the sides around (by relabeling), so besides (3.2.15), we will also get two additional formulas:

$$\text{Area}(\mathscr{T}) = \frac{1}{2}\cdot \text{side}_1 \cdot \text{side}_3 \cdot \sin\left(\text{angle facing side}_2\right); \tag{3.2.16}$$

$$\text{Area}(\mathscr{T}) = \frac{1}{2}\cdot \text{side}_2 \cdot \text{side}_3 \cdot \sin\left(\text{angle facing side}_1\right). \tag{3.2.17}$$

Example 3.2.3. Suppose we have a triangle $\triangle ABC$, with $a = 10$, $b = 8$ (both measured in inches), $\widehat{C} = 45°$, and we want to find its area.

Using (3.2.15), we can simply write

$$\text{Area}(\triangle ABC) = \frac{1}{2}ab\sin\widehat{C} = \frac{1}{2}\cdot 108\cdot\sin 45° = \frac{1}{2}\cdot 108\cdot\frac{\sqrt{2}}{2} = \\ = 20\sqrt{2}\text{ sq.in.} \simeq 28.28427125\text{ sq.in.}$$

It is also possible to derive another area formula, which does not involve the sine function. The neatest way to present the above formula is as follows.

Heron's Area Formula

Any triangle $\mathscr{T}$, with sides labeled side_1, side_2, and side_3, has area given by:

$$\text{Area}(\mathscr{T}) = \sqrt{\text{s}(\text{s}-\text{side}_1)(\text{s}-\text{side}_2)(\text{s}-\text{side}_3)}, \tag{3.2.18}$$

where s is the *semi-perimeter*: $\text{s} = \frac{1}{2}(\text{side}_1 + \text{side}_2 + \text{side}_3)$.

Proof. What follows is a long and tedious algebraic computation. On a first reading, you may want to skip it and go directly to Example 3.2.4 below.

Denote the three sides of the triangle simply by s_1, s_2, s_3, and let θ denote the angle facing s_3, so using (3.2.15) we have $\text{Area}(\mathscr{T}) = \dfrac{1}{2}s_1s_2\sin\theta$. Next, using the fact that $\sin\theta \geq 0$, we can safely say that $\sin\theta = \sqrt{1-\cos^2\theta}$, so now we can write

$$\text{Area}(\mathscr{T}) = \frac{1}{2}s_1s_2\sqrt{1-\cos^2\theta} \tag{3.2.19}$$

Using the Law of Cosine in the form (3.2.8), we know that $\cos\theta = \dfrac{s_1^2+s_2^2-s_3^2}{2s_1s_2}$, so using this as a replacement for $\cos\theta$ in (3.2.19), we get:

$$\begin{aligned}\text{Area}(\mathscr{T}) &= \frac{1}{2}s_1s_2\sqrt{1-\left(\frac{s_1^2+s_2^2-s_3^2}{2s_1s_2}\right)^2} = \frac{s_1s_2}{2}\sqrt{1-\frac{(s_1^2+s_2^2-s_3^2)^2}{(2s_1s_2)^2}} = \\ &= \frac{s_1s_2}{2}\sqrt{\frac{(2s_1s_2)^2}{(2s_1s_2)^2}-\frac{(s_1^2+s_2^2-s_3^2)^2}{(2s_1s_2)^2}} = \frac{s_1s_2}{2}\sqrt{\frac{(2s_1s_2)^2-(s_1^2+s_2^2-s_3^2)^2}{(2s_1s_2)^2}} = \\ &= \frac{s_1s_2\sqrt{(2s_1s_2)^2-(s_1^2+s_2^2-s_3^2)^2}}{2\cdot(2s_1s_2)} = \frac{1}{4}\sqrt{(2s_1s_2)^2-(s_1^2+s_2^2-s_3^2)^2}.\end{aligned} \tag{3.2.20}$$

Using the formula for the difference of squares, we can factor the expression under the radical as

$$\begin{aligned}(2s_1s_2)^2-\left(s_1^2+s_2^2-s_3^2\right)^2 &= \left[2s_1s_2+\left(s_1^2+s_2^2-s_3^2\right)\right]\cdot\left[2s_1s_2-\left(s_1^2+s_2^2-s_3^2\right)\right] = \\ &= \left[\left(s_1^2+s_2^2+2s_1s_2\right)-s_3^2\right]\cdot\left[s_3^2-\left(s_1^2+s_2^2-2s_1s_2\right)\right].\end{aligned} \tag{3.2.21}$$

Using the square of sum/difference formulas, we recognize $s_1^2+s_2^2+2s_1s_2 = (s_1+s_2)^2$ and $s_1^2+s_2^2-2s_1s_2 = (s_1-s_2)^2$, so the calculation from (3.2.21) continues as:

$$(2s_1^2s_2^2)^2-\left(s_1^2+s_2^2-s_3^2\right)^2 = \left[(s_1+s_2)^2-s_3^2\right]\cdot\left[s_3^2-(s_1-s_2)^2\right],$$

and then, when we use again the difference of squares formula, we can also write:

$$(2s_1^2s_2^2)^2-\left(s_1^2+s_2^2-s_3^2\right)^2=[s_1+s_2+s_3]\cdot[s_1+s_2-s_3]\cdot[s_3+s_1-s_2]\cdot[s_3+s_2-s_1]\,. \tag{3.2.22}$$

Since $s_1+s_2+s_3=2\mathbf{s}$, we can also write

$$\begin{aligned}
s_1+s_2-s_3&=(s_1+s_2+s_3)-2s_1=2\mathbf{s}-2s_1=2(\mathbf{s}-s_1),\\
s_3+s_1-s_2&=(s_1+s_2+s_3)-2s_2=2\mathbf{s}-2s_2=2(\mathbf{s}-s_2),\\
s_3+s_2-s_1&=(s_1+s_2+s_3)-2s_3=2\mathbf{s}-2s_3=2(\mathbf{s}-s_3),
\end{aligned}$$

so now (3.2.22) reads:

$$\begin{aligned}
(2s_1^2s_2^2)^2-\left(s_1^2+s_2^2-s_3^2\right)^2&=\left[2\mathbf{s}\right]\cdot\left[2(\mathbf{s}-s_1)\right]\cdot\left[2(\mathbf{s}-s_2)\right]\cdot\left[2(\mathbf{s}-s_3)\right]=\\
&=16\mathbf{s}(\mathbf{s}-s_1)(\mathbf{s}-s_2)(\mathbf{s}-s_3),
\end{aligned}$$

and then when we plug this under the radical in (3.2.20), the desired formula (3.2.18) follows. □

Example 3.2.4. Suppose we have a triangle $\triangle ABC$, with $a=10$, $b=9$, $c=7$ (all measured in inches), and we want to find its area.

We will use Heron's Formula (3.2.18), for which we first need the semi-perimeter:

$$\mathbf{s}=\frac{1}{2}\left(a+b+c\right)=\frac{1}{2}\left(10+7+9\right)=\frac{1}{2}\cdot 26=13.$$

Using this calculation, by Heron's Formula (3.2.18), we get

$$\begin{aligned}
\text{Area}(\triangle ABC)&=\sqrt{\mathbf{s}(\mathbf{s}-a)(\mathbf{s}-b)(\mathbf{s}-c)}=\sqrt{13(13-10)(13-9)(13-7)}=\\
&=\sqrt{936}\ \text{sq.in.}\simeq\sqrt{30.59411708}\ \text{sq.in.}
\end{aligned}$$

The Law of Sines

The second most important result in this section relates all sides and all the angles facing them, as follows.

The Law of Sines

In any triangle $\mathscr{T}$, with sides labeled side_1, side_2, and side_3, one has the following equalities:

$$\frac{\text{side}_1}{\sin\left(\text{angle facing side}_1\right)}=\frac{\text{side}_2}{\sin\left(\text{angle facing side}_2\right)}=\frac{\text{side}_3}{\sin\left(\text{angle facing side}_3\right)} \tag{3.2.23}$$

Proof. Start off with the Side-Angle-Side Area Formulas (3.2.15), (3.2.16), and (3.2.17), which yield the following two equalities:

$$\text{side}_1\cdot\text{side}_3\cdot\sin\left(\text{angle facing side}_2\right)=\text{side}_2\cdot\text{side}_3\cdot\sin\left(\text{angle facing side}_1\right); \tag{3.2.24}$$

$$\text{side}_1\cdot\text{side}_3\cdot\sin\left(\text{angle facing side}_2\right)=\text{side}_1\cdot\text{side}_2\cdot\sin\left(\text{angle facing side}_3\right). \tag{3.2.25}$$

If we divide both sides of (3.2.24) by side_3 we get

$$\text{side}_1\cdot\sin\left(\text{angle facing side}_2\right)=\text{side}_2\cdot\sin\left(\text{angle facing side}_1\right),$$

which is equivalent to the first equality from (3.2.23):

$$\frac{\text{side}_1}{\sin\left(\text{angle facing side}_1\right)} = \frac{\text{side}_2}{\sin\left(\text{angle facing side}_2\right)}.$$

Likewise, if we divide both sides of (3.2.25) by side_1 we get

$$\text{side}_3 \cdot \sin\left(\text{angle facing side}_2\right) = \text{side}_2 \cdot \sin\left(\text{angle facing side}_3\right),$$

which is equivalent to the second equality from (3.2.23):

$$\frac{\text{side}_2}{\sin\left(\text{angle facing side}_2\right)} = \frac{\text{side}_3}{\sin\left(\text{angle facing side}_3\right)}. \qquad \square$$

The Law of Sines has numerous applications to triangle solving, which we will discuss shortly. At this time, we give a more "theoretical" application, which tells us something about the ordering of the angles in a triangle.

Side-Angle Ordering Rule

If two sides in a triangle satisfy $\text{side}_1 > \text{side}_2$, then:

$$\text{angle facing side}_1 > \text{angle facing side}_2. \tag{3.2.26}$$

Proof. Let us denote the two sides simply by s_1, s_2, s_3, and the angles facing them by θ_1, θ_2, θ_3. With these simplified notations, we are given that $s_1 > s_2$, and we need to show that $\theta_1 > \theta_2$.

The first observation is that, since we have

$$\theta_1 + \theta_2 = 180^\circ - \theta_3 < 180^\circ,$$

we always have the inequalities

$$\theta_1 < 180^\circ - \theta_2; \tag{3.2.27}$$

$$\theta_2 < 180^\circ - \theta_1. \tag{3.2.28}$$

In particular, in the case when $\theta_1 \geq 90^\circ$, there is nothing to prove, because (3.2.27) would force $\theta_2 < 90^\circ \leq \theta_1$.

So the only other case to consider is when θ_1 is *acute*. In this case we notice that, when we rewrite the first equality in (3.2.23) in the from

$$\frac{s_1}{s_2} = \frac{\sin\theta_1}{\sin\theta_2},$$

then using the inequality $s_1 > s_2$ (which is the same as $\frac{s_1}{s_2} > 1$), it follows that $\frac{\sin\theta_1}{\sin\theta_2} > 1$, thus:

$$\sin\theta_1 > \sin\theta_2. \tag{3.2.29}$$

There are now two possibilities to consider:

(i) $0 < \theta_2 \leq 90^\circ$;

(ii) $90^\circ < \theta_2 < 180^\circ$.

In case (i) it follows that both θ_1 and θ_2 sit in the interval $[0^\circ, 90^\circ]$ on which sin is *increasing*, and then (3.2.29) will clearly force $\theta_1 > \theta_2$.

As for case (ii), we are going to show that it is *impossible*. Indeed, if we use the angle $\alpha = 180^\circ - \theta_2$ which by the formulas for supplements has $\sin\alpha = \sin\theta_2$, so now (3.2.29) will read

$$\sin\theta_1 > \sin\alpha,$$

and then arguing as above (now θ and α will be in $[0^\circ, 90^\circ]$), this will force

$$\theta_1 > \alpha = 180^\circ - \theta_2,$$

which clearly violates (3.2.27). □

Solving the ASA and the AAS Problems

The Law of Sines turns out to be very effective for solving triangles, when two angles are given, in which case the following procedure can be employed.

Solution of the ASA and AAS Problems based on Law of Sines

Given one side and two angles of a triangle, we find the third angle and the other two sides as follows.

I. Find the third angle, by subtracting the sum of the two given angles from 180°.

II. Once the third angle is found, set up the equalities (3.2.23) given by the Law of Sines, and by setting up two *proportion equations* (in which the given side appears as one of the numerators), solve for the missing sides.

Example 3.2.5. Suppose we want to solve the triangle $\triangle ABC$, in which we are given $a = 10$ cm, $\widehat{B} = 52^\circ$, and $\widehat{C} = 34^\circ$.

The procedure outlined above is carried on as follows.

I. We begin by finding the third angle

$$\widehat{A} = 180^\circ - \widehat{B} - \widehat{C} = 180^\circ - 52^\circ - 34^\circ = 94^\circ.$$

II. We now set up the Law of Sines $\dfrac{a}{\sin\widehat{A}} = \dfrac{b}{\sin\widehat{B}} = \dfrac{c}{\sin\widehat{C}}$, by replacing all known quantities:

$$\frac{10}{\sin 94^\circ} = \frac{b}{\sin 52^\circ} = \frac{c}{\sin 34^\circ}.$$

To compute b we solve the proportion equation $\dfrac{10}{\sin 94^\circ} = \dfrac{b}{\sin 52^\circ}$ by cross-multiplication an division, which yields

$$b = \frac{10\cdot\sin 52^\circ}{\sin 94^\circ} \simeq 7.899349956 \text{ cm}.$$

To compute c we solve the proportion equation $\dfrac{10}{\sin 104^\circ} = \dfrac{c}{\sin 34^\circ}$ by cross-multiplication an division, which yields

$$c = \frac{10\cdot\sin 34^\circ}{\sin 94^\circ} \simeq 5.605583955 \text{ cm}.$$

Using the Law of Sines to Find Angles

If we pay close attention to Example 3.2.5, we see that we can also use proportion equations to find sines of (unknown) angles. For instance, if we have a triangle $\triangle ABC$, in which we know a, $\widehat{A}$ and b, using the Law of Sines, which gives us the proportion $\dfrac{a}{\sin \widehat{A}} = \dfrac{b}{\sin \widehat{B}}$, we can find:

$$\sin \widehat{B} = \frac{b \sin \widehat{A}}{a}.$$

In principle, the above equality should be enough for finding the angle $\widehat{B}$, but we need to be extremely careful! This is due to the following important (and quite subtle) fact.

PRECAUTIONARY FACT. If θ represents an *angle in a triangle*, then *the value of* $\sin \theta$ *alone, does not determine the value of* θ *uniquely!* So, *without any additional information*, given

$$\sin \theta = \textit{number}, \tag{3.2.30}$$

there are *two* possible values of θ:
(a) one *acute* angle: $\theta_1 = \arcsin(\textit{number})$;
(b) one *obtuse* angle: $\theta_2 = 180° - \arcsin(\textit{number})$.
The only exceptions to these rules are:
(i) If *number* ≤ 0, or *number* > 1, there is *no possible value for* θ;
(ii) If *number* $= 1$, then only *one value* is possible: $\theta = 90°$.

CLARIFICATION. As seen in the Precautionary Fact above, the main drawback of working with sine, when dealing with angles in a triangles, is the fact that the equation (3.2.30) may have *two* solutions. This clearly contrasts what happens with a cosine: As it turns out, if θ an angle in a triangle, an equation of the form "$\cos \theta = \textit{number}$" has exactly *one* solution: $\theta = \arccos(\textit{number})$. (The only exceptions are the cases when *number* ≤ -1, or *number* > 1, when we have no solution).

In other words, sine *does not differentiate between acute and obtuse angles*, but cosine *does!*. For this reason, using the Law of Cosine is always preferred, when solving triangles.

"Shortcuts" for the SSS and SAS Problems, using the Law of Sines

The methods for solving the SSS and the SAS problem, outlined earlier (both of which were based on the Law of Cosine), can be slightly improved using the Law of Sines, with a careful application of the Precautionary Fact stated above. Although the outline below has four steps, the method is often less time consuming than the earlier ones, because it involves fewer keystrokes on the calculator.

"Hybrid" Solutions of the SSS and SAS Problems

When solving an SSS or an SAS problem, proceed as follows.

I. Find *one missing angle* (in the SSS problem), or *one missing side* (in the SAS problem), using the corresponding Derived Law of Cosine. *At the end of this step, we will have "in hand" three sides and one angle*.

II. Set up the Law of Sines, with the information acquired at the end of Step I, and *order the two missing angles* based on the *order of the sides facing them*. Identify the *smaller of the two missing angles*. This particular angle is necessarily *acute*.

III. Find the *value* of the sine of the smaller of the two missing angles, using a proportion equation, then compute this angle using:

$$\textit{smaller angle} = \arcsin(\textit{value}). \tag{3.2.31}$$

IV. Once one missing angle is found, the second missing angle is found by subtracting the two known angles from $180°$.

CLARIFICATION. Since the angle identified in Step III is acute, the formula (3.2.31) is "safe" to use, as exaplained in the Precautionary Fact above.

Example 3.2.6. Suppose we want to solve the same triangle $\triangle ABC$, given in Example 3.2.1, in which we were given $a = 7$, $b = 5$ and $c = 3$ (all measured in inches).

I. We start off as in Example 3.2.1, by finding one of the missing angles using the Law of Cosine. For example, we can compute $\widehat{A} = 120°$. (See Example 3.2.1 for details on this calculation.)

II. At this point, although we could also try to find one of the other angles again using the Law of Cosine, we now proceed by setting up the Law of Sines $\dfrac{a}{\sin\widehat{A}} = \dfrac{b}{\sin\widehat{B}} = \dfrac{c}{\sin\widehat{C}}$, which gives:

$$\frac{7}{\sin 120°} = \frac{5}{\sin\widehat{B}} = \frac{3}{\sin\widehat{C}}.$$

Since $b = 5$ and $c = 3$, the smaller of the missing angles is $\widehat{C}$, which is then acute.

III. Using the proportion equation $\dfrac{7}{\sin 120°} = \dfrac{3}{\sin\widehat{C}}$, we can find

$$\sin\widehat{C} = \frac{3\sin 120°}{7} \simeq 0.371153744,$$

and then we can compute: $\widehat{C} = \arcsin(0.371153744) \simeq 21.7867893°$.

IV. With the angles $\widehat{A} = 120°$ and $\widehat{C} \simeq 21.7867893°$ "in hand," we can find the third angle:

$$\widehat{B} = 180° - \widehat{A} - \widehat{C} \simeq 180° - 120° - 21.7867893° \simeq 38.2132107°.$$

Example 3.2.7. Suppose we want to solve the same triangle $\triangle ABC$, given in Example 3.2.2, $a = 10$, $b = 5$ (both measured in inches), and $\widehat{C} = 35°$.

I. We start off as in Example 3.2.2, by finding the missing side $c \simeq 6.563900942$ in, using the Law of Cosine. (See Example 3.2.2 for details on this calculation.)

II. With the information available we set up the Law of Sines $\frac{a}{\sin \widehat{A}} = \frac{b}{\sin \widehat{B}} = \frac{c}{\sin \widehat{C}}$, which yields:

$$\frac{10}{\sin \widehat{A}} = \frac{5}{\sin \widehat{B}} = \frac{6.563900942}{\sin 35^\circ}.$$

Since $a = 10$ and $b = 5$, the smaller of the missing angles is $\widehat{B}$, which is then acute.

III. Using the proportion equation $\frac{5}{\sin \widehat{B}} = \frac{6.563900942}{\sin 35^\circ}$, we can find

$$\sin \widehat{B} = \frac{5 \sin 35^\circ}{6.563900942} \simeq 0.436917346,$$

and then we can compute: $\widehat{B} = \arcsin(0.436917346) \simeq 25.9073605^\circ$.

IV. With the angles $\widehat{B} \simeq 25.9073605^\circ$ and $\widehat{C} = 35^\circ$ “in hand,” we can find the third angle:

$$\widehat{A} = 180^\circ - \widehat{B} - \widehat{C} \simeq 180^\circ - 25.9073605^\circ - 35^\circ \simeq 119.0926395^\circ.$$

Solving the SSA Problem

The SSA Problem can be tricky! As we shall see, such a problem has three possible outcomes:

(A) *no solution;*
(B) *one solution;*
(C) *two solutions.*

As it turns out, we can solve the SSA problem in two ways, using either the Law of Sines (the preferred method), or using the Law of Cosine (which many textbooks ignore!). In either method, after several steps, we will reduce the problem to a “familiar” case, of one of the types discussed earlier (ASA, AAS, or SSS).

Solution of the SSA based on Law of Sines

Given two sides in a triangle, and an angle facing one of the given sides, we solve the triangle as follows.

I. (Preparation) Compare the given sides and figure out the *order of the angles facing them*. In some peculiar cases, you can bypass Steps II and III.

II. Set up the Law of Sines, and find the *value* of the sine of the *missing angle that faces one of the given sides*, by solving the proportion equation. If *value* > 1, then we stop here! The Problem has no solution.

III. With the *value* found in Step II, find the missing angle, which can be *either one* of:

$$\text{angle}_1 = \arcsin(\textit{value}). \tag{3.2.32}$$

$$\text{angle}_2 = 180^\circ - \arcsin(\textit{value}). \tag{3.2.33}$$

Using the information from Step I, decide if *one*, or *both* angles above are acceptable.

IV. With each one of the angles found in Step III, we have an AAS Problem. Solve each one of these problems as outlined earlier:
   (i) Find the third angle by subtracting the two known angles from $180°$.
   (ii) Go back to the Law of Sines and solve a proportion equation to find the missing side.

If after Step III you are left with *two* AAS problems, solve them separately. The given problem will have *two* solutions, which must be written separately. (This situation is illustrated in Example 3.2.8.)

Do not skip Step I! There are instances (illustrated in Examples 3.2.11, 3.2.12), in which you can very quickly get your answer, without the need of Steps II and III.

Example 3.2.8. Suppose we want to solve triangle $\triangle ABC$, given $a = 10$, $b = 8$ (both measured in inches), and $\widehat{B} = 40°$.

I. Since $a > b$, it follows that $\widehat{A} > \widehat{B}$.

II. Set up the Law of Sines

$$\frac{a}{\sin \widehat{A}} = \frac{b}{\sin \widehat{B}} = \frac{c}{\sin \widehat{C}} \tag{3.2.34}$$

by plugging in the given values, which yields:

$$\frac{10}{\sin \widehat{A}} = \frac{8}{\sin 40°} = \frac{c}{\sin \widehat{C}}.$$

Using the proportion equation $\dfrac{10}{\sin \widehat{A}} = \dfrac{8}{\sin 40°}$, we can solve by cross-multiplication and division, which gives:

$$\sin \widehat{A} = \frac{10 \sin 40°}{8} \simeq 0.803484512. \tag{3.2.35}$$

III. The possible solutions of (3.2.35) are:

$$\widehat{A}_1 = \arcsin(0.803484512) \simeq 53.46414901°; \tag{3.2.36}$$

$$\widehat{A}_2 = 180° - \arcsin(0.803484512) \simeq 126.535851°. \tag{3.2.37}$$

By Step I we know that $\widehat{A} > \widehat{B}$, and clearly both (3.2.36) and (3.2.37) are acceptable. This means that our problem will have *two solutions*.

IV. For the *first* solution, we use (3.2.36) to get the third angle

$$\widehat{C}_1 = 180° - \widehat{A}_1 - \widehat{B} \simeq 180° - 53.46414901° - 40° \simeq 86.53585099°. \tag{3.2.38}$$

Now if we go back to the Law of Sines (3.2.34), we also get $\dfrac{8}{\sin 40°} = \dfrac{c}{\sin 86.53585099°}$, so the value of c given by the *first* solution is:

$$c_1 = \frac{8 \sin 86.53585099°}{\sin 40°} \simeq 12.423040969 \text{ in.} \tag{3.2.39}$$

IV. For the *second* solution, we use (3.2.37) to get the third angle

$$\widehat{C}_2 = 180^\circ - \widehat{A}_2 - \widehat{B} \simeq 180^\circ - 126.535851^\circ - 40^\circ \simeq 13.46414901^\circ. \tag{3.2.40}$$

Now if we go back to the Law of Sines (3.2.34), we also get $\dfrac{8}{\sin 40^\circ} = \dfrac{c}{\sin 13.46414901^\circ}$, so the value of c given by the *second* solution is:

$$c_2 = \frac{8\sin 13.46414901^\circ}{\sin 40^\circ} \simeq 2.897839169 \text{ in.} \tag{3.2.41}$$

Conclusion: Our problem has *two* solutions:

$$\text{Solution}_1 : \begin{cases} \widehat{A}_1 & \simeq 53.46414901^\circ \\ \widehat{C}_1 & \simeq 86.53585099^\circ \\ c_1 & \simeq 12.423040969 \text{ in.} \end{cases} \qquad \text{Solution}_2 : \begin{cases} \widehat{A}_2 & \simeq 126.535851^\circ \\ \widehat{C}_2 & \simeq 13.46414901^\circ \\ c_2 & \simeq 2.897839169 \text{ in.} \end{cases}$$

Example 3.2.9. Suppose we want to solve triangle $\triangle ABC$, given $a = 7$, $b = 10$ (both measured in inches), and $\widehat{B} = 40^\circ$.

I. Since $a < b$, it follows that $\widehat{A} < \widehat{B}$.

II. Set up the Law of Sines

$$\frac{a}{\sin \widehat{A}} = \frac{b}{\sin \widehat{B}} = \frac{c}{\sin \widehat{C}} \tag{3.2.42}$$

by plugging in the given values, which yields:

$$\frac{7}{\sin \widehat{A}} = \frac{10}{\sin 40^\circ} = \frac{c}{\sin \widehat{C}}.$$

Using the proportion equation $\dfrac{7}{\sin \widehat{A}} = \dfrac{10}{\sin 40^\circ}$, we can solve by cross-multiplication and division, which gives:

$$\sin \widehat{A} = \frac{7\sin 40^\circ}{10} \simeq 0.449951327. \tag{3.2.43}$$

III. The possible solutions of (3.2.43) are:

$$\widehat{A}_1 = \arcsin(0.449951327) \simeq 26.74056117^\circ; \tag{3.2.44}$$

$$\widehat{A}_2 = 180^\circ - \arcsin(0.449951327) \simeq 153.2594388^\circ. \tag{3.2.45}$$

By Step I we know that $\widehat{A} < \widehat{B}$, and is is obvious that the solution (3.2.45) is not acceptable. Therefore, we can only continue with (3.2.44), which means that our problem will have *only one solution*.

IV. Using our only possible value for $\widehat{A}$, given by (3.2.44) our third angle is:

$$\widehat{C} = 180^\circ - \widehat{A} - \widehat{B} \simeq 180^\circ - 26.74056117^\circ - 40^\circ \simeq 113.2594388^\circ. \tag{3.2.46}$$

Now if we go back to the Law of Sines (3.2.34), we also get $\dfrac{10}{\sin 40^\circ} = \dfrac{c}{\sin 113.2594388^\circ}$, so the value of c is:

$$c = \frac{10\sin 113.2594388^\circ}{\sin 40^\circ} \simeq 14.2928419 \text{ in.} \tag{3.2.47}$$

Conclusion: Our problem has *one* solution:

$$\begin{cases} \widehat{A} & \simeq 26.74056117^\circ \\ \widehat{C} & \simeq 113.2594388^\circ \\ c & \simeq 14.2928419 \text{ in.} \end{cases}$$

Example 3.2.10. Suppose we want to solve triangle $\triangle ABC$, given $a = 10$, $c = 6$ (both measured in inches), and $\widehat{C} = 80^\circ$.

I. Since $a > c$, it follows that $\widehat{A} > \widehat{C}$.

II. Set up the Law of Sines

$$\frac{a}{\sin \widehat{A}} = \frac{b}{\sin \widehat{B}} = \frac{c}{\sin \widehat{C}} \tag{3.2.48}$$

by plugging in the given values, which yields:

$$\frac{10}{\sin \widehat{A}} = \frac{b}{\sin \widehat{B}} = \frac{6}{\sin 80^\circ}.$$

Using the proportion equation $\frac{10}{\sin \widehat{A}} = \frac{6}{\sin 80^\circ}$, we can solve by cross-multiplication and division, which gives:

$$\sin \widehat{A} = \frac{10 \sin 80^\circ}{6} \simeq 1.641346255. \tag{3.2.49}$$

Since the value we got is greater than 1, the equation $\sin \widehat{A} = 1.641346255$ cannot be solved, so our problem has *no solution*.

Example 3.2.11. Suppose we want to solve triangle $\triangle ABC$, given $a = 10$, $b = 8$ (both measured in inches), and $\widehat{B} = 100^\circ$.

I. Since $a > b$, it follows that $\widehat{A} > \widehat{B}$. However, since $\widehat{B}$ is *obtuse*, this would force $\widehat{A}$ to be obtuse, as well, which is clearly impossible. (No triangle can have two obtuse angles!) So we can quickly conclude that this problem has *no solution*.

Example 3.2.12. Suppose we want to solve triangle $\triangle ABC$, given $a = 5$, $b = 5$ (both measured in inches), and $\widehat{B} = 77^\circ$.

I. Since $a = b$, it follows immediately that our triangles is *isosceles*, thus $\widehat{A} = \widehat{B}$, so we immediately get $\widehat{A} = 77^\circ$, so we can skip directly to the final step.

IV. $\widehat{C} = 180^\circ - \widehat{A} - \widehat{B} = 180^\circ - 77^\circ - 77^\circ = 26^\circ$. Now if we go back to the Law of Sines (3.2.34), we also get $\frac{5}{\sin 77^\circ} = \frac{c}{\sin 26^\circ}$, so the value of c is:

$$c = \frac{5 \sin 26^\circ}{\sin 77^\circ} \simeq 2.249510543 \text{ in.} \tag{3.2.50}$$

Conclusion: Our problem has *one* solution:

$$\begin{cases} \widehat{A} & = 77^\circ \\ \widehat{C} & = 26^\circ \\ c & \simeq 2.249510543 \text{ in.} \end{cases}$$

As mentioned earlier, it is also possible to solve the SSA problem using the Law of Cosine. The only drawback of this approach is the fact that it uses the Quadratic Formula:

Solution of the SSA based on Law of Cosine

Given two sides in a triangle, and an angle facing one of the given sides, we solve the triangle as follows.

I. Set up one form of the Law of Cosine (3.2.3), or (3.2.4), or (3.2.5), in which the given angle contributes. Treat this identity as an *equation in which the unknown is the missing side*.

II. Solve the equation set up in Step I. (The equation will always be quadratic.) Collect *both* solutions. If one solution is negative, eliminate it!

III. With each one of the solution(s) found in Step II, we have an SSS Problem. Solve each one of these problems as outlined earlier:
(i) Find one of the missing angles, using the Law of Cosine.
(ii) Find the third angle by subtracting the two known angles from $180°$.

Example 3.2.13. Suppose we want to solve the same triangle $\triangle ABC$ from Example 3.2.8, which has $a = 10$, $b = 8$ (both measured in inches), and $\widehat{B} = 40°$. When working this Example using the above method, we will not finish it completely: we will only illustrate the first two steps (to confirm we are on the right track), and leave the third one to the reader.

I. The equation given by the Law of Cosine, in which $\widehat{B}$ participates is:

$$b^2 = a^2 + c^2 - 2a\,c\cos\widehat{B}.$$

We treat this as an equation which has c as the unknown. By subtracting b^2 and replacing all given values, we can write the above as a quadratic equation

$$10^2 + c^2 - 2 \cdot 10\,c\cos 40° - 8^2 = 0,$$

which reads:

$$c^2 + (-20\cos 40°)c + 36 = 0.$$

II. Using the Quadratic Formula, the above equation has two solutions:

$$c = \frac{-(-20\cos 40°) \pm \sqrt{(-20\cos 40°)^2 - 4 \cdot 36}}{2 \cdot 1} = 10\cos 40° \pm \sqrt{100\cos^2 40° - 36} \simeq$$
$$\simeq 7.660444431 \pm 4.762605262.$$

These two possible values of c clearly match the values we found in our calculations (3.2.39) and (3.2.41) from Example 3.2.8.

Exercises

In each one of Exercises 1-6 you are asked to find the area of the triangle $\triangle ABC$, based on the given information. (Do not solve the triangle!)

1. $a = 12$ cm, $b = 14$ cm, $c = 7$ cm. Use **exact values**, then round to nearest 0.001.

2. $a = 10.3$ cm, $b = 8.2$ cm, $c = 5.4$ cm. Round to nearest 0.001.

3. $a = 7$ cm, $b = 6$ cm, $\widehat{C} = 40°$. Round to nearest 0.001.

4. $a = 12$ cm, $\widehat{B} = 60°$, $c = 4$ cm. Use **exact value**, then round to nearest 0.001.

5. $a = 20$ cm, $\widehat{B} = 30°$, $c = 40$ cm. Use **exact values**.

6. $\widehat{A} = 135°$, $b = 12$ cm, $c = 16$ cm. Use **exact value**, then round to nearest 0.001.

In each one of Exercises 7-18 you are asked to solve the triangle $\triangle ABC$, based on the given information. Round you answers to the nearest 0.001.

7. $a = 10$ cm, $b = 9$ cm, $c = 7$ cm.

8. $a = 10.5$ cm, $b = 9.2$ cm, $c = 5$ cm.

9. $a = 1.7$ cm, $b = 5$ cm, $c = 3.5$ cm.

10. $a = 20$ cm, $b = 8$ cm, $c = 11$ cm.

11. $a = 7$ cm, $b = 6$ cm, $\widehat{C} = 40°$.

12. $a = 12$ cm, $\widehat{B} = 50°$, $c = 14$ cm.

13. $a = 22$ cm, $\widehat{B} = 130°$, $c = 17$ cm.

14. $\widehat{A} = 105°$, $b = 13$ cm, $c = 15$ cm.

15. $\widehat{A} = 35°$, $b = 13$ cm, $\widehat{C} = 25°$.

16. $\widehat{A} = 47°$, $\widehat{B} = 29°$, $c = 8$ cm.

17. $\widehat{A} = 43°$, $\widehat{B} = 75°$, $a = 12$ cm.

18. $a = 20$ cm, $\widehat{B} = 59°$, $\widehat{C} = 100°$.

In each one of Exercises 19-22 you are asked to solve the triangle $\triangle ABC$, based on the given information. Each one of these problems may have two, one, or no solutions! In case of two solutions, write both of them. Round you answers to the nearest 0.001.

19. $a = 4$ cm, $b = 7$ cm, $\widehat{A} = 27°$.

20. $a = 4$ cm, $b = 7$ cm, $\widehat{A} = 37°$.

21. $a = 4$ cm, $b = 7$ cm, $\widehat{A} = 47°$.

22. $a = 7$ cm, $b = 4$ cm, $\widehat{A} = 57°$.

In each one of Exercises 23-26 you are asked to find the area of the triangle $\triangle ABC$, based on the given information. (Do not solve the triangle!)

23. $\widehat{A} = 35°$, $b = 13$ cm, $\widehat{C} = 45°$. Round to nearest 0.001.

24. $\widehat{A} = 44°$, $\widehat{B} = 55°$, $c = 10$ cm. Round to nearest 0.001.

25. $\widehat{A} = 83°$, $\widehat{B} = 65°$, $a = 15$ cm. Round to nearest 0.001.

26. $a = 25$ cm, $\widehat{B} = 29°$, $\widehat{C} = 107°$. Round to nearest 0.001.

3.3 Polar Coordinates

In this section we introduce a new way to describe points in the plane, which essentially incorporates the key elements (direction and magnitude) of their position vectors in a more efficient manner, which is very useful especially in navigation and GPS.

One More Look at Vector Direction

We have discussed the notion of *direction* several times in this course. Our last treatment occurred in Section 3.1, when we agreed to identify this notion with *unit vectors*. Among all unit vectors, a particular one is of most importance to us:

$$\vec{\mathbf{u}}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The main reason we consider this unit vector important is the fact that the coordinates of any vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ in the plane can be obtained using the dot and skew product with $\vec{\mathbf{u}}_0$:

$$x = \vec{\mathbf{u}}_0 \bullet \vec{\mathbf{v}}$$
$$y = \vec{\mathbf{u}}_0 \wedge \vec{\mathbf{v}}.$$

If $\vec{\mathbf{v}}$ is *non-zero*, then using the *trigonometric formulas* of the dot and skew product, we know that these coordinates can also be computed using a certain angle, which from now on will be referred to using a different name, as follows.

The **standard direction angle** of a non-zero vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ is the *unique value τ in the interval $(-\pi, \pi]$*, that satisfies the equalities

$$x = \left\|\vec{\mathbf{v}}\right\| \cos\tau,$$
$$y = \left\|\vec{\mathbf{v}}\right\| \sin\tau.$$

Equivalently (see Section 3.1), the angle τ is the *turning angle of $\vec{\mathbf{u}}_0$ over $\vec{\mathbf{v}}$*.

CLARIFICATION. Any unit vector is completely characterized by its standard direction angle, because the above formulas force the unit vector to be of the form:

$$\vec{\mathbf{u}}_\tau = \begin{bmatrix} \cos\tau \\ \sin\tau \end{bmatrix}.$$

(The notation $\vec{\mathbf{u}}_\tau$ is consistent with the one used for the vector $\vec{\mathbf{u}}_0$, since $\cos 0 = 1$ and $\sin 0 = 0$.)

Using this notation, saying that an arbitrary non-zero vector $\vec{\mathbf{v}}$ has τ as its standard direction angle is the same as saying that *the unit direction vector of* $\vec{\mathbf{v}}$ *coincides with* $\vec{\mathbf{u}}_\tau$. Equivalently, this is the same as saying that we have the equality:

$$\vec{\mathbf{v}} = \|\vec{\mathbf{v}}\| \, \vec{\mathbf{u}}_\tau. \tag{3.3.1}$$

If we write our non-zero vector in coordinates $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$, then its magnitude is, of course, $\|\vec{\mathbf{v}}\| = \sqrt{x^2 + y^2}$, and its unit direction vector is

$$\vec{\mathbf{u}} = \frac{1}{\|\vec{\mathbf{v}}\|}\vec{\mathbf{v}} = \frac{1}{\sqrt{x^2+y^2}}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{x}{\sqrt{x^2+y^2}} \\ \dfrac{y}{\sqrt{x^2+y^2}} \end{bmatrix},$$

and then all we are saying can be summarized as follows.

The standard direction angle of a non-zero vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ *is the unique value* τ *in the interval* $(-\pi, \pi]$, *that satisfies the equalities:*

$$\cos\tau = \frac{x}{\sqrt{x^2+y^2}},$$
$$\sin\tau = \frac{y}{\sqrt{x^2+y^2}}.$$

Using the Vector Angle Formulas from Section 3.1, the standard direction angle can be computed as follows.

Standard Direction Angle Formulas

The standard direction angle τ of a *non-zero* vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ is given by:

$$\tau = (\text{sign of } y)\arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right)$$

(We agree that, when $y = 0$, the above sign is $+$.)

Example 3.3.1. Suppose two ships started from the same point on the map, and they navigated as follows. Ship 1 traveled 300 miles in the N$35°$E direction, and ship 2 traveled 400 miles in the S$40°$E direction. We wish to find the exact location of ship 2 in relation to ship 1, at the end of their respective trips.

The situation described in this example is depicted in the figure below, which shows the position vectors $\vec{\mathbf{v}}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{\mathbf{v}}_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ of the two ships at the end of their trips.

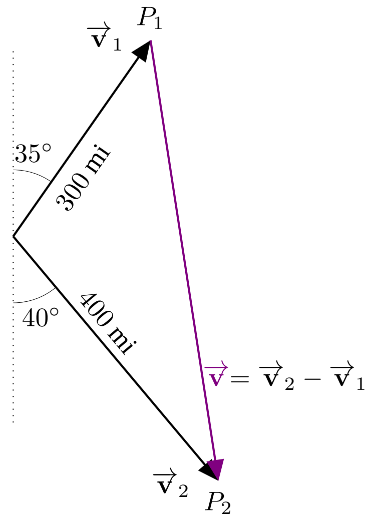


Figure 3.3.1

The vector that will give us the location of ship 2 relative to ship 1 will be the difference

$$\vec{\mathbf{v}} = \vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1,$$

so what we need to do is to analyze the vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$, and ultimately compute their coordinates.

When we consider the position vector $\vec{\mathbf{v}}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, we know that $\sqrt{x_1^2 + y_1^2} = \|\vec{\mathbf{v}}_1\| = 300.$

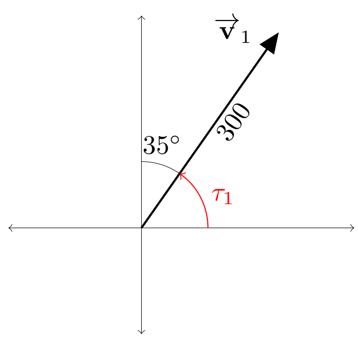


Figure 3.3.2

Furthermore, by the bearing notation features and the above diagram, it follows that the standard direction angle τ_1 of the vector $\vec{\mathbf{v}}_1$ is simply given by $\tau_1 = 55°$. Using (3.3.1) it follows that our vector can be written as

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \vec{\mathbf{v}}_1 = \|\vec{\mathbf{v}}_1\| \vec{\mathbf{u}}_{55°} = 300 \begin{bmatrix} \cos 55° \\ \sin 55° \end{bmatrix} = \begin{bmatrix} 300\cos 55° \\ 300\sin 55° \end{bmatrix}.$$

When we consider the position vector $\vec{\mathbf{v}}_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, we know that $\sqrt{x_2^2 + y_2^2} = \|\vec{\mathbf{v}}_2\| = 400.$

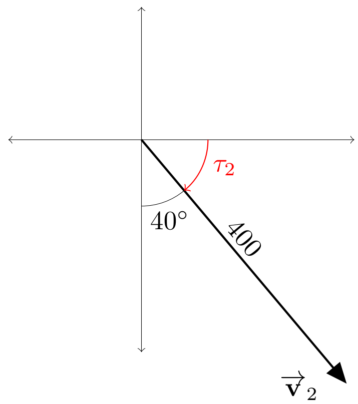


Figure 3.3.3

Furthermore, by the bearing notation features and the above diagram, it follows that the standard direction angle τ_2 of the vector $\vec{\mathbf{v}}_2$ is simply given by $\tau_2 = -50°$. Using (3.3.1) it follows that our vector can be written as

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \vec{\mathbf{v}}_2 = \|\vec{\mathbf{v}}_2\| \vec{\mathbf{u}}_{-50°} = 400 \begin{bmatrix} \cos(-50°) \\ \sin(-50°) \end{bmatrix} = \begin{bmatrix} 400\cos(-50°) \\ 400\sin(-50°) \end{bmatrix}.$$

With the vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ now computed, the vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ will be given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_2 \end{bmatrix} = \begin{bmatrix} 400\cos(-50°) - 300\cos 55° \\ 400\sin(-50°) - 300\sin 55° \end{bmatrix} \simeq \begin{bmatrix} 85.04211297 \\ -552.1633905 \end{bmatrix}.$$

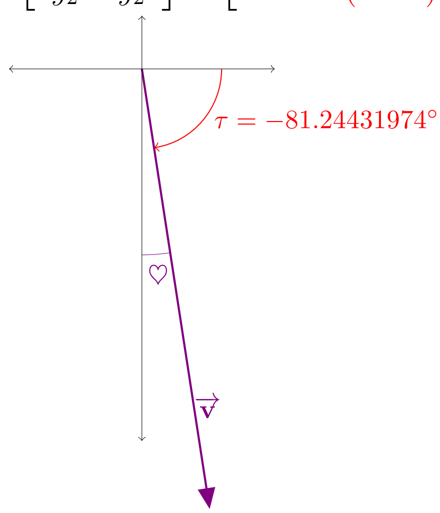


Figure 3.3.4

The magnitude of the vector $\vec{\mathbf{v}}$ is:

$$\|\vec{\mathbf{v}}\| = \sqrt{x^2 + y^2} \simeq \sqrt{85.04211297^2 + (-552.1633905)^2} \simeq 558.6739396.$$

Since $y < 0$, the direction angle of the vector $\vec{\mathbf{v}}$ is:

$$\tau = -\arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right) = -\arccos\left(\frac{85.04211297}{558.6739396}\right) \simeq -81.24431974°.$$

Since this angle is in the fourth quadrant, the bearing will be of the form S♡E, where $\heartsuit \simeq 90^\circ - 81.24431974^\circ \simeq 8.75568026^\circ$. So our conclusion is as follows: the position of ship 2 is 558.6739396 miles away from ship 1, in the $\mathrm{S}8.75568026^\circ\mathrm{E}$ direction.

Polar Coordinates

The main consequence of "playing" with direction angles is the fact that, given any pair (x, y) of real numbers, we can always find an angle τ, such that

$$\begin{cases} x &= \left(\sqrt{x^2+y^2}\right)\cos\tau \\ y &= \left(\sqrt{x^2+y^2}\right)\sin\tau \end{cases}$$

Furthermore, if $(x, y) \neq (0, 0)$, we also learned that, if demand that τ be in the interval $(-\pi, \pi]$, then τ is in fact unique, because it must coincide with the *standard direction angle of the vector* $\begin{bmatrix} x \\ y \end{bmatrix}$.

If we do not like vectors, then based on what we learned in Section 2.2, we can also characterize τ as the unique value in the interval $(-\pi, \pi]$ that represents the *measure of a rotation angle in standard position, that has the point $P(x, y)$ on its terminal side*.

At this point, we would like to loosen some of the above restrictions (and forget about vectors, as well!) and as a result we introduce the following terminology.

> For a point $P(x, y)$ in the coordinate plane, a **polar coordinate representation of P** is an ordered pair $(\boldsymbol{r}, \theta)$ of real numbers, satisfying:
>
> $$\begin{cases} x &= \boldsymbol{r}\cos\theta \\ y &= \boldsymbol{r}\sin\theta \end{cases} \tag{3.3.2}$$

CLARIFICATIONS. If our point $P(x, y)$ is *distinct from the origin* $(0, 0)$, then we always have a particular polar coordinate representation of it by setting

$$\begin{cases} \boldsymbol{r} &= \sqrt{x^2+y^2} \\ \theta &= \tau, \text{ the standard direction angle of the vector } \begin{bmatrix} x \\ y \end{bmatrix} \end{cases} \tag{3.3.3}$$

which we refer to as the **standard** polar coordinate representation.

Each point has many polar coordinate representations. This is due to the fact that both cosine and sine functions are periodic with period 2π, so if we have a particular polar coordinate representation $(\boldsymbol{r}, \theta)$ of P, then so are all pairs of the form

$$(\boldsymbol{r}, \theta + 2n\pi),\ n \text{ integer}.$$

Furthermore, the *origin* O, which mathematicians call the **pole**, has a huge list of polar coordinate representations: $\boldsymbol{r} = 0$ and $\theta =$ *any* number! (More on how non-unique polar coordinate representations are will be discussed a little later.)

ADDITIONAL CLARIFICATION. Given a polar coordinate pair $(\boldsymbol{r}, \theta)$, there are two ways of locating the point P represented by it. The easiest way is, of course, to compute the coordinates of P using the definition (3.3.2).

However, the point P can also be constructed *geometrically*, as follows.

I. Construct the (unique) *rotation angle in standard position, which has θ as its rotation measure*.

II. Depending on the *sign of $\boldsymbol{r}$*, pick up P as follows.

(A) If $\boldsymbol{r} = 0$, pick up P to be the *pole O*.

(B) If $\boldsymbol{r} > 0$, pick up P to be the unique point on *terminal side*, with $\text{dist}(P, O) = \boldsymbol{r}$.

(C) If $\boldsymbol{r} < 0$, pick up P to be the unique point on *opposite of terminal side*, with $\text{dist}(P, O) = -\boldsymbol{r}$.

In either case, the distance from P to the pole O is:

$$\text{dist}(P, O) = |\boldsymbol{r}|. \tag{3.3.4}$$

The *geometric* construction outlined above is useful only when we need to plot points given in polar coordinates. When using this construction for plotting points by hand, we use a "polar graphing" paper, which consists of a grid made up of *circles* centered at the origin, and *rays* emanating from the origin, as seen in the Example below.

Example 3.3.2. Consider the following points represented by the following polar coordinate pairs: $A\left(5, \dfrac{\pi}{4}\right)$, $B\left(4, \dfrac{3\pi}{2}\right)$, $C(-6, 0)$, $D(6, \pi)$. Assume we are asked to plot these points (geometrically), as well as to locate them by means of their usual *rectangular* coordinates.

For the plotting problem, use the "polar graphing paper" shown below, in which the circles have integer radii $1, 2, 3, \ldots$, and the rays mark directions that are multiples of $\frac{\pi}{12}$:

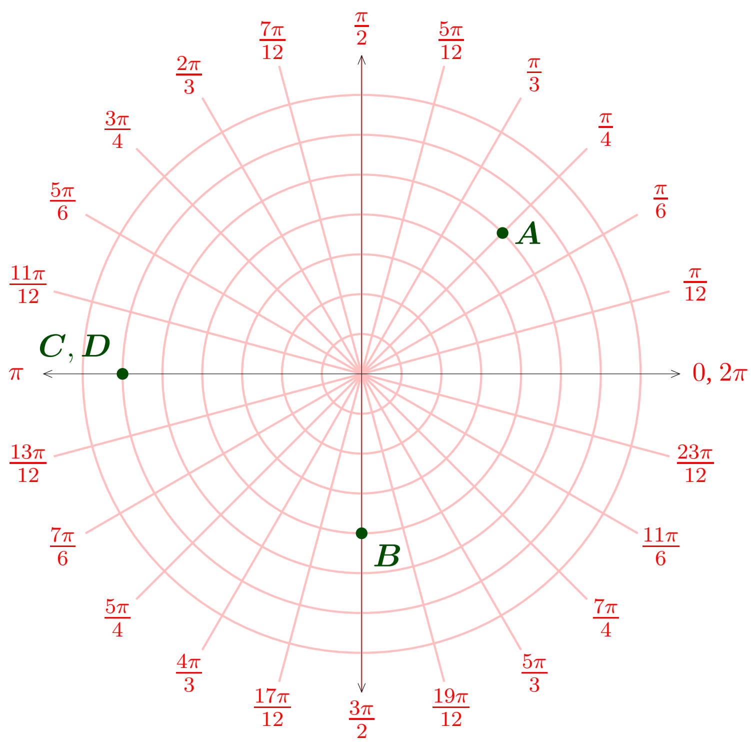


Figure 3.3.5

Geometrically, the point A sits on the ray $\theta = \frac{\pi}{4}$ at distance 5 from the origin. Similarly, the point B sits on the ray $\theta = \frac{3\pi}{2}$ – which is nothing else but the *negative y-axis* – at distance 4 from the origin. Next, the point C sits on the ray $\theta = \pi$ – which is nothing else but the *negative x-axis* – at distance 6 from the origin. Finally, since the point D has $r = -6$, it sits on the ray opposite to the ray $\theta = 0$, at distance 6 from the origin. Since the ray the point D sits on is again the *negative x-axis*, it follows that D coincides with C.

If one want to located our points in rectangular coordinates, we can directly use formulas (3.3.2).

(a) The rectangular coordinates of the point A are $x = 5\cos\dfrac{\pi}{4} = \dfrac{5}{\sqrt{2}}$ and $y = 5\sin\dfrac{\pi}{4} = \dfrac{5}{\sqrt{2}}$.

(b) The rectangular coordinates of the point B are $x = 4\cos\dfrac{3\pi}{2} = 4\cdot 0 = 0$ and $y = 4\sin\dfrac{3\pi}{2} = 4\cdot(-1) = -5$.

(c) The rectangular coordinates of the point C (and D) are $x = 6\cos\pi = 6\cdot(-1) = -6$ and $y = 4\sin\pi = 6\cdot 0 = 0$.

It is worth pointing out that, the calculations for B and C (and D) are not really necessary, because their rectangular coordinates can be easily obtained from their geometric construction.

Rectangular-to-Polar Conversion

As we have already seen in Example 3.3.2 above, the process of converting *from polar to rectangular coordinates* is pretty straightforward: we simply use the definition (3.3.2). The remaining question is then the conversion in reverse: given the point $P(x, y)$ presented in rectangular coordinates, how do we find all(!) its polar coordinate representations? As we have already seen, this task is a complicated one, because there are many possible answers.

Rectangular-to-Polar Coordinate Conversion Method

Given a point $P(x, y)$, in order to find its polar coordinate representations, we proceed as follows.

I. If *P coincides with the pole*, that is, $x = y = 0$, then all its polar coordinate representations are of the form: $\boldsymbol{r} = 0$, $\theta =$ any number.

II. If P is *distinct from the pole*, we first find its *standard* polar coordinate representation $(\sqrt{x^2+y^2}, \tau)$, given by (3.3.3), and then all polar coordinate representations of P will be of the form

(a) $(\sqrt{x^2+y^2}, \tau + 2n\pi)$, n integer, or

(b) $(-\sqrt{x^2+y^2}, \tau + \pi + 2n\pi)$, n integer.

In this case, no matter what polar coordinate representation $(\boldsymbol{r}, \theta)$ we compute, the following identities always hold:

$$\begin{cases} \boldsymbol{r}^2 &= x^2 + y^2 \\ \cos\theta &= \dfrac{x}{\boldsymbol{r}} \\ \sin\theta &= \dfrac{y}{\boldsymbol{r}} \end{cases} \tag{3.3.5}$$

In particular, whether we are in case I or case II, we always have the equality:

$$\boldsymbol{r} = \pm\sqrt{x^2+y^2}. \tag{3.3.6}$$

CLARIFICATIONS. The reason for the matching of π with the "$-$" sign in (b) is the fact that adding π changes the sign of both sine and cosine.

The identities (3.3.5) are reminiscent of the well known formulas from Section 2.2, which we used for computing trigonometric functions in coordinates, with one big addition: *the number $\boldsymbol{r}$ is* *allowed to be negative!*

Example 3.3.3. Start with the point $P(-1, \sqrt{3})$ and let us find all its polar coordinate representations $(\boldsymbol{r}, \theta)$ with $-3\pi \le \theta \le 3\pi$.

Since the quantity $\sqrt{x^2 + y^2}$ is needed, no matter what we do, we start with its computation:

$$\sqrt{x^2 + y^2} = \sqrt{(-1)^2 + (\sqrt{3})^2} = \sqrt{1+3} = 2. \tag{3.3.7}$$

Next we compute the angle τ from the *standard* polar coordinate representation, which is computed using the Standard Direction Angle Formulas. Since $y = \sqrt{3} > 0$, it follows that

$$\tau = \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right) = \arccos\left(\frac{-1}{2}\right) = \pi - \arccos\left(\frac{1}{2}\right) = \pi - \frac{\pi}{3} = \frac{2\pi}{3}.$$

Using the above method, we conclude that all polar coordinate representations of P will be of the form

(a) $\boldsymbol{r} = 2$ and $\theta = \frac{2\pi}{3} + 2n\pi$, n integer, or

(b) $\boldsymbol{r} = -2$ and $\theta = \frac{2\pi}{3} + \pi + 2n\pi$, n integer.

From the list (a) only $n = -1, 0, 1$ will work; From the list (b) only $n = -2, -1, 0$ will work. This means that all desired polar coordinate representations of our point $P(-1, \sqrt{3})$ are (according to our two lists):

(a) $\left(2, \frac{2\pi}{3} - 2\pi\right) = \left(2, -\frac{4\pi}{3}\right)$, $\left(2, \frac{2\pi}{3}\right)$, $\left(2, \frac{2\pi}{3} + 2\pi\right) = \left(2, \frac{8\pi}{3}\right)$,

(b) $\left(-2, \frac{2\pi}{3} + \pi - 4\pi\right) = \left(-2, -\frac{7\pi}{3}\right)$, $\left(-2, \frac{2\pi}{3} + \pi - 2\pi\right) = \left(-2, -\frac{\pi}{3}\right)$, and $\left(-2, \frac{2\pi}{3} + \pi\right) = \left(-2, \frac{5\pi}{3}\right)$.

Graphs of Polar Equations

Polar coordinates are useful for describing certain curves by nice and concise equations, which are of the form:

$$\textit{(left) expression in } \boldsymbol{r} \textit{ and } \theta = \textit{(right) expression in } \boldsymbol{r} \textit{ and } \theta$$

which we refer to as **polar equations**.

As is the case with *graphs of functions*, a particularly useful polar equation is one that looks like:

$$\boldsymbol{r} = \textit{function of } \theta,$$

which can be easily plotted with the help of a (good) graphing calculator. However, when a graphing calculator with polar graphing capabilities is not available, certain easy polar equations can also be graphed by hand, by the plotting points using the geometric technique illustrated in Example 3.3.2.

Example 3.3.4. Suppose we are asked to graph the polar equation

$$\boldsymbol{r} = 2 + 2\cos\theta.$$

When attempting to graph this polar equation by hand, it is a good idea to start off by graphing the same equation in *rectangular* coordinates. Since the function $2 + 2\cos\theta$ is periodic, with period 2π, it suffices to graph only the part corresponding to the interval $[0, 2\pi]$:

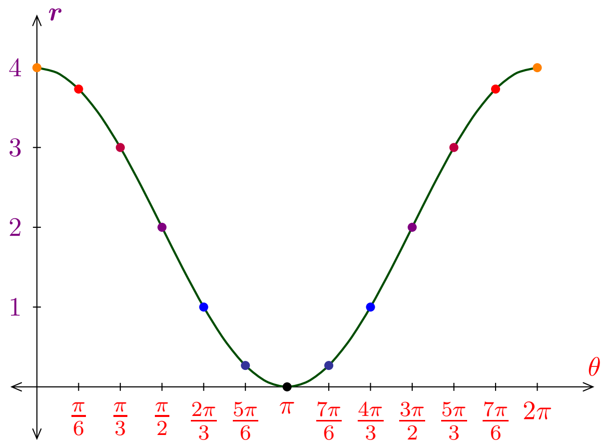


Figure 3.3.6

As the above graph suggests, as θ *increases from* 0 *to* π, the corresponding point on the curve given in polar coordinates *gets closer and closer to the origin*. At $\theta = 0$, the point is 4 units away from the origin; at $\theta = \pi/3$, the point is 3 units away from the origin; at $\theta = \pi/2$, the point is 2 units away from the origin; at $\theta = 2\pi/3$, the point is 1 unit away from the origin; finally, at $\theta = \pi$, the point is at the origin. On the other hand, as *increases from* π *to* 2π, the corresponding point on the curve given in polar coordinates *gets further and further away from the origin*. Based on these findings, on "polar graphing paper" our curve will look like:

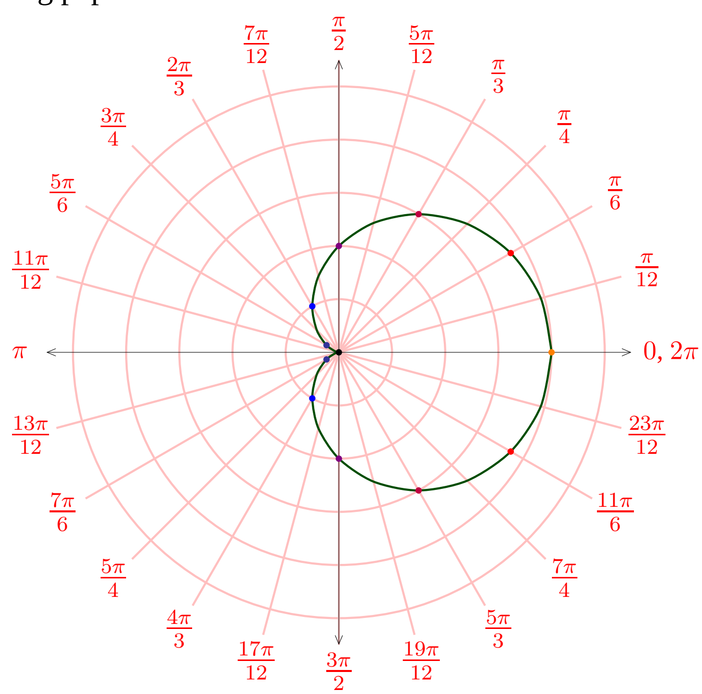


Figure 3.3.7

The graph shown above is a special case of a *Limaçon*. Other "nice" polar equations are given below.

Example 3.3.5. Below is a list of some interesting polar equations that yield very beautiful pictures:

(i) Archimede's Spiral: $\boldsymbol{r} = \theta$.

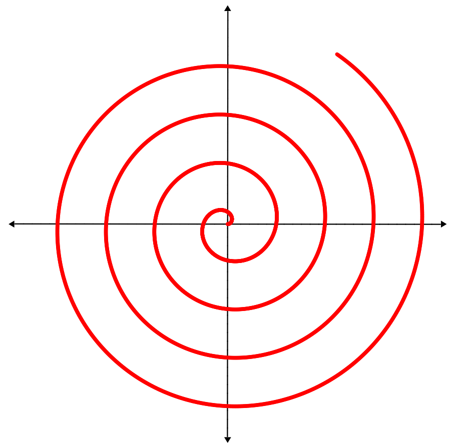


Figure 3.3.8

(ii) Limaçons: $\boldsymbol{r} = a(k + \cos\theta)$, $a > 0$. The pictures shown below depict the corresponding graphs for various values of k.

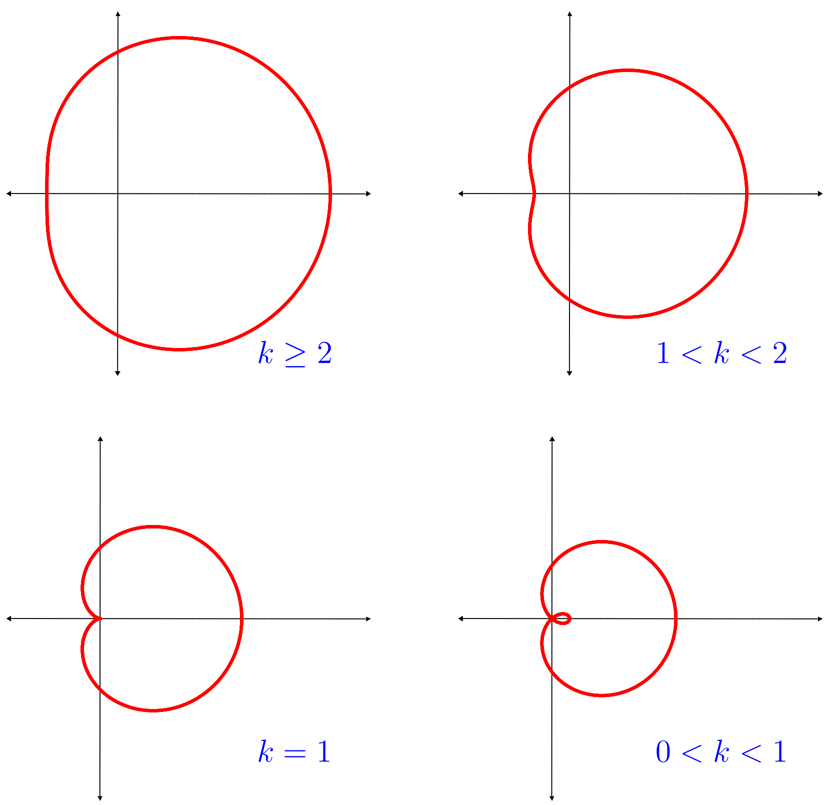


Figure 3.3.9

(iii) Four-leaf clover: $\boldsymbol{r} = a\cos(2\theta)$, $a > 0$.

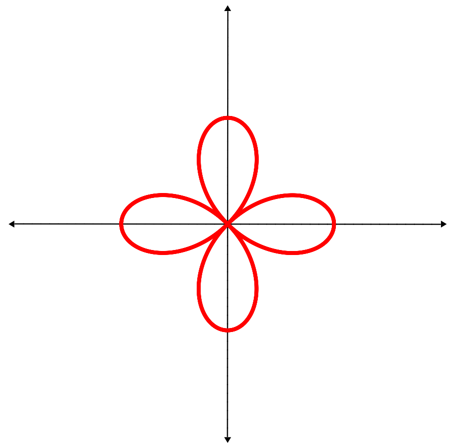


Figure 3.3.10

(iv) You name this! $r = \sin(2\theta/5)$

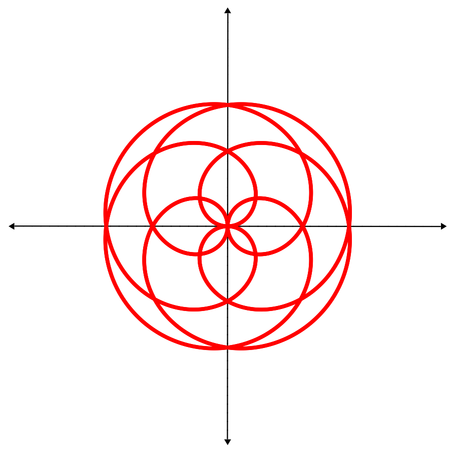


Figure 3.3.11

▶ You are encouraged to "play" with these graphs, especially with Limaçons, where you may plug in other values for k, including negative values. You may also replace $\cos$ with $-\cos$, $\sin$, or $-\sin$. More examples of Limaçons appear in Exercises 4-10.

The whole point about using polar equations is this: Although it is possible to write equations in *rectangular coordinates* for curves such as those given in the above Example, they will be extremely complicated! On the other hand, their polar equations are particularly nice, and they allow us to gather some geometric information quite effectively.

However, there are curves for which *both* polar equations and rectangular coordinate equations are not too complicated, in which case switching from one form to another can benefit us. Among such curves, we recognize some familiar ones, such as lines and (certain) circles. When we wish to convert equations from one format to another, we should always use the following general principles.

A. Rectangular-to-Polar Equation Conversion: To convert from an equation in x and y to a polar equation, we simply replace $x = r\cos\theta$ and $y = r\sin\theta$.

B. Polar-to-Rectangular Equation Conversion: To convert a polar equation to an equation in x and y, do the following (in this order!):

(i) Replace every occurrence of $\cos\theta$ with $\dfrac{x}{r}$, and every occurrence of $\sin\theta$ with $\dfrac{y}{r}$. If these replacements eliminate θ, the resulting equation is an equation in x, y and r only.

(ii) Simplify the equation, then replace every occurrence of r^2 by $x^2 + y^2$. If r has not been eliminated, replace it with $\pm\sqrt{x^2+y^2}$.

C. Decide Pole Status: For either conversion, when handling equations, make sure you analyze the case when $r = 0$, which represents the pole (that is, the origin). Make sure you decide whether the pole is or is not a point on the given curve.

Example 3.3.6. Suppose we want to find the polar equation of the line

$$3x - 4y = 5.$$

Upon replacing $x = r\cos\theta$ and $y = r\sin\theta$, we can write our equation in the form

$$3r\cos\theta - 4r\sin\theta = 5,$$

which we can factor as

$$r(3\cos\theta - 4\sin\theta) = 5. \tag{3.3.8}$$

We may think then of dividing to get

$$r = \frac{5}{3\cos\theta - 4\sin\theta}.$$

Is this safe? The only issue we have to be concerned with is whether this transformation looses some points. All we have to observe here is the fact that (3.3.8) clearly excludes the possibility that

$$3\cos\theta - 4\sin\theta = 0,$$

so dividing by $3\cos\theta - 4\sin\theta$ is indeed safe, that is, it will not "loose" any points.

Example 3.3.7. Suppose we want to rectangular equation of

$$r = 6\cos\theta + 14\sin\theta.$$

We start off by replacing $\cos\theta = \dfrac{x}{r}$, and $\sin\theta = \dfrac{y}{r}$, so our new equation will be

$$r = \frac{6x}{r} + \frac{14y}{r}, \tag{3.3.9}$$

which we can re-write as:

$$r = \frac{6x + 14y}{r}. \tag{3.3.10}$$

We can cross-multiply to get:

$$r^2 = 6x + 14y, \tag{3.3.11}$$

which finally becomes (by replacing $r^2 = x^2 + y^2$):

$$x^2 + y^2 = 6x + 14y. \tag{3.3.12}$$

Have we been careless here? Well, when we made our first replacement, which lead us to (3.3.9) we were assuming that $r \neq 0$, which means that we "lost" the pole, that is the origin, which has coordinates $x = y = 0$. However, when we look at the final equation (3.3.12), we see that the pole is "taken care of," because $x = y = 0$ do satisfy (3.3.12).

Exercises

1. A ship leaves port A and must reach port B, which is located 1000 miles from A in the S$70°$E direction. In order to avoid a hurricane, the ship started by sailing 600 miles in the N$50°$E direction, reaching point C.

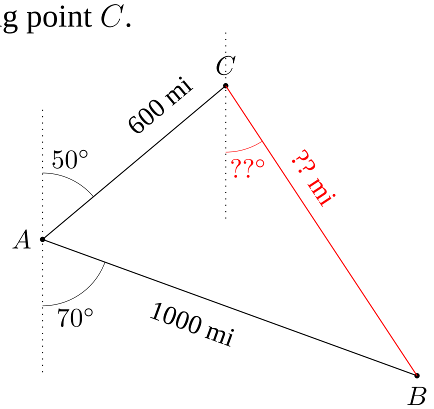


Figure 3.3.12

How far is the ship from the destination port B? In what direction must the ship sail from point C in order to reach its destination?

2. A tornado was spotted 10 miles north-east of town A, and is moving in the S$70°$E direction at 40 miles per hour. Suppose town B is located 30 miles east of A.

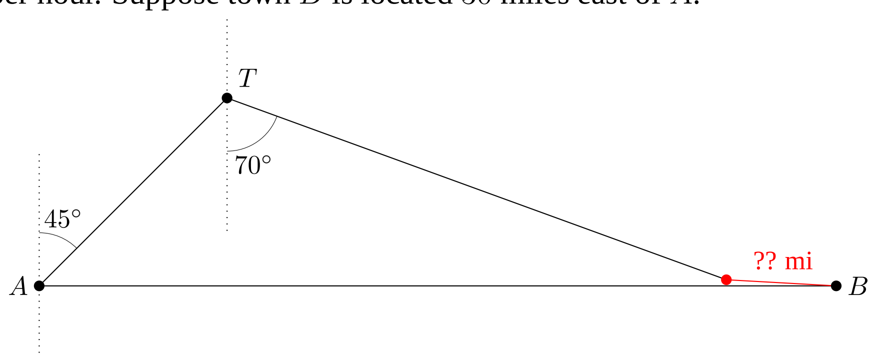


Figure 3.3.13

How close will be the tornado from town B in half an hour? (Here north-east means of course the N$45°$E direction.)

3. Suppose we are in the desert and we drive 50 miles in the N$30°$E direction (and reach point A), then we drive 60 miles in the S$30°$E direction (and reach point B), and continue 30 miles in the N$20°$W direction (and reach point C).

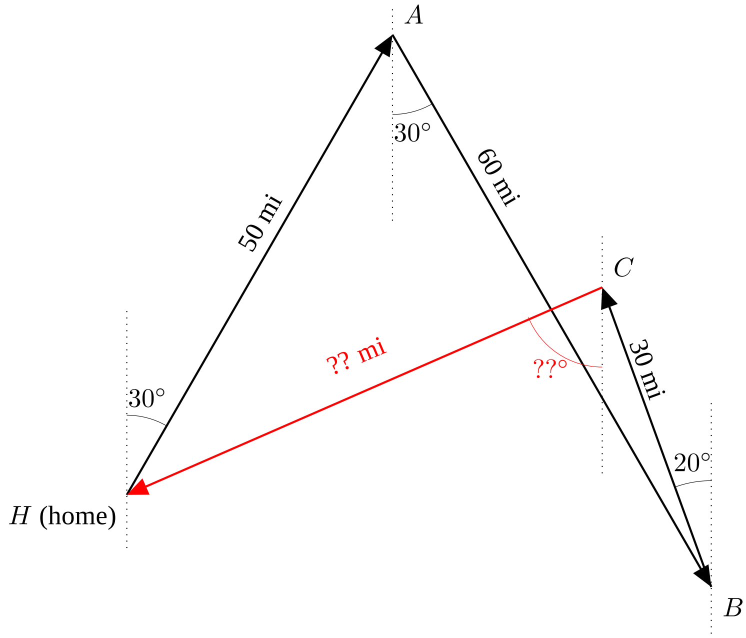


Figure 3.3.14

How far are we from home (our starting point)? In what direction do we need to drive to get home?

In each one of Exercises 4-10 you are given a polar equation of the form

$$r = f(\theta),$$

and as in Example 3.3.4, you are asked to do the following.

(i) Complete a table of values for the function f, for θ of the form $\frac{m\pi}{6}$, $m = 0, 1, 2, \ldots, 12$. (Use a calculator.)

(ii) Sketch the graph of $r = f(t)$ in normal coordinates.

(iii) Using the above information as a guide, plot the points given in (i) geometrically on a "polar graphing paper," then sketch the entire graph. You can transfer the picture from your graphing calculator.

All these equations represent the so-called *limaçons*. Compare your graphs to the models given in Example 3.3.5.(i).

4. $r = 1 + \cos\theta$.

5. $r = 1 - \cos\theta$.

6. $r = 2 + 2\sin\theta$.

7. $r = 1 - \sin\theta$.

8. $r = 1 + 2\cos\theta$.

9. $r = 3 + 2\cos\theta$.

10. $r = 4 + \cos\theta$.

11. Find all polar coordinate representations $(\boldsymbol{r}, \theta)$ of the point $A(-2, 0)$ with $-4\pi \leq \theta \leq 4\pi$. Use **exact** values.

12. Find all polar coordinate representations $(\boldsymbol{r}, \theta)$ of the point $B(-2, 2)$ with $-2\pi \leq \theta \leq 2\pi$. Use **exact** values.

13. Find all polar coordinate representations $(\boldsymbol{r}, \theta)$ of the point $P(-2, -2)$ with $-3\pi \leq \theta \leq 3\pi$. Use **exact** values.

In Exercises 14-18 you are asked to find the polar equation of the curve, given its rectangular coordinate equation. Whenever possible, write the polar equation in the form

$$r = f(\theta),$$

14. $x^2 + y^2 = 16$

15. $x^2 = y^2 - y$.

16. $3x - 7y = 10$.

17. $y = 4x^2$.

18. $x = 4y^2$.

In Exercises 19-24 you are asked to find the rectangular coordinate equation of the curve, given its polar equation.

19. $r = \cos\theta$

20. $r = 3\sin\theta$.

21. $r = 4\cos\theta - 5\sin\theta$.

22. $r = \dfrac{1}{\cos\theta + 3\sin\theta}$.

23. $r\sec\theta = 2$.

24. $r(\cos\theta - 2r\sin^2\theta) = 1$.

Chapter 4

Trigonometry for Combinations of Angles

In this Chapter we learn how to handle combinations of angles, such as sums, differences, doubles and halves.

4.1 Sums and Differences of Angles

In this section we introduce several key formulas that allow us to compute the trigonometric functions of the sum and the difference of two angles.

Before you begin this section, please read Appendix B, which reviews *matrix arithmetic*.

The Angle Difference Formulas for Sine and Cosine

The easiest way to compute the trigonometric functions of an *angle difference* $\alpha-\beta$ is to realize it as a *turning angle*, which can be done as follows:

> *Up to a multiple of 2π, the difference $\alpha-\beta$ coincides with the turning angle of the unit vector $\vec{\mathbf{u}}_\beta = \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}$ over the unit vector $\vec{\mathbf{u}}_\alpha = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$.*

CLARIFICATION. If τ denotes the turning angle of $\vec{\mathbf{u}}_\beta$ over $\vec{\mathbf{u}}_\alpha$, then the vector $\vec{\mathbf{u}}_\alpha$ is obtained by *τ-rotating the vector* $\vec{\mathbf{u}}_\beta$. At the same time, if we rotate $\vec{\mathbf{u}}_\beta$ by the angle $\alpha-\beta$, we get the same vector $\vec{\mathbf{u}}_\alpha$. So the τ-rotation and the $(\alpha-\beta)$-rotation transformations are the same(!), thus[13] τ and $\alpha-\beta$ differ by a multiple of 2π.

Using the turning angle formulas from Section 3.1, combined with the fact that $\left\|\vec{\mathbf{u}}_\alpha\right\| = \left\|\vec{\mathbf{u}}_\beta\right\| = 1$, we get the formulas:

$$\cos(\alpha-\beta) = \vec{\mathbf{u}}_\beta \bullet \vec{\mathbf{u}}_\alpha \tag{4.1.1}$$

$$\sin(\alpha-\beta) = \vec{\mathbf{u}}_\beta \wedge \vec{\mathbf{u}}_\alpha \tag{4.1.2}$$

When we use the dot and skew product formulas in coordinates, the above equalities yield:

[13] We can also think this fact as a statement about two rotation angles (both having $\vec{\mathbf{u}}_\beta$ as their *initial* sides), which are **coterminal**.

The Angle Difference Formulas for Cosine and Sine

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta+\sin\alpha\sin\beta; \tag{4.1.3}$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta-\cos\alpha\sin\beta; \tag{4.1.4}$$

As we see, the cosine and sine of a difference of two angles are given by complicated formulas. Equalities like "$\cos(\alpha-\beta) = \cos\alpha - \cos\beta$" or "$\sin(\alpha-\beta) = \sin\alpha - \sin\beta$" are **WRONG!** (They are what we call **false identities**!)

CLARIFICATION. What we see from (4.1.3) and (4.1.4) is that: *in order to be able to compute the trigonometric functions of an angle difference* $\alpha-\beta$, we need to know *all four* values of $\cos\alpha$, $\sin\alpha$, $\cos\beta$, and $\sin\beta$.

Example 4.1.1. Suppose we are given the following information:

(a) α is in Quadrant III, and has $\tan\alpha = \dfrac{3}{4}$.

(b) β is in Quadrant II, and has $\sin\beta = \dfrac{12}{13}$.

Given this information, we wish to find sine and cosine of $\alpha-\beta$, as well as the quadrant where $\alpha-\beta$ sits in.

Based on the Clarification above, our "wish list" is:

$$\cos\alpha = ? \qquad \sin\alpha = ?$$
$$\cos\beta = ? \qquad \sin\beta = \frac{12}{13}.$$

We will fill in our "wish list" using the techniques introduced in Section 2.2. (When we first dealt with these type problems, we learned of two techniques: the algebraic method and the coordinate method. For the benefit of the reader, we will illustrate both.)

To find sine and cosine of α, we use the coordinate method. We build α as an angle in standard position, and we pick a point $P(x,y)$ on the terminal side of α, which we know that must satisfy $\dfrac{y}{x} = \tan\alpha$. Based on the given information on α, a correct choice would be $x=-4$ and $y=-3$, which will then give us $r = \sqrt{x^2+y^2} = \sqrt{(-4)^2+(-3)^2} = \sqrt{16+9} = \sqrt{25} = 5$, from which we quickly get: $\cos\alpha = \dfrac{x}{r} = -\dfrac{4}{5}$ and $\sin\alpha = \dfrac{y}{r} = -\dfrac{3}{5}$.

As for β, we already have its sine, so we can use the algebraic method for finding its cosine:

$$\cos\beta = \pm\sqrt{1-\sin^2\beta} = \pm\sqrt{1-\left(\frac{12}{13}\right)^2} = \pm\sqrt{1-\frac{144}{169}} = \pm\sqrt{\frac{169}{169}-\frac{144}{169}} =$$
$$= \pm\sqrt{\frac{169-144}{169}} = \pm\sqrt{\frac{25}{169}} = \pm\frac{5}{13}.$$

Since β is in Quadrant II, where cosine is *negative*, the correct value is: $\cos\beta = -\dfrac{5}{13}$.

If we go back to our "wish list," we now have all we need:

$$\cos\alpha = -\frac{4}{5} \qquad \sin\alpha = -\frac{3}{5}$$
$$\cos\beta = -\frac{5}{13} \qquad \sin\beta = \frac{12}{13},$$

so using the angle difference formulas we can compute:

$$\begin{aligned}
\cos(\alpha-\beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta = \left(-\frac{4}{5}\right)\left(-\frac{5}{13}\right) + \left(-\frac{3}{5}\right)\left(\frac{12}{13}\right) = \\
&= \frac{20}{65} - \frac{36}{65} = \frac{20-36}{65} = -\frac{16}{65}; \\
\sin(\alpha-\beta) &= \sin\alpha\cos\beta - \cos\alpha\sin\beta = \left(-\frac{3}{5}\right)\left(-\frac{5}{13}\right) - \left(-\frac{4}{5}\right)\left(\frac{12}{13}\right) = \\
&= \frac{15}{65} + \frac{48}{65} = \frac{15+48}{65} = \frac{63}{65}.
\end{aligned}$$

Since both $\cos(\alpha-\beta)$ and $\sin(\alpha-\beta)$ are *negative*, it follows that $\alpha-\beta$ is in Quadrant III.

The Angle Sum Formulas for Sine and Cosine

Based on the Angle Difference Formulas, we can also deal with *sums*. After all, we can always write $\alpha+\beta = \alpha-(-\beta)$. Using the *formulas for negatives*, we know that $\cos(-\beta) = \cos\beta$ and $\sin(-\beta) = -\sin\beta$, and therefore, the cosine and sine of the sum are given by:

The Angle Sum Formulas for Sine and Cosine

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta; \tag{4.1.5}$$
$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta; \tag{4.1.6}$$

Example 4.1.2. Suppose we want to compute the cosine and sine of the angle

$$\psi = \arcsin\left(\frac{8}{17}\right) + \arccos\left(-\frac{24}{25}\right),$$

and find the quadrant ψ sits in.

We start off by naming the first angle $\arcsin\left(\frac{8}{17}\right) = \alpha$, and the second angle $\arccos\left(-\frac{24}{25}\right) = \beta$, so the angle we are interested in is $\psi = \alpha + \beta$.

Using the formulas from Section 2.6 (see Trigonometric Functions of Inverses), we can quickly compute:

$$\begin{aligned}
\sin\alpha &= \sin\left[\arcsin\left(\frac{8}{17}\right)\right] = \frac{8}{17}; \\
\cos\alpha &= \cos\left[\arcsin\left(\frac{8}{17}\right)\right] = \sqrt{1-\left(\frac{8}{17}\right)^2} = \sqrt{1-\frac{64}{289}} = \sqrt{\frac{225}{289}} = \frac{15}{17}; \\
\cos\beta &= \cos\left[\arccos\left(-\frac{24}{25}\right)\right] = -\frac{24}{25}; \\
\sin\beta &= \sin\left[\arccos\left(-\frac{24}{25}\right)\right] = \sqrt{1-\left(-\frac{24}{25}\right)^2} = \sqrt{1-\frac{576}{625}} = \sqrt{\frac{49}{625}} = \frac{7}{25}.
\end{aligned}$$

With our "wish list" filled, using the angle sum formulas we can compute:

$$\begin{aligned}\cos\psi = \cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta &= \left(\frac{15}{17}\right)\left(-\frac{24}{25}\right) - \left(\frac{8}{17}\right)\left(\frac{7}{25}\right) = \\ &= -\frac{360}{425} - \frac{56}{425} = \frac{-360-56}{425} = -\frac{416}{425}; \\ \sin\psi = \sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta &= \left(\frac{8}{17}\right)\left(-\frac{24}{25}\right) + \left(\frac{15}{17}\right)\left(\frac{7}{25}\right) = \\ &= -\frac{192}{425} + \frac{105}{425} = \frac{-192+105}{425} = -\frac{87}{425}.\end{aligned}$$

Since both $\cos\psi$ and $\sin\psi$ are *negative*, it follows that ψ is in Quadrant III.

(Optional) Rotation Matrices

There is a neat way to memorize the angle sum formulas, which employs *rotation matrices*, which we have already encountered several times (see Sections 2.1, 2.2, and 3.1). These special matrices are those of the form:

$$\mathbf{R}_\tau = \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix}. \tag{4.1.7}$$

Their main feature, upon which everything learned about trigonometric functions builds, is the following.

Rotation Principle

Given a rotation measure τ, the vector obtained by applying a τ-*rotation* to a vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ is given by the product:

$$\mathbf{R}_\tau\vec{\mathbf{v}} = \begin{bmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\tau - y\sin\tau \\ x\sin\tau + y\cos\tau \end{bmatrix}.$$

In particular, the unit vector $\vec{\mathbf{u}}_\tau = \begin{bmatrix} \cos\tau \\ \sin\tau \end{bmatrix}$, which represents the *first column of* the rotation matrix $\mathbf{R}_\tau$, and is given by $\vec{\mathbf{u}}_\tau = \mathbf{R}_\tau\vec{\mathbf{u}}_0$, is the result of applying a τ-rotation to the vector $\vec{\mathbf{u}}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Since *applying a rotation* is the same as *multiplying by a rotation matrix*, we clearly see that rotation matrices obey the following important rule.

Product Rule for Rotation Matrices

The product of two rotation matrices is again a rotation matrix, namely

$$\mathbf{R}_{\alpha+\beta} = \mathbf{R}_\alpha \cdot \mathbf{R}_\beta. \tag{4.1.8}$$

When we write down the Product Rule (4.1.8) explicitly (by identifying the coefficients of all three matrices involved in it), we get:

$$\begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \tag{4.1.9}$$

When we take a closer look at the matrix product $\mathbf{R}_\alpha \cdot \mathbf{R}_\beta$ that appears in the right-hand side of (4.1.9), especially to the entries that sit in the first column of $\mathbf{R}_\alpha \cdot \mathbf{R}_\beta$, we get exactly formulas (4.1.5) and (4.1.6).

Application: Expressions of the Form $E(x) = a\cos kx + b\sin kx$

Using the Angle Difference Formula for Cosine, there is a neat way to simplify expressions of the form:

$$E(x) = a\cos kx + b\sin kx.$$

The main idea is to make such expressions match the cosine of a difference. Of course, this is only possible, if we divide by $\sqrt{a^2+b^2}$, so a better expression we should first consider is:

$$\frac{a}{\sqrt{a^2+b^2}}\cos kx + \frac{b}{\sqrt{a^2+b^2}}\sin kx.$$

We know, from Section 3.3, that we can always find an angle τ, such that

$$\begin{cases} \cos\tau &= \dfrac{a}{\sqrt{a^2+b^2}} \\ \sin\tau &= \dfrac{b}{\sqrt{a^2+b^2}} \end{cases}$$

for instance we can take τ to be *standard direction angle of the vector* $\vec{\mathbf{v}} = \begin{bmatrix} a \\ b \end{bmatrix}$, which is given by the formula:

$$\tau = (\text{sign of } b)\arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right). \tag{4.1.10}$$

(As in Section 3.1, we agree that when $b = 0$, the above sign is $+$.) With this choice of τ, we can write

$$\frac{a}{\sqrt{a^2+b^2}}\cos kx + \frac{b}{\sqrt{a^2+b^2}}\sin kx = \cos kx\cos\tau + \sin kx\cos\tau = \cos(kx-\tau),$$

so the original expression we wanted to simplify can now be written as follows.

> With τ is given by (4.1.10), the expression $E(x) = a\cos kx + b\sin kx$ can also be presented as:
>
> $$E(x) = a\cos kx + b\sin kx = \sqrt{a^2+b^2}\cos(kx-\tau). \tag{4.1.11}$$

Example 4.1.3. Suppose we are given the function $f(x) = 3\cos 5x - 2\sin 5x$, and we are asked to find its maximum value, as well as the x-values where $f(x)$ attains its maximum.

The idea is to re-write $f(x)$ as in (4.1.11). We start off by identifying the coefficients: $a = 3$, $b = -2$. For future use, we also compute at this time the value:

$$\sqrt{a^2+b^2} = \sqrt{3^2+(-2)^2} = \sqrt{13}.$$

Since b is negative, the angle τ given by (4.1.10) is:

$$\tau = -\arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right) = -\arccos\left(\frac{3}{\sqrt{13}}\right).$$

We defer the calculation of τ until later. However, for the record, we point point that, by construction (remember that τ is a standard direction angle of the vector $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$) we already know that

$$-\pi \le \tau \le \pi. \tag{4.1.12}$$

With τ more-or-less in hand, using (4.1.11) we can now write our given function as

$$f(x) = \sqrt{13}\cos(5x - \tau). \tag{4.1.13}$$

With this alternative presentation of $f(x)$, it is now clear that its maximum value is $\sqrt{13}$. Furthermore, the x-values where this maximum is attained are the solutions of the equation

$$\cos(5x - \tau) = 1.$$

Using the substitution $5x - \tau = y$, this equation becomes $\cos y = 1$, and all its solutions are[14] $y = 2n\pi$, n integer, so when we go back to the substitution, we get $5x - \tau = 2n\pi$, which yields $5x = \tau + 2n\pi$, thus the x-values where $f(x)$ attains its maximum value ($\sqrt{13}$) are:

$$x = \frac{\tau}{5} + \frac{2n\pi}{5},\ n \text{ integer}. \tag{4.1.14}$$

It should be pointed out here that, when we plug in $n = 0$, we get the special value

$$x = \frac{\tau}{5} \simeq -.117600521,$$

which is *the closest x-value to zero, at which $f(x)$ attains its maximum value ($\sqrt{13}$).* This is due to the fact that, by dividing all sides in (4.1.12) by 5, it follows that $\frac{\tau}{5}$ is in the interval $\left[-\frac{\pi}{5}, \frac{\pi}{5}\right]$, so adding any non-zero multiple of $\frac{2\pi}{5}$ will result in a number outside this interval.

Using presentations like (4.1.11), we can also develop a method for solving equations of the form

$$a\cos kx + b\sin kx = c. \tag{4.1.15}$$

Assume both a and b are non-zero. If we set τ to be given by (4.1.10), then the solutions of the equation (4.1.15) are as follows.

I. If $c > \sqrt{a^2 + b^2}$, or $c < -\sqrt{a^2 + b^2}$, then there are no solutions.

[14] There is no need to use arccos for this equation! We know that the solutions are the “peaks” of cosine!

II. If $c = \sqrt{a^2+b^2}$, then the solutions are:

$$x = \frac{1}{k}\left[\tau + 2n\pi\right],\ n \text{ integer.}$$

III. If $c = -\sqrt{a^2+b^2}$, then the solutions are:

$$x = \frac{1}{k}\left[\tau + \pi + 2n\pi\right],\ n \text{ integer.}$$

III. If $-\sqrt{a^2+b^2} < c < \sqrt{a^2+b^2}$, then the solutions are:

$$x = \frac{1}{k}\left[\tau \pm \arccos\left(\frac{c}{\sqrt{a^2+b^2}}\right) + 2n\pi\right],\ n \text{ integer.}$$

CLARIFICATION. With τ given by (4.1.10), the equation (4.1.15) is equivalent to:

$$\cos(kx - \tau) = \frac{c}{\sqrt{a^2+b^2}},$$

which after the substitution $kx - \tau = u$ becomes

$$\cos u = \frac{c}{\sqrt{a^2+b^2}}.$$

Depending of the four possibilities listed, we immediately get:

I. If $\frac{c}{\sqrt{a^2+b^2}} > 1$, or $\frac{c}{\sqrt{a^2+b^2}} < -1$, then there are no solutions.

II. If $\frac{c}{\sqrt{a^2+b^2}} = 1$, then:

$$u = 2n\pi,\ n \text{ integer.}$$

III. If $\frac{c}{\sqrt{a^2+b^2}} = -1$, then:

$$u = \pi + 2n\pi,\ n \text{ integer.}$$

III. If $-1 < \frac{c}{\sqrt{a^2+b^2}} < 1$, then:

$$u = \pm\arccos\left(\frac{c}{\sqrt{a^2+b^2}}\right) + 2n\pi,\ n \text{ integer.}$$

Going back to the substitution, we clearly have $kx = \tau + u$, and then all the given formulas are clear, by dividing everything by k.

Example 4.1.4. Let us solve the equation $\cos 3t + \sin 3t = 1$.

Our angle τ can be easily guessed here. Since $a = b = 1$, we clearly have $\tau = \frac{\pi}{4}$, so we can write

$$\cos t + \sin t = \sqrt{2}\left[\frac{1}{\sqrt{2}}\cos 3t + \frac{1}{\sqrt{2}}\sin 3t\right] = \sqrt{2}\left[\cos 3t \cos\frac{\pi}{4} + \sin 3t \sin\frac{\pi}{4}\right] = \sqrt{2}\cos\left(3t - \frac{\pi}{4}\right).$$

Our equation becomes $\sqrt{2}\cos\left(3t - \frac{\pi}{4}\right) = 1$, which can be rewritten as

$$\cos\left(3t - \frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}.$$

Using the substitution $3t - \frac{\pi}{4} = u$, this equation reduces to

$$\cos u = \frac{1}{\sqrt{2}},$$

which yields: $u = \pm\arccos\left(\frac{1}{\sqrt{2}}\right) + 2n\pi = \pm\frac{\pi}{4} + 2n\pi$, n integer. Going back to our substitution, we now get

$$3t - \frac{\pi}{4} = \pm\frac{\pi}{4} + 2n\pi,$$

so when we add $\frac{\pi}{4}$ to both sides we get

$$3t = \frac{\pi}{4} \pm \frac{\pi}{4} + 2n\pi,$$

which after dividing everything by 3 yields

$$t = \frac{1}{3}\left(\frac{\pi}{4} \pm \frac{\pi}{4} + 2n\pi\right), \ n \text{ integer}.$$

If we wish, we can break these solutions into two lists:

(a) $t = \frac{1}{3}\left(\frac{\pi}{4} + \frac{\pi}{4} + 2n\pi\right) = \frac{\pi}{6} + \frac{2n\pi}{3}$, n integer;

(b) $t = \frac{1}{3}\left(\frac{\pi}{4} - \frac{\pi}{4} + 2n\pi\right) = \frac{2n\pi}{3}$, n integer.

Example 4.1.5. Suppose we are asked to find all solutions of the equation

$$2\cos 2x + \sin 2x = 1,$$

that are in the interval $[-\pi, \pi]$.

We will transform the left-hand side of our equation in a manner similar to what we did in Example 4.1.3. We start off by identifying the coefficients: $a = 2$, $b = 1$. For future use, we also compute at this time the value:

$$\sqrt{a^2 + b^2} = \sqrt{2^2 + 1^2} = \sqrt{5}.$$

Since b is positive, the angle τ given by (4.1.10) is:

$$\tau = \arccos\left(\frac{a}{\sqrt{a^2 + b^2}}\right) = \arccos\left(\frac{2}{\sqrt{5}}\right).$$

We defer the calculation of τ until later. With τ more-or-less in hand, using (4.1.11) we can now write $2\cos 2x + \sin 2x = \sqrt{5}\cos(2x - \tau)$, so now our equation becomes: $\sqrt{5}\cos(2x - \tau) = 1$, which we can also write as:

$$\cos(2x - \tau) = \frac{1}{\sqrt{5}}.$$

Using the substitution $2x - \tau = w$, our equation reads:

$$\cos w = \frac{1}{\sqrt{5}},$$

which has as its solutions:

$$w = \pm \arccos\left(\frac{1}{\sqrt{5}}\right) + 2n\pi, \ n \text{ integer}.$$

As we did with τ, we will defer the calculation of $\arccos\left(\frac{1}{\sqrt{5}}\right)$. Let us just give it a name for now: call it ϕ. If we go back to our substitution, we get $2x - \tau = \pm\phi + 2n\pi$, which yields $2x = \tau \pm \phi + 2n\pi$, and then we get:

$$x = \frac{1}{2}\left[\tau \pm \phi + 2n\pi\right] = \frac{\tau \pm \phi}{2} + n\pi, \ n \text{ integer}.$$

We can now start our computations, and split the solutions into the following two lists (depending on the $\pm$ sign):

(a) $x = \dfrac{\tau + \phi}{2} + n\pi \simeq 0.785398163 + n\pi$, n integer. From this list, it is clear that only the values given by $n = 0$ and $n = -1$ will yield numbers in $[-\pi, \pi]$:

$$x_1 \simeq 0.785398163;$$
$$x_2 \simeq 0.785398163 - \pi \simeq -2.35619449.$$

(b) $x = \dfrac{\tau - \phi}{2} + n\pi \simeq -0.321750054 + n\pi$, n integer. From this list, it is clear that only the values given by $n = 0$ and $n = 1$ will yield numbers in $[-\pi, \pi]$:

$$x_3 \simeq -0.321750054;$$
$$x_4 \simeq -0.321750054 + \pi \simeq 2.819842099.$$

The Tangent Formulas

If we pay special attention to the Angle Sum Formulas, we see that we can write

$$\tan(\alpha+\beta) = \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} = \frac{\sin\alpha\cos\beta + \cos\alpha\sin\beta}{\cos\alpha\cos\beta - \sin\alpha\sin\beta}. \tag{4.1.16}$$

We can push this a little bit by force-factoring $\cos\alpha\cos\beta$ both upstairs and downstairs, so we can write

$$\text{Numerator} = \cos\alpha\cos\beta\left[\frac{\sin\alpha\cos\beta}{\cos\alpha\cos\beta} + \frac{\cos\alpha\sin\beta}{\cos\alpha\cos\beta}\right] = \cos\alpha\cos\beta\left[\tan\alpha + \tan\beta\right];$$
$$\text{Denominator} = \cos\alpha\cos\beta\left[\frac{\cos\alpha\cos\beta}{\cos\alpha\cos\beta} - \frac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}\right] = \cos\alpha\cos\beta\left[1 - \tan\alpha\tan\beta\right].$$

This way, going back to (4.1.16), we arrive to the following two interesting formulas (the one for differences is obtained the same way):

Angle Sum/Difference Formulas for Tangent

$$\tan(\alpha+\beta) = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta};$$
$$\tan(\alpha-\beta) = \frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta}.$$

What is nice about these two formulas is the fact that, if we only want to know tangent of a sum or a difference of two angles, we do not need all four values of sines and cosines.

Example 4.1.6. Suppose we want to compute $\tan[\arctan(10) + \arctan(7)]$. If we set up $\arctan(10) = \alpha$ and $\arctan(7) = \beta$, then $\tan\alpha = 10$ and $\tan\beta = 7$, so we can immediately compute

$$\tan[\arctan(10) + \arctan(7)] = \tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta} = \frac{10+7}{1 - 10\cdot 7} = -\frac{17}{69}.$$

In this particular example, we can also locate the quadrant of $\alpha+\beta$, as follows. We know that, since both α and β are in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and their tangents are positive, these two angles belong in fact to the interval $\left(0, \frac{\pi}{2}\right)$, thus $0 < \alpha, \beta < \frac{\pi}{2}$, so if we add them, we get $0 < \alpha+\beta < \pi$, so $\alpha+\beta$ can only be in Quadrant I or II. Since $\tan(\alpha+\beta)$ is negative, it follows that $\alpha+\beta$ is in Quandrant II.

Exercises

In Exercises 1-8 you are asked to prove several well known identities, using the addition and subtraction formulas.

1. $\sin\left(t + \frac{\pi}{2}\right) = \cos t$.

2. $\sin\left(t - \frac{\pi}{2}\right) = -\cos t$.

3. $\cos\left(t + \frac{\pi}{2}\right) = -\sin t$.

4. $\cos\left(t - \frac{\pi}{2}\right) = \sin t$.

5. $\sin(t+\pi) = -\sin t$.

6. $\sin(t-\pi) = -\sin t$.

7. $\cos(t+\pi) = -\cos t$.

8. $\cos(t-\pi) = -\cos t$.

In Exercises 9-13 you are asked to find the **exact** values of sine, cosine, and tangent, for several angles, denoted by α, that can be easily computed out of the "familiar" angles $30°$, $45°$, $60°$ and $90°$. For instance, the exact value of $\cos 75°$ is: $\cos 30° \cos 45° - \sin 30° \sin 45° = \frac{\sqrt{3}}{2} \cdot \frac{1}{\sqrt{2}} - \frac{1}{2} \cdot \frac{1}{\sqrt{2}} = \frac{\sqrt{3}-1}{2\sqrt{2}}$.

9. $\alpha = 75°$. (Above we found cosine; find also sine and tangent.)

10. $\alpha = 15°$.

11. $\alpha = 105°$.

12. $\alpha = 165°$.

13. $\alpha = 195°$.

In Exercises 14-17 you are asked to find the **exact** values of sine, cosine, and tangent, for a sum or difference of two angles, based on some information.

14. Given α in Quadrant II, with $\sin \alpha = \frac{3}{5}$, and β in Quadrant III, with $\tan \beta = \frac{12}{5}$, find sine, cosine and tangent of the angle $\gamma = \alpha + \beta$, as well as the quadrant of γ.

15. Same problem as above, except that $\gamma = \alpha - \beta$.

16. Given α, β in Quadrant II, with $\tan \alpha = -\frac{7}{24}$, and $\cot \beta = -\frac{8}{15}$, find sine, cosine and tangent of the angle $\gamma = \alpha + \beta$, as well as the quadrant of γ.

17. Same problem as above, except that $\gamma = \alpha - \beta$.

18. Find the exact value of $\sin \left[\arcsin\left(\frac{1}{2}\right) + \arcsin\left(\frac{\sqrt{2}}{2}\right) \right]$.

19. Find the exact value of $\sin \left[\arcsin\left(\frac{1}{3}\right) - \arccos\left(-\frac{2}{3}\right) \right]$.

20. Find the exact value of $\cos \left[\arcsin\left(-\frac{3}{5}\right) + \arctan\left(-\frac{12}{13}\right) \right]$.

21. Find the exact value of $\tan [\arctan(5) + \arctan(-2)]$.

22. Find the exact value of $\tan \left[\arctan(4) - \arctan\left(\frac{1}{4}\right) \right]$.

23. Find all solutions of the equation $\cos t + \sqrt{3} \sin t = 1$. Use **exact** value.

24. Find all solutions of the equation $\cos t - \sqrt{3} \sin t = -1$. Use **exact** value.

25. Find all solutions of the equation $\cos t + \sin t = \dfrac{1}{\sqrt{2}}$. Use **exact** value.

26. Find all solutions of the equation $-\cos t + \sin t = \dfrac{1}{\sqrt{2}}$. Use **exact** value.

4.2 Multiples of Angles

In this section we use the Angle Addition and Subtraction Formulas to obtain several formulas that are very useful for computing multiples and fractions of angles.

The Double Angle Formulas

The formulas for double angles are obtained by specializing the Angle Addition Formulas

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta+\cos\alpha\sin\beta, \tag{4.2.1}$$

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta-\sin\alpha\sin\beta, \tag{4.2.2}$$

$$\tan(\alpha+\beta) = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}, \tag{4.2.3}$$

to the case when $\beta = \alpha$. The package below summarizes all the possible versions of the resulting identities.

Double Angle Formulas

$$\sin 2\alpha = 2\sin\alpha\cos\alpha = \frac{2\tan\alpha}{1+\tan^2\alpha} \tag{4.2.4}$$

$$\cos 2\alpha = \cos^2\alpha-\sin^2\alpha = 2\cos^2\alpha-1 = 1-2\sin^2\alpha = \frac{1-\tan^2\alpha}{1+\tan^2\alpha} \tag{4.2.5}$$

$$\tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha} \tag{4.2.6}$$

CLARIFICATIONS. The first equalities in (4.2.4), (4.2.5) and (4.2.6) follow directly from the Angle Addition Formulas (4.2.1), (4.2.2) and (4.2.3) shown above.

The second and third equalities in (4.2.5) follow by replacing $\sin^2\alpha = 1-\cos^2\alpha$, or replacing $\cos^2\alpha = 1-\sin^2\alpha$.

As for the third equalities in (4.2.4) and (4.2.5), they follow from the identity $\dfrac{1}{1+\tan^2\alpha} = \dfrac{1}{\sec^2\alpha} = \cos^2\alpha$, so if we compute the final right-hand sides of both (4.2.5) and (4.2.6), we get

$$\frac{2\tan\alpha}{1+\tan^2\alpha} = 2\tan\alpha\cdot\frac{1}{1+\tan^2\alpha} = \frac{2\sin\alpha}{\cos\alpha}\cdot\cos^2\alpha = 2\sin\alpha\cos\alpha = \sin 2\alpha;$$

$$\frac{1-\tan^2\alpha}{1+\tan^2\alpha} = (1-\tan^2\alpha)\cdot\frac{1}{1+\tan^2\alpha} = \left(1-\frac{\sin^2\alpha}{\cos^2\alpha}\right)\cdot\cos^2\alpha = \cos^2\alpha-\sin^2\alpha = \cos 2\alpha.$$

ADDITIONAL CLARIFICATIONS. What we see from (4.2.5), (4.2.4) and (4.2.6) is that, in order to be able to compute the trigonometric functions of a double angle 2α, then depending on what function we need to compute, we need to complete one of the following "wish lists"

(A) Find $\tan\alpha$: this will easily give us *all* trigonometric functions of the double angle 2α.
(B) If we only want to compute $\cos 2\alpha$, then *one* of the values $\sin\alpha$ or $\cos\alpha$ suffices.
(C) If need to find $\sin 2\alpha$, then as an alternative to (A), we may also use the first equality in (4.2.4), if we know *both* $\sin\alpha$ and $\cos\alpha$.

Example 4.2.1. Suppose we know that α is an angle in Quadrant III with $\cos\alpha = -\frac{5}{13}$, and we are asked to compute sine and cosine of $2\,\alpha$, as well as the quadrant $2\,\alpha$ sits in.

The cosine of $2\,\alpha$ can be easily computed solely out of $\cos\alpha$, using (4.2.5):

$$\cos 2\alpha = 2\cos^2\alpha - 1 = 2\left(-\frac{5}{13}\right)^2 - 1 = 2\cdot\frac{25}{169} - 1 = \frac{50}{169} - \frac{169}{169} = -\frac{119}{169}.$$

As for the sine of $2\,\alpha$, we either need $\tan\alpha$, or $\sin\alpha$. Of course, knowing one of these values will give us the other one, so it is up to us to decide which one we want to compute. We opt here to compute the sine:

$$\begin{aligned}\sin\alpha &= \pm\sqrt{1-\cos^2\alpha} = \pm\sqrt{1-\left(-\frac{5}{13}\right)^2} = \pm\sqrt{1-\frac{25}{169}} = \\ &= \pm\sqrt{\frac{169}{169}-\frac{25}{169}} = \pm\sqrt{\frac{144}{169}} = \pm\frac{12}{13}.\end{aligned}$$

Since α is in Quadrant III, it follows that $\sin\alpha$ is negative, so the correct value is: $\sin\alpha = -\frac{12}{13}$, and then using the first formula from (4.2.4), we get:

$$\sin 2\alpha = 2\sin\alpha\cos\alpha = 2\left(-\frac{12}{13}\right)\left(-\frac{5}{13}\right) = \frac{120}{169}.$$

Since $\cos 2\alpha$ is *negative*, and $\sin 2\alpha$ is *positive*, it follows that 2α is in *Quadrant II*.

Example 4.2.2. Suppose we know that $\tan\alpha = -2$, and we are asked to compute sine and cosine of $2\,\alpha$, as well as the quadrant $2\,\alpha$ sits in. (Notice that we are not given any information about the Quadrant of α!)

In this case, our knowledge of $\tan\alpha$ allows us to compute both sine and cosine of $2\,\alpha$ *directly*:

$$\begin{aligned}\sin 2\alpha &= \frac{2\tan\alpha}{1+\tan^2\alpha} = \frac{2(-2)}{1+(-2)^2} = \frac{-4}{1+4} = -\frac{4}{5};\\ \cos\ 2\alpha &= \frac{1-\tan^2\alpha}{1+\tan^2\alpha}\frac{1-(-2)^2}{1+(-2)^2} = \frac{1-4}{1+4} = -\frac{3}{5}.\end{aligned}$$

Since both $\sin 2\alpha$ and $\cos 2\alpha$ are *negative*, it follows that 2α is in *Quadrant III*.

The Triple Angle Formulas

The Double Angle Formulas can be combined with the Angle Addition Formulas to yield the following interesting formulas for triple angles.

Triple Angle Formulas

$$\sin 3\alpha = 3\sin\alpha - 4\sin^3\alpha \tag{4.2.7}$$
$$\cos 3\alpha = 4\cos^3\alpha - 3\sin\alpha \tag{4.2.8}$$
$$\tan 3\alpha = \frac{3\tan\alpha - \tan^3\alpha}{1-3\tan^2\alpha} \tag{4.2.9}$$

CLARIFICATION. Formula (4.2.7) follows from (4.2.1), (4.2.4) and (4.2.5):

$$\begin{aligned}\sin 3\alpha &= \sin(2\alpha+\alpha) = \sin 2\alpha\cos\alpha + \cos 2\alpha\sin\alpha = \\ &= 2\sin\alpha\cos^2\alpha + (1-2\sin^2\alpha)\sin\alpha = \\ &= 2\sin\alpha\,(1-\sin^2\alpha) + (1-2\sin^2\alpha)\sin\alpha = \\ &= 2\sin\alpha - 2\sin^3\alpha + \sin\alpha - 2\sin^3\alpha = \\ &= 3\sin\alpha - 4\sin^3\alpha.\end{aligned}$$

Formula (4.2.8) follows from (4.2.2), (4.2.4) and (4.2.5):

$$\begin{aligned}\cos 3\alpha &= \cos(2\alpha+\alpha) = \cos 2\alpha\cos\alpha - \sin 2\alpha\sin\alpha = \\ &= (2\cos^2\alpha - 1)\cos\alpha - 2\sin^2\alpha\cos\alpha = \\ &= 2\cos^3\alpha - \cos\alpha) - 2(1-\cos^2\alpha)\cos\alpha = \\ &= 2\cos^3\alpha - \cos\alpha) - 2\cos\alpha + 2\cos^3\alpha = \\ &= 4\cos^3\alpha - 3\cos\alpha.\end{aligned}$$

Formula (4.2.9) follows from (4.2.3) and (4.2.6):

$$\begin{aligned}\tan 3\alpha = \tan(2\alpha+\alpha) = \frac{\tan 2\alpha + \tan\alpha}{1-\tan 2\alpha\cdot\tan\alpha} &= \frac{\dfrac{2\tan\alpha}{1-\tan^2\alpha}+\tan\alpha}{1-\dfrac{2\tan\alpha}{1-\tan^2\alpha}\cdot\tan\alpha} = \\ = \frac{\dfrac{2\tan\alpha}{1-\tan^2\alpha}+\dfrac{\tan\alpha(1-\tan^2\alpha)}{1-\tan^2\alpha}}{\dfrac{1-\tan^2\alpha}{1-\tan^2\alpha}-\dfrac{2\tan^2\alpha}{1-\tan^2\alpha}} &= \frac{\dfrac{3\tan\alpha-\tan^3\alpha}{1-\tan^2\alpha}}{\dfrac{1-3\tan^2\alpha}{1-\tan^2\alpha}} = \\ = \frac{3\tan\alpha-\tan^3\alpha}{1-\tan^2\alpha}\cdot\frac{1-\tan^2\alpha}{1-3\tan^2\alpha} &= \frac{3\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha}.\end{aligned}$$

Below is a very interesting application of the Triple Angle Formulas, which illustrates how we can use them to obtain the **exact** values of a new angle.

Example 4.2.3. Consider the angle $\alpha = 36^\circ = \frac{\pi}{5}$ angle. Since $3\alpha + 2\alpha = 5\alpha = 180^\circ = \pi$, it follows that $\cos 3\alpha = \cos(\pi - 2\alpha) = -\cos 2\alpha$, that is,

$$\cos 3\alpha + \cos 2\alpha = 0, \tag{4.2.10}$$

so if we let $\cos\alpha = x$, the using (4.2.8) and (4.2.5), then the equality (4.2.10) reads:

$$4x^3 - 3x + 2x^2 - 1 = 0. \tag{4.2.11}$$

Since we also have $3\cdot\dfrac{3\pi}{5}+2\cdot\dfrac{3\pi}{5}=3\pi$, it follows that the angle $\alpha=\dfrac{3\pi}{5}(=108°)$ satisfies $\cos 3\alpha = \cos(3\pi-2\alpha)=\cos(\pi-2\alpha)=-\cos 2\alpha$, so α also satisfies (4.2.10). Clearly, $\alpha=\pi(=180°)$ also satisfies (4.2.10), so the values $x_1=\cos 36°$, $x_2=\cos 108°$ and $x_3=\cos 180°=-1$ are *all the solutions* of the cubic equation. (4.2.11). This means that the polynomial on the left-hand side of (4.2.11) factors as

$$4x^3-3x+2x^2-1=4(x-\cos 36°)(x-\cos 108°)(x+1),$$

which means that

$$(x-\cos 36°)(x-\cos 108°)=\frac{4x^3-3x+2x^2-1}{x+1}=4x^2-2x-1.$$

This means that the numbers $\cos 36°$ and $\cos 108°$ are precisely the two solutions of the quadratic equation

$$4x^2-2x-1=0.$$

The two solutions of these equations are easily obtained using the Quadratic Formula:

$$x=\frac{2\pm\sqrt{(-2)^{-}4(4)(-1)}}{2\cdot 4}=\frac{2\pm\sqrt{20}}{8}=\frac{2\pm 2\sqrt{5}}{8}=\frac{2(1\pm\sqrt{5})}{8}=\frac{1\pm\sqrt{5}}{4}.$$

Since $x_1=\cos 36°>0$ and $x_2=\cos 108°<0$, it follows that the correct way to match these values with the above solutions is: $x_1=\dfrac{1+\sqrt{5}}{4}$ and $x_2=\dfrac{1-\sqrt{5}}{4}$. Since $\sin 36°=\sqrt{1-\cos^2 36°}$, the two exact values we are interested in are:

$$\begin{cases}\cos 36° &= \dfrac{1+\sqrt{5}}{4}\\ \sin 36° &= \sqrt{1-\left(\dfrac{1+\sqrt{5}}{4}\right)^2}=\dfrac{\sqrt{10-2\sqrt{5}}}{4}\end{cases}$$

The Half Angle Formulas

We now take another look at the Double Angle Formulas, especially at the last equality in (4.2.4) and the second through fourth equalities in (4.2.5).

For instance, if we consider the second and the third equalities in (4.2.5), they give us the equalities

$$2\cos^2\alpha=1+\cos 2\alpha \text{ and } 2\sin^2\alpha=1-\cos 2\alpha,$$

which allow us to conclude that:

$$\cos^2\alpha=\frac{1+\cos 2\alpha}{2}; \tag{4.2.12}$$

$$\sin^2\alpha=\frac{1-\cos 2\alpha}{2}. \tag{4.2.13}$$

Another "game we can play" is to compute the two numerators in right-hand sides in (4.2.12) and (4.2.13) using the last equality in (4.2.5):

$$1+\cos 2\alpha = 1+\frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{1+\tan^2\alpha}{1+\tan^2\alpha}+\frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{2}{1+\tan^2\alpha}; \tag{4.2.14}$$

$$1-\cos 2\alpha = 1-\frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{1+\tan^2\alpha}{1+\tan^2\alpha}-\frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{2\tan^2\alpha}{1+\tan^2\alpha};. \tag{4.2.15}$$

When we combine these two identities with the last equality in (4.2.4), we obtain

$$\frac{\sin 2\alpha}{1+\cos 2\alpha} = \frac{\dfrac{2\tan\alpha}{1+\tan^2\alpha}}{\dfrac{2}{1+\tan^2\alpha}} = \frac{2\tan\alpha}{1+\tan^2\alpha}\cdot\frac{1+\tan^2\alpha}{2} = \tan\alpha; \tag{4.2.16}$$

$$\frac{1-\cos 2\alpha}{\sin 2\alpha} = \frac{\dfrac{2\tan^2\alpha}{1+\tan^2\alpha}}{\dfrac{2\tan\alpha}{1+\tan^2\alpha}} = \frac{2\tan^2\alpha}{1+\tan^2\alpha}\cdot\frac{1+\tan^2\alpha}{2\tan\alpha} = \tan\alpha. \tag{4.2.17}$$

If we start now with some angle θ, and we set up $\alpha=\theta/2$, then the above formulas yield:

Half Angle Formulas

$$\sin(\theta/2) = \pm\sqrt{\frac{1-\cos\theta}{2}}; \tag{4.2.18}$$

$$\cos(\theta/2) = \pm\sqrt{\frac{1+\cos\theta}{2}}; \tag{4.2.19}$$

$$\tan(\theta/2) = \pm\sqrt{\frac{1-\cos\theta}{1+\cos\theta}} = \frac{\sin\theta}{1+\cos\theta} = \frac{1-\cos\theta}{\sin\theta}. \tag{4.2.20}$$

CLARIFICATIONS. Formulas (4.2.18) and (4.2.19) follow directly from (4.2.12) and (4.2.13). By taking the ratio $\dfrac{\sin(\theta/2)}{\cos(\theta/2)}$, the first equality in (4.2.20) follows immediately. The other two equalities from (4.2.20) follow from (4.2.16) and (4.2.17).

If we need to compute the sine or cosine of a half angle, we must resolve the $\pm$ sign in each of the formulas (4.2.18) and (4.2.18). This meas that our "wish list" must include the *quadrant information on* $\theta/2$. However, when we only asked to compute the tangent of a half-angle, we have the option of using the other formulas from (4.2.20), which do not require this information.

Example 4.2.4. Suppose we hand to compute sine and cosine of $\theta/2$, given that $\pi<\theta<2\pi$ and $\cos\theta=-\dfrac{3}{5}$.

Since we are given $\cos\theta$, all we need is the quadrant information on $\theta/2$. Based on the given inequalities involving θ, which we can divide by 2, we get:

$$\frac{\pi}{2}<\frac{\theta}{2}<\pi,$$

which tells us, among other things, that *$\theta/2$ is in Quadrant II*, so using (4.2.18) and (4.2.19), with the correct signs, we get

$$\sin(\theta/2) = +\sqrt{\frac{1-\cos\theta}{2}} = \sqrt{\frac{1-\left(-\frac{3}{5}\right)}{2}} = \sqrt{\frac{1+\frac{3}{5}}{2}} = \sqrt{\frac{\frac{8}{5}}{2}} = \sqrt{\frac{4}{5}} = \frac{2}{\sqrt{5}};$$
$$\cos(\theta/2) = -\sqrt{\frac{1+\cos\theta}{2}} = -\sqrt{\frac{1+\left(-\frac{3}{5}\right)}{2}} = -\sqrt{\frac{1-\frac{3}{5}}{2}} = -\sqrt{\frac{\frac{2}{5}}{2}} = -\sqrt{\frac{1}{5}} = -\frac{1}{\sqrt{5}}.$$

Example 4.2.5. Suppose we hand to compute sine and cosine of $\theta/2$, given that $270^\circ < \theta < 360^\circ$ and $\sin\,\theta = -\frac{5}{13}$.

Our "wish list" is a little longer here, because we need

(i) Quadrant information on $\theta/2$;
(ii) $\cos\theta$; since we are only given $\sin\theta$, we also need
(iii) Quadrant information on θ !!!

Start off with (iii). Based on the given information, we immediately get that θ sits in Quadrant IV, so $\cos\theta$ is *positive*, so we have

$$\cos\theta = +\sqrt{1-\sin^2\theta} = +\sqrt{1-\left(-\frac{5}{13}\right)^2} = +\sqrt{1-\frac{25}{169}} = +\sqrt{\frac{144}{169}} = +\frac{12}{13}.$$

Based on the given inequalities for θ, we get (upon dividing by 2) the inequalities $\frac{270^\circ}{2} < \frac{\theta}{2} < \frac{360^\circ}{2}$, that is,

$$135^\circ < \frac{\theta}{2} < 180^\circ,$$

which clearly tells us that $\theta/2$ is in Quadrant II, so using (4.2.18) and (4.2.19), with the correct signs, we get

$$\sin(\theta/2) = +\sqrt{\frac{1-\cos\theta}{2}} = \sqrt{\frac{1-\frac{12}{13}}{2}} = \sqrt{\frac{\frac{1}{13}}{2}} = \sqrt{\frac{1}{26}} = \frac{1}{\sqrt{26}};$$
$$\cos(\theta/2) = -\sqrt{\frac{1+\cos\theta}{2}} = -\sqrt{\frac{1+\frac{12}{13}}{2}} = -\sqrt{\frac{\frac{25}{13}}{2}} = -\sqrt{\frac{25}{26}} = -\frac{5}{\sqrt{26}}.$$

Example 4.2.6. Using Half Angle Formulas, we can compute the **exact** value of halves of the familiar angles. For instance, we can consider the angle $\frac{\pi}{8} = 22.5^\circ$, which is in Quadrant I, so its sine and cosine will be:

$$\sin\frac{\pi}{8} = +\sqrt{\frac{1-\cos\frac{\pi}{4}}{2}} = \sqrt{\frac{1-\frac{1}{\sqrt{2}}}{2}};$$

$$\cos\frac{\pi}{8} = +\sqrt{\frac{1+\cos\frac{\pi}{4}}{2}} = \sqrt{\frac{1+\frac{1}{\sqrt{2}}}{2}}.$$

Even though the above values do not look "nice" at all, they are nevertheless **exact**.

Trigonometric Equations Involving Multiple Angles

Using the Double (or Triple) Angle Formulas it is possible to solve more complicated equations, as seen in the following Example.

Example 4.2.7. Suppose we want to sole the equation

$$\sin 2t - \sin t = 0.$$

Using the Double Angle, we can replace $\sin 2t = 2\sin t\cos t$, so the above equation becomes

$$2\sin t\cos t - \sin t = 0.$$

Now we can factor the left-hand side, so our equation becomes

$$\sin t(2\cos t - 1) = 0,$$

and then we can split into two equations:

(A) $\sin t = 0$, which has as solutions: $t = n\pi$, n integer;

(B) $2\cos t - 1 = 0$, which is equivalent to $\cos t = \dfrac{1}{2}$, so it has has as solutions (see Section 2.7):

$$t = \pm\arccos\left(\frac{1}{2}\right) + 2n\pi = \pm\frac{\pi}{3} + 2n\pi,\ n \text{ integer}.$$

Exercises

In Exercises 1-7 you are asked to find the sine, cosine and tangent of the double angle 2θ, as well as the quadrant 2θ sits in, based on the given information on θ. Use **exact** values.

1. $\cos\theta = \dfrac{8}{17}$, θ in Quadrant IV.

2. $\sin\theta = -\dfrac{7}{25}$, θ in Quadrant III.

3. $\tan\theta = -7$.

4. $\theta = \arccos\left(-\dfrac{12}{13}\right)$

5. $\theta = \arcsin\left(-\dfrac{35}{37}\right)$

6. $\cot\theta = -3$.

7. $\theta = \arctan(-4)$.

In Exercises 8-12 you are asked to find the sine, cosine and tangent of the half angle $\theta/2$, based on the given information on θ. Use **exact** values.

8. $\sin\theta = -\frac{3}{5}$, $\frac{3\pi}{2} < \theta < \frac{5\pi}{2}$.

9. $\cos\theta = -\frac{1}{4}$, $\pi < \theta < 2\pi$.

10. $\tan\theta = -\frac{7}{24}$, $-3\pi < \theta < -2\pi$.

11. $\cos\theta = -\frac{1}{2}$, $180° < \theta < 360°$.

12. $\theta = \arccos\left(-\frac{5}{13}\right)$

In Exercises 13-19 you are asked to find all solutions of the given trigonometric equation

13. $\sin 2x = \sqrt{3}\sin x$. Use **exact values**.

14. $\sin x + \cos 2x = 1$. Use **exact values**.

15. $\sin 3x + \sin x = 0$. Use **exact values**.

16. $\cos 2x - \tan x = 1$. Use **exact values**.

17. $\sin 2t - \tan x = 0$. Use **exact values**.

18. $\tan 2x = \tan x$ Use **exact values**.

19*. $\cos 2x + \cos x = \frac{1}{\sqrt{2}}$. Use **exact values**.

In Exercises 20-26 you are asked to verify the given identity.

20. $\sin 4t = 4\sin t\cos t\cos 2t$.

21. $(\sin u + \cos u)^2 = 1 + \sin 2u$.

22. $\cos^4 x - \sin^4 x = \cos 2x$.

23. $\frac{\sec^2 t}{2 - \sec^2 t} = \sec 2t$.

24. $\cos 4t = 8\cos^4 t - 8\cos^2 t + 1$.

25. $2\csc 2u = \csc u\sec u$.

26. $\tan(t/2) = \csc t - \cot t$.

4.3 From Sums to Products and Back

In this section we build on the Angle Addition and Subtraction Formulas to several useful identities that involve sums and products of sines and cosines.

The Product-to-Sum Identities

Let us start off with the Angle Addition and Subtraction Formula for cosines, and put them side by side:

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta+\sin\alpha\sin\beta, \tag{4.3.1}$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta-\sin\alpha\sin\beta. \tag{4.3.2}$$

When we add these two identities, the product $\sin\alpha\sin\beta$ cancels out, so we simply get:

$$\cos(\alpha-\beta) + \cos(\alpha+\beta) = 2\cos\alpha\cos\beta. \tag{4.3.3}$$

Likewise, when we subtract (4.3.2) from (4.3.1), the the product $\cos\alpha\cos\beta$ cancels out, so now we get:

$$\cos(\alpha-\beta) - \cos(\alpha+\beta) = 2\sin\alpha\sin\beta. \tag{4.3.4}$$

When we put the Angle Addition and Subtraction Formulas for sines side by side:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta+\cos\alpha\sin\beta, \tag{4.3.5}$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta-\cos\alpha\sin\beta, \tag{4.3.6}$$

and we add them, the product $\cos\alpha\sin\beta$ cancels out, so now we get:

$$\sin(\alpha+\beta) + \sin(\alpha-\beta) = 2\sin\alpha\cos\beta. \tag{4.3.7}$$

So now, when we read each one of the formulas (4.3.3), (4.3.4), (4.3.7) from right to left, then after dividing each one of them by 2, we obtain the following formula package.

Product-to-Sum Identities

$$\cos\alpha\cos\beta = \tfrac{1}{2}\big[\cos(\alpha-\beta) + \cos(\alpha+\beta)\big]; \tag{4.3.8}$$
$$\sin\alpha\sin\beta = \tfrac{1}{2}\big[\cos(\alpha-\beta) - \cos(\alpha+\beta)\big]; \tag{4.3.9}$$
$$\sin\alpha\cos\beta = \tfrac{1}{2}\big[\sin(\alpha+\beta) + \sin(\alpha-\beta)\big]. \tag{4.3.10}$$

CLARIFICATIONS. At a first glance, the formulas (4.3.8) and (4.3.9) may appear to be unbalanced. What happens if we flip α and β? The left-hand sides of (4.3.8) and (4.3.9) will be unchanged, but what happens in the right-hand side? As it turns out, since $\cos$ is *even*, we can always replace

$$\cos(\beta-\alpha) = \cos(\alpha-\beta),$$

so the right-hand sides of (4.3.8) and (4.3.9) will not change either.

As for the formula (4.3.10), when we flip α and β, we should take into account the fact that $\sin$ is *odd*, so the correct replacement is

$$\sin(\beta-\alpha) = -\sin(\alpha-\beta),$$

so in addition to (4.3.10), we also have another possible expansion:

$$\cos\alpha\sin\beta = \tfrac{1}{2}\big[\sin(\alpha+\beta) - \sin(\alpha-\beta)\big]. \tag{4.3.11}$$

Deciding, whether to use (4.3.10) or (4.3.11) to expand of product of one sine and one cosine, is up to you.

Application: Fourier Expansion

The main application of the Product-to-Sum Formulas is concerned with certain types of trigonometric sums described as follows.

For any integer $N \geq 0$, a **Fourier function of order N** is a sum of the form:

$$\begin{aligned} \mathsf{F}(x) = a_0 + a_1\cos x + a_2\cos 2x + a_3\cos 3x + \ \dots\ + a_N\cos Nx +\\ + b_1\sin x + b_2\sin 2x + b_3\sin 2x + \ \dots\ + b_N\sin Nx, \end{aligned} \tag{4.3.12}$$

where $a_0, a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N$ are constants, with either a_N or b_N non-zero.

What is not at all obvious is the fact that, whenever a function $\mathsf{F}(x)$ can be written as a sum as above, then the integer N, as well as the coefficients $a_0, a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N$, are *uniquely determined*. With this uniqueness in mind (which will be clarified in Section 5.2), the sum that appears in the right-hand side of (4.3.12) is what we call the **Fourier expansion of F**.

Right now, the only thing we are concerned with is *finding* Fourier expansions for various functions.

Example 4.3.1. Suppose we are asked to find the Fourier expansion of

$$f(t) = (1 + \sin 3t)\sin^2 t.$$

We start off by reducing $\sin^2 t$ using formula (4.3.9):

$$\sin^2 t = \frac{1}{2}\big[\cos(t-t) - \cos(t+t)\big] = \frac{1}{2}\big[\cos\ 0 - \cos 2t\big] = \frac{1}{2} - \frac{1}{2}\cos 2t.$$

(Here we also used the obvious identity $\cos 0 = 1$.) Now we can replace the factor $\sin^2 t$ back in $f(t)$, and obtain (by "folding"):

$$f(t) = (1 + \sin 3t)\left[\frac{1}{2} - \frac{1}{2}\cos 2t\right] = \frac{1}{2} - \frac{1}{2}\cos 2t + \frac{1}{2}\sin 3t - \frac{1}{2}\sin 3t\cos 2t, \tag{4.3.13}$$

and all we have to take care of is the last term. Using (4.3.11), and leaving $-\dfrac{1}{2}$ aside for a moment, we can write:

$$\sin 3t\cos 2t = \frac{1}{2}\big[\sin(2t+3t) + \sin(3t-2t)\big] = \frac{1}{2}\big[\sin 5t + \sin t\big] = \frac{1}{2}\sin 5t + \frac{1}{2}\sin t,$$

so when we go back to (4.3.13), we get:

$$f(t) = \frac{1}{2} - \frac{1}{2}\cos 2t + \frac{1}{2}\sin 3t - \frac{1}{2}\left[\frac{1}{2}\sin 5t + \frac{1}{2}\sin t\right] = \frac{1}{2} - \frac{1}{2}\cos 2t + \frac{1}{2}\sin 3t - \frac{1}{4}\sin 5t - \frac{1}{4}\sin t.$$

Of course, when we want to arrange the Fourier expansion more neatly, we can (and should) write

$$f(t) = \frac{1}{2} - \frac{1}{2}\cos 2t - \frac{1}{4}\sin t + \frac{1}{2}\sin 3t - \frac{1}{4}\sin 5t, \tag{4.3.14}$$

so our function $f(t)$ is a *Fourier function of order* 5.

CLARIFICATION: (READ THIS ONLY IF YOU ARE AN ENGINEERING STUDENT!). Why does one care about Fourier expansions? Consider for instance the function f from Example 4.3.1, and think of it as representing the wave function of some sound produced by several "instruments." (Think of these "instruments" for instance, as musical instruments in a band. You can also think of them as keys on a piano keyboard.) Assume the pitch of "instrument k" is the function $p_k(t) = \sin kt$, where k is some positive integer. Besides these "instruments" we also have to account for "constant sounds," which are constant functions. When we look closely at the Fourier expansion (4.3.14), we see that our sound is put together using a "constant sound" $c = \frac{1}{2}$, and "instruments" that use $k = 1, 2, 3, 5$. To see the *volume* of each "instrument" that contributes to (4.3.14), we need to rearrange all terms with negative coefficients, as well as those that have cosine functions. We do this using *shifting formulas* like $-\sin\alpha = \sin(\alpha - \pi)$ and $\cos\alpha = \sin\left(\alpha + \frac{\pi}{2}\right) = -\sin\left(\alpha - \frac{\pi}{2}\right)$. Using such identities, the Fourier expansion (4.3.14) can be rearranged as:

$$\begin{aligned}
f(t) &= \frac{1}{2} - \frac{1}{2}\cos 2t - \frac{1}{4}\sin t + \frac{1}{2}\sin 3t - \frac{1}{4}\sin 5t = \\
&= \frac{1}{2} + \frac{1}{2}\sin\left(2t - \frac{\pi}{2}\right) + \frac{1}{4}\sin(t-\pi) + \frac{1}{2}\sin 3t + \frac{1}{4}\sin(5t - \pi) = \\
&= \frac{1}{2} + \frac{1}{2}\sin\left(2\left(t - \frac{\pi}{4}\right)\right) + \frac{1}{4}\sin(t-\pi) + \frac{1}{2}\sin 3t + \frac{1}{4}\sin\left(5\left(t - \frac{\pi}{5}\right)\right) = \\
&= \frac{1}{2} + \frac{1}{2}p_2\left(t - \frac{\pi}{4}\right) + \frac{1}{4}p_1(t-\pi) + \frac{1}{2}p_3(t) + \frac{1}{4}p_5\left(t - \frac{\pi}{5}\right).
\end{aligned}$$

Based on this calculation, we now see that the "sound" $f(t)$ is assembled using:

- the "constant sound" at volume $c = \frac{1}{2}$, together with
- "instrument 1" at volume $\frac{1}{4}$, shifted by π;
- "instrument 2" at volume $\frac{1}{2}$, shifted by $\frac{\pi}{4}$;
- "instrument 3" at volume $\frac{1}{2}$;
- "instrument 5" at volume $\frac{1}{4}$, shifted by $\frac{\pi}{5}$.

The Sum-to-Product Identities

Suppose we "play" with the Product-to-Sum Formulas (4.3.8), (4.3.9), (4.3.10) and (4.3.11) by multiplying all identities by 2. When reading what we get from right to left, we get the following identities

$$\cos(\alpha-\beta) + \cos(\alpha+\beta) = 2\cos\alpha\cos\beta; \tag{4.3.15}$$

$$\cos(\alpha-\beta) - \cos(\alpha+\beta) = 2\sin\alpha\sin\beta; \tag{4.3.16}$$

$$\sin(\alpha+\beta) + \sin(\alpha-\beta) = 2\sin\alpha\cos\beta; \tag{4.3.17}$$

$$\sin(\alpha+\beta) - \sin(\alpha-\beta) = 2\sin\beta\cos\alpha. \tag{4.3.18}$$

Suppose now we use the substitution

$$\begin{cases} \alpha+\beta &= u \\ \alpha-\beta &= v \end{cases}$$

which can be solved as a system of equations in α and β to give:

$$\begin{cases} \alpha &= \dfrac{u+w}{2} \\ \\ \beta &= \dfrac{u-v}{2} \end{cases}$$

Using these computations, when we go back to the identities (4.3.15), (4.3.16), (4.3.17) and (4.3.18), we now get the following formula package.

Sum-to-Product Identities

$$\cos u+\cos v = 2\cos\left(\frac{u+v}{2}\right)\ \cos\left(\frac{u-v}{2}\right); \qquad (4.3.19)$$

$$\cos u-\cos v = 2\sin\left(\frac{u+v}{2}\right)\ \sin\left(\frac{v-u}{2}\right); \qquad (4.3.20)$$

$$\sin u+\sin v = 2\sin\left(\frac{u+v}{2}\right)\ \cos\left(\frac{u-v}{2}\right); \qquad (4.3.21)$$

$$\sin u-\sin v = 2\sin\left(\frac{u-v}{2}\right)\ \cos\left(\frac{u+v}{2}\right). \qquad (4.3.22)$$

Application: Equations

CLARIFICATION. The Sum-to-Product Identities are very useful tools for *factoring sums or differences of sines or cosines*. This is particularly useful when we want to solve certain equations, as seen in the Examples below, which shows how certain equations can be reduced (after substitutions) to *x-intercept equations*, which are those of the form $\sin ? = 0$, or $\cos ? = 0$.

When solving x-intercept equations of the form $\sin ? = 0$, or $\cos ? = 0$ we *do not need to use inverse trigonometric functions* as we did in Section 2.7! The solutions are *much simpler!*

I. The solutions of $\sin ? = 0$ are: $n\pi$, n integer.

II. The solutions of $\cos ? = 0$ are: $\dfrac{\pi}{2} + n\pi$, n integer.

Example 4.3.2. Suppose we want to find all solutions of the equation

$$\sin 3x = \sin x. \qquad (4.3.23)$$

Subtract the right-hand side, so our equation is the same as:

$$\sin 3x - \sin x = 0. \qquad (4.3.24)$$

Using (4.3.22) we can factor

$$\sin 3x - \sin x = 2\ \sin\left(\frac{3x-x}{2}\right)\ \cos\left(\frac{3x+x}{2}\right) = 2\ \sin x\ \cos 2x,$$

so our equation becomes

$$2\ \sin x\ \cos 2x = 0.$$

By setting each factor equal to zero (of course, we cannot have $2 = 0$), the above equation breaks into two possibilities:

(A) $\sin x = 0$, which yields $x = n\pi$, n integer.

(B) $\cos 2x = 0$, which yields $2x = \dfrac{\pi}{2} + n\pi$, thus

$$x = \frac{1}{2}\left(\frac{\pi}{2} + n\pi\right) = \frac{\pi}{4} + \frac{n\pi}{2},\ n \text{ integer}.$$

Example 4.3.3. Suppose we want to find all solutions of the equation

$$\sin 3x + \cos 5x = 0. \tag{4.3.25}$$

There are many ways to solve this equation, based on transforming the left-hand side of (4.3.25) either into a sum/difference of sines, or into a sum/difference of cosines. For instance, using the well-known identity $\sin\alpha = \cos\left(\frac{\pi}{2} - \alpha\right)$, our equation can be presented as:

$$\cos\left(\frac{\pi}{2} - 3x\right) + \cos 5x = 0. \tag{4.3.26}$$

Using (4.3.19) we can factor

$$\begin{aligned}\cos\left(\frac{\pi}{2} - 3x\right) + \cos 5x &= 2\ \cos\left(\frac{\frac{\pi}{2} - 3x + 5x}{2}\right)\ \cos\left(\frac{\frac{\pi}{2} - 3x - 5x}{2}\right) = \\ &= 2\ \cos\left(\frac{\frac{\pi}{2} + 2x}{2}\right)\ \cos\left(\frac{\frac{\pi}{2} - 8x}{2}\right) = 2\ \cos\left(\frac{\pi}{4} + x\right)\ \cos\left(\frac{\pi}{4} - 4x\right),\end{aligned}$$

so our equation becomes

$$2\ \cos\left(\frac{\pi}{4} + x\right)\ \cos\left(\frac{\pi}{4} - 4x\right) = 0.$$

By setting each factor equal to zero (of course, we cannot have $2 = 0$), the above equation breaks into two possibilities:

(A) $\cos\left(\dfrac{\pi}{4} + x\right)$, which yields $\dfrac{\pi}{4} + x = \dfrac{\pi}{2} + n\pi$, thus

$$x = -\frac{\pi}{4} + \frac{\pi}{2} + n\pi = \frac{\pi}{4} + n\pi,\ n \text{ integer}.$$

(B) $\cos\left(\dfrac{\pi}{4} - 4x\right) = 0$, which yields $\dfrac{\pi}{4} - 4x = \dfrac{\pi}{2} + n\pi$, which is the same as $4x = \dfrac{\pi}{4} - \left(\dfrac{\pi}{2} + n\pi\right) = \dfrac{\pi}{4} - \dfrac{\pi}{2} - n\pi = -\dfrac{\pi}{4} - n\pi$ thus

$$x = \frac{1}{4}\left(-\frac{\pi}{4} - n\pi\right) = -\frac{\pi}{16} - \frac{n\pi}{4},\ n \text{ integer}.$$

Application: Identities

Another application of the Sum-to-Product Identities concerns certain trigonometric identities which involve ratios, as illustrated in the following Example.

Example 4.3.4. Suppose we are asked to prove the identity:

$$\frac{\sin 5x + \sin 3x}{\cos 5x + \cos 3x} = \tan 4x.$$

We start off with the left-hand side, and use the Sum-to-Product Identities to factor both the numerator and the denominator:

$$\text{LHS} = \frac{\sin 5x + \sin 3x}{\cos 5x + \cos 3x} = \frac{2\sin\left(\frac{5x+3x}{2}\right)\cos\left(\frac{5x-3x}{2}\right)}{2\cos\left(\frac{5x+3x}{2}\right)\cos\left(\frac{5x-3x}{2}\right)} = \frac{2\sin\ 4x\ \cos\ x}{2\cos\ 4x\ \cos\ x}.$$

Now we can finish everything by simplifying $2\ \cos\ x$:

$$\text{LHS} = \frac{2\sin\ 4x\ \cos\ x}{2\cos\ 4x\ \cos\ x} = \frac{\sin\ 4x}{\cos\ 4x} = \tan\ 4x = \text{RHS}.$$

Exercises

In Exercises 1-7 you are asked to find the Fourier expansion of the given function.

1. $f(t) = \sin^3 t$.

2. $f(t) = \cos^3 t$.

3. $f(t) = \sin^4 t$.

4. $f(t) = \cos^4 t$.

5. $f(t) = \sin^2 t + \cos 2t$.

6. $f(t) = (1 + \sin t)^3$.

7. $f(t) = (2\sin t - \cos t)^2$.

In Exercises 8-12 you are asked to find all solutions of the given trigonometric equation.

8. $\sin 4x - \sin 2x = 0$. Use **exact values**.

9. $\sin 4x - \cos 2x = 0$. Use **exact values**.

10. $\cos x + \cos 3x = 0$. Use **exact values**.

11. $\cos 5x - \cos 3x = 0$. Use **exact values**.

12*. $\sin x + \sin 5x = \sin 3x$. Use **exact values**. (HINT: Factor LHS, then subtract RHS, then factor again.)

In Exercises 13-18 you are asked to prove the given identity.

13. $\dfrac{\sin 4t + \sin 6t}{\cos 4t - \cos 6t} = \cot t$.

14. $\dfrac{\sin t + \sin 4t + \sin 7t}{\cos t + \cos 4t + \cos 7t} = \tan 4t$.

15. $\dfrac{\sin u + \sin v}{\sin u - \sin v} = \dfrac{\tan\left(\dfrac{u+v}{2}\right)}{\tan\left(\dfrac{u-v}{2}\right)}$.

16. $\dfrac{\cos u - \cos v}{\cos u + \cos v} = \tan\left(\dfrac{u+v}{2}\right)\tan\left(\dfrac{v-u}{2}\right)$.

17. $\sin 2x + \sin 4x + \sin 6x = 4\cos x \cos 2x \sin 3x$.

18*. $\dfrac{\sin x + \sin 2x + \sin 3x + \sin 4x + \sin 5x}{\cos x + \cos 2x + \cos 3x + \cos 4x + \cos 5x} = \tan 3x$.

Chapter 5

Applications of Trigonometry in Algebra

In this Chapter we discuss an important application of Trigonometry to Algebra, specifically dealing with calculations involving *complex numbers*.

5.1 Complex Number Arithmetic

A **complex number** is an expression of the form

$$z = a + b\,i,$$

where a and b are real numbers, and i is a special symbol, called the *purely imaginary unit*. The two coefficients a and b are referred to by the following names:

(i) The number a is called the *real part of z*, and is denoted by $\mathrm{Re}\, z$;

(ii) The number b is called the *imaginary part of z*, and is denoted by $\mathrm{Im}\, z$.

CLARIFICATIONS. Two complex numbers are *equal*, if and only if *their real parts are equal, and their imaginary parts are equal*, so:

$$a_1 + b_1 i = a_2 + b_2 i \text{ means: } \begin{cases} a_1 = a_2 \\ \text{and} \\ b_1 = b_2 \end{cases}$$

The set of all complex numbers is denoted by $\mathbb{C}$. We will always agree that the set $\mathbb{R}$ is a *subset of* $\mathbb{C}$. In other words, if we start with some real number a, we can also view it as a complex number $z = a + 0\,i$, with imaginary part equal to zero.

Addition, Subtraction and Multiplication with Complex Numbers

The easiest way to understand how addition, subtraction and multiplication of complex numbers work is to operate with them as if they are *polynomials, with i as a variable*, but we *simplify the powers of i by the rule:*

$$i^2 = -1. \tag{5.1.1}$$

Example 5.1.1. Consider the complex numbers $z_1 = 2 + 3\,i$ and $z_2 = 4 - 5\,i$.

To add these numbers we think them as first degree polynomials in i, so their sum can also the thought as a first degree polynomial in i, thus (by grouping the like terms) we get:

$$z_1 + z_2 = (2 + 3\,i) + (4 - 5\,i) = 2 + 3\,i + 4 - 5\,i = 6 + (3 - 5)i = 6 - 2\,i.$$

When we multiply our two numbers, the result can be though as a second degree polynomial in i, which (by "folding") is given as:

$$z_1 \cdot z_2 = (2 + 3\,i)(4 - 5\,i) = 8 + 12\,i - 10\,i - 15\,i^2 = 8 + 2\,i - 15\,i^2.$$

Our calculation is not done here! Using the simplification rule (5.1.1), we must continue:

$$z_1 \cdot z_2 = 8 + 2\,i - 15\,i^2 = 8 + 2\,i - 15(-1) = 8 + 2\,i + 15 = 23 + 2\,i.$$

If we wish, we can also write down the general rules for computing sums, differences and products.

Arithmetic Formulas for Complex Numbers

$$(a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_1) i \tag{5.1.2}$$

$$(a_1 + b_1 i) - (a_2 + b_2 i) = (a_1 - a_2) + (b_1 - b_1) i \tag{5.1.3}$$

$$(a_1 + b_1 i) \cdot (a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) i \tag{5.1.4}$$

Conjugate and Modulus

To each complex number one associates two quantities – another complex number and a non-negative real number, which are defined as follows.

To each complex number $z = a + b\,i$, one associates the following numbers.

- The **complex conjugate** of z is the complex number:
$$\bar{z} = a - b\,i.$$
- The **modulus** (a.k.a **absolute value**) of z is the non-negative real number:
$$|z| = \sqrt{a^2 + b^2}.$$

The associated numbers satisfy the identity:

$$z\bar{z} = a^2 + b^2 = |z|^2. \tag{5.1.5}$$

CLARIFICATION. Using the modulus, the condition that $z = 0$ is equivalent to the condition that $|z| = 0$. In particular, the condition that two complex numbers z_1 and z_2 are *equal*, is equivalent to: $|z_1 - z_2| = 0$.

The Division Operation

One of the most useful features of complex numbers is the fact that the *division operation* is also possible. Of course, it is very easy to *divide by non-zero real numbers*. Indeed, if we start with some complex number $z = a + b\,i$, and with some non-zero real number c, then:

$$\frac{z}{c} = \frac{a + b\,i}{c} = \frac{a}{c} + \frac{b}{c}\,i. \tag{5.1.6}$$

When we want to divide by an *arbitrary* (non-real) complex number, the calculation reduces to the above case, since by (5.1.5) we we simply write:

$$\frac{z_1}{z_2} = \frac{z_1\bar{z}_2}{z_2\bar{z}_2} = \frac{z_1\bar{z}_2}{|z_2|^2}, \tag{5.1.7}$$

where the denominator is now *real*. Of course, in order for the ratio $\dfrac{z_1}{z_2}$ to exist, we must assume that $z_2 \neq 0$, which is the same as $|z_2| > 0$.

If we wish, we can expand (5.1.7), which in combination with the "real" division formula (5.1.6) yields the following.

Division Formula for Complex Numbers

Assuming $a_2 + b_2\boldsymbol{i} \neq 0$,
$$\frac{a_1 + b_1\boldsymbol{i}}{a_2 + b_2\boldsymbol{i}} = \frac{a_1a_2 + b_1b_2}{a_2^2 + b_2^2} + \frac{-a_1b_2 + a_2b_1}{a_2^2 + b_2^2}\boldsymbol{i} \tag{5.1.8}$$

Example 5.1.2. Suppose we want to divide $\boldsymbol{z}_1 = 3 - 4\,\boldsymbol{i}$ by $\boldsymbol{z}_2 = 1 - 3\,\boldsymbol{i}$. According to the Division Formula, we get:
$$\frac{\boldsymbol{z}_1}{\boldsymbol{z}_2} = \frac{3 - 4\,\boldsymbol{i}}{1 + 3\,\boldsymbol{i}} = \frac{3 \cdot 1 + (-4) \cdot 3}{1^2 + 3^2} + \frac{-3 \cdot 3 + 1 \cdot (-4)}{1^2 + 3^2}\boldsymbol{i} =$$
$$= \frac{3 - 12}{10} + \frac{-9 - 4}{10}\,\boldsymbol{i} = -\frac{9}{10} - \frac{13}{10}\,\boldsymbol{i}.$$

The Fundamental Theorem of Algebra

The most remarkable feature of complex numbers, which distinguishes $\mathbb{C}$ (the set of all complex numbers) from other number sets, such as $\mathbb{R}$, or $\mathbb{Q}$, is the following statement.

The Fundamental Theorem of Algebra

Any polynomial with complex coefficients has at least one complex zero. In other words, any equation of the form
$$P(x) = 0,$$
with P a polynomial with complex coefficients, has at least one complex solution.

CLARIFICATION. The polynomials mentioned in the preceding theorem are expressions of the form
$$P(x) = A_0 + A_0x + A_2x^2 + \ldots + A_nx^n,$$
with $A_0, A_1, \ldots, A_n$ complex coefficients. The integer $n \geq 0$, which represents the largest exponent that appears with a non-zero coefficient, is called the *degree* of the polynomial, and the coefficient A_n is called the *leading coefficient*. In instances when we do not want to be very precise about our polynomial, so we only care about specifying the degree and the leading coefficient, we will present the polynomial like "$P(x) = A_nx^n +$ lower terms."

The most significant application of the Fundamental Theorem of Algebra concerns the *factorization* of polynomials, as indicated in the following statement.

The Complete Factorization Theorem

Any polynomial with complex coefficients of degree n factors completely as product of first degree polynomials:
$$P(x) = A_nx^n + \text{lower terms} = A_n(x - \boldsymbol{z}_1)(x - \boldsymbol{z}_2)\cdots(x - \boldsymbol{z}_n).$$

Example 5.1.3. Suppose we are asked to completely factor the polynomial $P(x) = 4x^2 + 4x + 5$.

By the Complete Factorization Theorem, we know that $P(x) = 4(x - \boldsymbol{z}_1)(x - \boldsymbol{z}_1)$, where the $\boldsymbol{z}$'s are the solutions of $P(x) = 0$, that is, of the equation
$$4x^2 + 4x + 5 = 0. \tag{5.1.9}$$

Since this is a second degree equation, we can try to use the Quadratic Formula:

$$x = \frac{-4 \pm \sqrt{4^2 - 4 \cdot 4 \cdot 5}}{2 \cdot 4} = \frac{-4 \pm \sqrt{-64}}{8}.$$

If we were to solve (5.1.9) *over the real numbers*, we should *stop!* However, since we allow ourselves to work over the complex numbers, we can continue our calculation by finding a suitable replacement for the quantity "$\pm\sqrt{-64}$." In this case the correct replacement will simply be "$\pm 8\,\boldsymbol{i}$" the solutions of (5.1.9) are:

$$x = \frac{-4 \pm \sqrt{4^2 - 4 \cdot 4 \cdot 5}}{2 \cdot 4} = \frac{-4 \pm 8\,\boldsymbol{i}}{8} = -\frac{4}{8} \pm \frac{8}{8}\boldsymbol{i} = -\frac{1}{2} \pm \boldsymbol{i}.$$

Using these two solutions, our polynomial factors as:

$$P(x) = 4\left(x - \left(-\frac{1}{2} + \boldsymbol{i}\right)\right)\left(x - \left(-\frac{1}{2} - \boldsymbol{i}\right)\right) = 4\left(x + \frac{1}{2} - \boldsymbol{i}\right)\left(x + \frac{1}{2} + \boldsymbol{i}\right).$$

The square root operation cannot be appropriately defined for numbers other than non-negative real numbers. For this reason, the usual Quadratic Formula for a second degree equation like

$$ax^2 + bx + c = 0,$$

should be slightly modified, in order to accommodate the cases when the discriminant $D = b^2 - 4ac$ fails to be *non-negative real number* (which is the only case when $\sqrt{D}$ is "honestly" defined).

So in the general case (which allows even for complex values of D), the correct version of the Quadratic Formula is:

$$x = \frac{-b \pm \boldsymbol{\delta}}{2a}, \tag{5.1.10}$$

where $\boldsymbol{\delta}$ is *one particular solution* of the equation

$$\boldsymbol{\delta}^2 = b^2 - 4ac.$$

It should be noted that, if $\boldsymbol{\delta}$ is one particular solution of the above equation, then the only other solution is $-\boldsymbol{\delta}$.

Example 5.1.4. Suppose we want to solve the second degree equation

$$x^2 - 2x + 4 + 4\,\boldsymbol{i} = 0. \tag{5.1.11}$$

We start off by computing the discriminant

$$D = (-2)^2 - 4 \cdot 1 \cdot (4 + 4\,\boldsymbol{i}) = 4 - 16 - 16\,\boldsymbol{i} = -12 - 16\,\boldsymbol{i}.$$

Since the discriminant is complex, in order to solve our equation, we first need to solve the equation

$$\boldsymbol{\delta}^2 = -12 - 16\,\boldsymbol{i}. \tag{5.1.12}$$

We solve this equation by setting up $\boldsymbol{\delta} = u + v\,\boldsymbol{i}$ as a complex number (with u and v real), so now we get

$$-12 - 16\,\boldsymbol{i} = (u + v\,\boldsymbol{i})^2 = u^2 - v^2 + 2uv\,\boldsymbol{i},$$

so when we match real and imaginary parts we get the system of equations

$$\begin{cases} u^2 - v^2 &= -12 \\ 2uv &= -16 \end{cases} \tag{5.1.13}$$

Using the second equation, we can substitute $v = -\dfrac{16}{2u} = -\dfrac{8}{u}$, so when we plug into the first equation, we get

$$u^2 - \frac{64}{u^2} = -12,$$

which after multiplying everything by u^2 yields $u^4 - 64 = -12u^2$, which is the same as

$$u^4 + 12u^2 - 64 = 0.$$

By subtituting $y^2 = t$, this equation reduces to

$$t^2 + 12t - 64 = 0,$$

which can be solved using the (usual) Quadratic Formula:

$$t = \frac{-12 \pm \sqrt{12^2 - 4 \cdot 1 \cdot (-64)}}{2 \cdot 1} = \frac{12 \pm \sqrt{400}}{2} = \frac{12 \pm 20}{2}.$$

The two solutions are $t_1 = \dfrac{-12+20}{2} = 4$ and $t_2 = \dfrac{-12-20}{2} = -16$. However, when we go back to the subsitution $u^2 = t$, we see that only the first one will work: $u^2 = 4$. Since we are interested only in *one* solution of (5.1.12), we will pick the easiest one: $u = 2$, and then using the substitution from the second equation of (5.1.13), we can match $u = 2$ with $v = -4$. In conclusion, *one particular* solution of (5.1.12) is:

$$\boldsymbol{\delta} = 2 - 4\,\boldsymbol{i},$$

so now, when we use the the Quadratic Formula (5.1.10), the solutions of our equation (5.1.11) will be:

$$x = \frac{-(-2) \pm (2 - 4\,\boldsymbol{i})}{2 \cdot 1} = \frac{2 \pm (2 - 4\,\boldsymbol{i})}{2} = \frac{2\left[1 \pm (1 - 2\,\boldsymbol{i})\right]}{2} = 1 \pm (1 - 2\,\boldsymbol{i}),$$

so the two solutions of (5.1.11) are $x_1 = 1 + (1 - 2\,\boldsymbol{i}) = 2 - 2\,\boldsymbol{i}$ and $x_1 = 1 - (1 - 2\,\boldsymbol{i}) = 2\,\boldsymbol{i}$.

CLARIFICATION. As we see from the above example, solving even simple equations of the form $z^2 =$ *number* becomes quite tedious, when we work over the complex numbers. This task becomes even more complicated, when we want to solve power equations of the form

$$z^n = \textit{number}.$$

In Section 5.3 we will learn how to handle such equations using Trigonometry!

Conformal Matrix Presentation of $\mathbb{C}$

THIS TOPIC IS OPTIONAL!

As it turns out, the arithmetic of complex numbers – especially the multiplication operation – is best understood if we consider the so-called *conformal matrices*, which are those of the form

$$\mathbf{Z} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \tag{5.1.14}$$

Given such a matrix, we can write it as a sum

$$\mathbf{Z} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} + \begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

so when we consider the *identity* matrix $\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the special matrix $\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then any matrix of the form (5.1.14) is now presented as

$$\mathbf{Z} = a\,\mathbf{I} + b\,\mathbf{J}.$$

The point here is that $\mathbf{J}$ satisfies a *matrix* version of (5.1.1), which looks like:

$$\mathbf{J}^2 = -\mathbf{I}, \tag{5.1.15}$$

in which the left-hand side is the *matrix product* $\mathbf{J}^2 = \mathbf{J} \cdot \mathbf{J}$.

To summarize everything, we now have a *one-to-one correspondence*

$$\boldsymbol{z} = a + b\,\boldsymbol{i} \longleftrightarrow \mathbf{Z} = a\,\mathbf{I} + b\,\mathbf{J},$$

between the set $\mathbb{C}$ of all complex numbers, and the set $\mathscr{K}$ of all matrices of the form (5.1.14), which *matches the operations of addition, subtraction and multiplication*. In particular, instead of memorixing the multiplication formula (5.1.4), we can think that the multiplication of complex numbers works as follows. Given two complex numbers $\boldsymbol{z}_1 = a_1 + b_1\boldsymbol{i}$ and $\boldsymbol{z}_2 = a_2 + b_2\boldsymbol{i}$, in order to multiply them, we do the following:

(i) assemble the associated matrices $\mathbf{Z}_1 = \begin{bmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{bmatrix}$ and $\mathbf{Z}_2 = \begin{bmatrix} a_2 & -b_2 \\ b_2 & a_2 \end{bmatrix}$;

(ii) compute the matrix product $\mathbf{Z} = \mathbf{Z}_1\mathbf{Z}_2$; the matrix $\mathbf{Z}$ will also have the form (5.1.14), that is, $\mathbf{Z} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$;

(iii) by collecting the entries in the first column of $\mathbf{Z}$, we recover the product of our given complex numbers: $\boldsymbol{z}_1\boldsymbol{z}_2 = a + b\,\boldsymbol{i}$.

For example, if we look at the two complex numbers $\boldsymbol{z}_1 = 2 + 3\,\boldsymbol{i}$ and $\boldsymbol{z}_2 = 4 - 5\,\boldsymbol{i}$. given in Example 5.1.1, one way to compute their product is to consider the matrix product:

$$\begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 & -(-5) \\ -5 & 4 \end{bmatrix} = \begin{bmatrix} 2\cdot 4 + (-3)\cdot(-5) & 2\cdot 5 + (-3)\cdot 4 \\ 3\cdot 4 + 2\cdot(-5) & 3\cdot 5 + 2\cdot 4 \end{bmatrix} = \begin{bmatrix} 23 & -2 \\ 2 & 23 \end{bmatrix},$$

so we can directly conclude that

$$\boldsymbol{z}_1 \cdot \boldsymbol{z}_2 = (2 + 3\,\boldsymbol{i})(4 - 5\,\boldsymbol{i}) = 23 + 2\,\boldsymbol{i}.$$

Exercises

1. Compute $(2 - 3\,\boldsymbol{i})(5 + 6\,\boldsymbol{i})$.

2. Compute $\boldsymbol{i}^{100}$, $\boldsymbol{i}^{101}$, $\boldsymbol{i}^{102}$, and $\boldsymbol{i}^{103}$.

3. Compute $(2 - 3\,\boldsymbol{i})^3$.

4. Compute $\dfrac{5 + 3\,\boldsymbol{i}}{2 + 3\,\boldsymbol{i}}$.

5. Compute $\dfrac{1 - 2\,\boldsymbol{i}}{1 + 2\,\boldsymbol{i}}$.

6. Compute the modulus $|2 + 3\,\boldsymbol{i}|$.

7. Compute the modulus $|(1 - 2\,\boldsymbol{i})^2|$.

8*. Show that, for any complex number $\boldsymbol{z}$, one has the equality

$$|\bar{\boldsymbol{z}}| = |\boldsymbol{z}|.$$

9*. Show that, for any two complex numbers $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, one has the equality

$$|\boldsymbol{z}_1\boldsymbol{z}_2| = |\boldsymbol{z}_1| \cdot |\boldsymbol{z}_2|.$$

10. Find all complex solutions of the equation: $z^2 = 5 - 12\,\boldsymbol{i}$.

11. Find all complex solutions of the equation: $z^2 = 4 + 4\,\boldsymbol{i}$.

5.2 Geometry of Complex Numbers

In this section we learn how to interpret complex numbers *geometrically*. At one point, Trigonometry will also enter the picture, and this will make complex numbers even more interesting.

Geometric Representations of Complex Numbers

Any complex number $\boldsymbol{z} = a + b\,\boldsymbol{i}$ can be represented geometrically as a *point $P(a, b)$ in the coordinate plane*. Using this presentation, the modulus of he complex number is simply

$$|\boldsymbol{z}| = \text{dist}(P, O),$$

the distance from P to the origin.

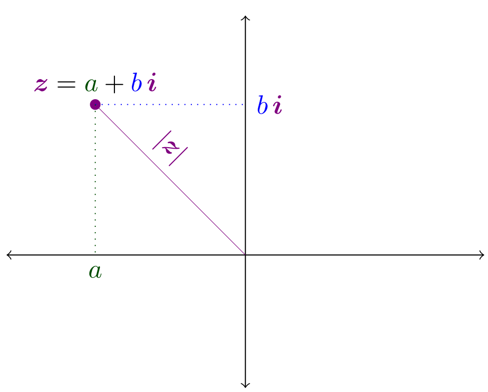


Figure 5.2.1

Another way to interpret complex numbers is to think of them as *vectors* $\vec{\mathbf{v}} = \begin{bmatrix} a \\ b \end{bmatrix}$. The main benefits of vector presentations of complex numbers are:

I. *Complex number addition* corresponds to *vector addition*.

II. For any complex number $z = a + b\,i$, the associated vector $\vec{\mathbf{v}} = \begin{bmatrix} a \\ b \end{bmatrix}$ has magnitude:

$$\|\vec{\mathbf{v}}\| = |z|.$$

Polar Representations of Complex Numbers

Whether we interpret a complex number $z = a + b\,i$ either as a *point* $P(a, b)$, or a *vector* $\vec{\mathbf{v}} = \begin{bmatrix} a \\ b \end{bmatrix}$, besides the modulus $|z|$, which can be interpreted either as a *distance*, or a *vector magnitude*, there is another object that we can consider, namely the *direction angle* τ of the vector $\vec{\mathbf{v}}$. As we learned in Section 3.3, the combination consisting of $\|\vec{\mathbf{v}}\|$ and τ is intimately related to the *polar coordinate representations of* $P(a, b)$, which are pairs (r, θ) that satisfy $a = r\,\cos\,\theta$ and $b = r\,\sin\,\theta$.

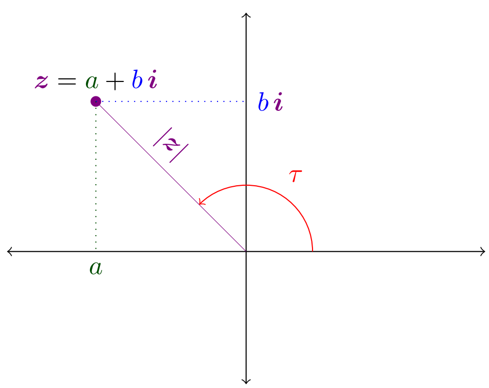


Figure 5.2.2

With this association in mind, we introduce the following terminology.

A **polar representation** of a complex number z is any presentation of it as a product

$$z = r(\cos\theta + i\sin\theta), \tag{5.2.1}$$

with $r \geq 0$.

The main features of such a representation are:

I. The number r is *uniquely determined*, namely:

$$r = |z| = \sqrt{(\text{Re}\,z)^2 + (\text{Im}\,z)^2}.$$

II. If $z \neq 0$, the number θ is *unique, up to an integer multiple of* 2π. More precisely, if one considers the angle

$$\tau = (\text{sign of Im}\,z)\arccos\left(\frac{\text{Re}\,z}{|z|}\right), \tag{5.2.2}$$

(when $\text{Im}\, z = 0$, that is, when z is *real*, we take the sign to be $+$), then θ must be of the form: $\theta = \tau + 2n\pi$, n integer.

Clarifications and Additional Terminology. Everything follows from what we learned in Section 3.3, because the condition that (r, θ) satisfies the equality (5.2.1) is equivalent to the condition that (r, θ) is a *polar coordinate representation of point* $P(a, b)$, with $r \geq 0$.

The angles θ which appear in all polar representations of z are called **arguments of z**.

Among all possible arguments, we have a particular one: the angle τ, given by (5.2.2). This special angle is referred to as the **principal argument of z**. The particular polar representation

$$z = |z|(\cos\tau + i\sin\tau), \tag{5.2.3}$$

which uses this particular angle, which is the only argument belonging to the interval $(-\pi, \pi]$, is referred to as the **principal polar representation of z**.

For any angle θ, there is a "lazy" notation for the complex number that appears in parentheses in the right-hand side of (5.2.1): we denote the complex number $\cos\theta + i\sin\theta$ simply by $\textbf{cis}\,\theta$.

Example 5.2.1. Suppose we are asked to find *all* (including the *principal*) polar representations of the complex number $z = 4 - 4i$.

We start of with the computation of the modulus:

$$r = |z| = \sqrt{(\text{Re}\, z)^2 + (\text{Im}\, z)^2} = \sqrt{4^2 + (-4)^2} = \sqrt{32} = 4\sqrt{2}.$$

Then we compute the principal argument (using the fact that $\text{Im}\, z$ is *negative*):

$$\tau = -\arccos\left(\frac{\text{Re}\, z}{|z|}\right) = -\arccos\left(\frac{4}{4\sqrt{2}}\right) = -\arccos\left(\frac{1}{\sqrt{2}}\right) = -\frac{\pi}{4}.$$

So the *principal polar representation of z* is:

$$4 - 4i = 4\sqrt{2}\,\textbf{cis}\left(-\frac{\pi}{4}\right).$$

It is worth pointing out that, since our number is relatively easy to plot as a point in the coordinate plane, the principal argument τ can also be found *geometrically*, by inspecting the picture:

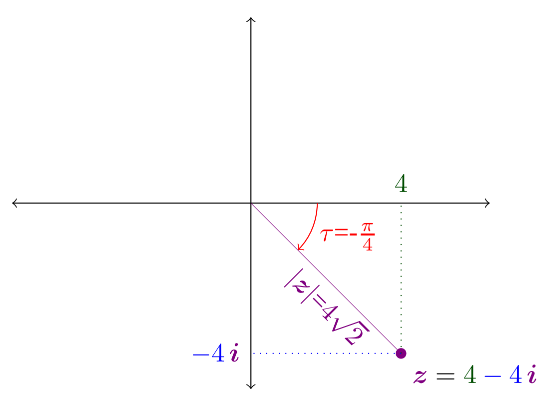


Figure 5.2.3

As for *all polar representations of z*, they are:

$$4 - 4\,\boldsymbol{i} = 4\sqrt{2}\,\mathbf{cis}\left(-\frac{\pi}{4}+2n\pi\right),\ n \text{ integer}.$$

As far as the complex numbers of the form $\mathbf{cis}\,\theta$ are concerned, the following three features are very useful.

$$\left|\mathbf{cis}\,\theta\right| = 1. \tag{5.2.4}$$

$$\overline{\mathbf{cis}\,\theta} = \mathbf{cis}\,(-\theta) = \frac{1}{\mathbf{cis}\,\theta}. \tag{5.2.5}$$

CLARIFICATION. The equality (5.2.4) follows immediately from the definition:

$$\left|\mathbf{cis}\,\theta\right| = \sqrt{(\mathrm{Re}\,\mathbf{cis}\,\theta)^2 + (\mathrm{Im}\,\mathbf{cis}\,\theta)^2} = \sqrt{\cos^2\theta + \sin^2\theta} = 1.$$

The first equality in (5.2.5) follows again directly from the definition and from the formulas for negatives:

$$\mathbf{cis}\,(-\theta) = \cos(-\theta) + \boldsymbol{i}\sin(-\theta) = \cos\theta - \boldsymbol{i}\sin\theta = \overline{\cos\theta + \boldsymbol{i}\sin\theta} = \overline{\mathbf{cis}\,\theta}.$$

Using (5.2.4), the equality between the left-most and the right-most sides of (5.2.5) now follows immediately from the easy division formula (5.1.7) from Section 5.1.

Multiplication and Division With Polar Representations

In Section 5.1 we learned that complex number multiplication corresponds to matrix multiplications, under the correspondence

$$a + b\,\boldsymbol{i} \longleftrightarrow \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

When we specialize this to a complex number of the form $\mathbf{cis}\,\theta$, the corresponding matrix is precisely the *rotation matrix* $\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, so using the Product Rule for Rotation Matrices, we immediately get the identity:

$$\mathbf{cis}\,\alpha\,\mathbf{cis}\,\beta \;=\; \mathbf{cis}\,(\alpha+\beta). \tag{5.2.6}$$

CLARIFICATION. Of course, we do not need the matrix correspondence to derive (5.2.6), because we can derive it directly from the Angle Addition Formulas for cosine and sine. However, we can also *use* (5.2.6) to help us *memorize the Angle Addition Formulas*, which we can now present in a very nice and concise way as:

$$\cos(\alpha+\beta) + \boldsymbol{i}\sin(\alpha+\beta) = (\cos\alpha + \boldsymbol{i}\sin\alpha)(\cos\beta + \boldsymbol{i}\sin\beta) \tag{5.2.7}$$

Similarly, if we rewrite the above formula with β replaced by $-\beta$, we also get a formula which we can use for memorizing the *Angle Subtraction Formulas*

$$\cos(\alpha-\beta) + \boldsymbol{i}\sin(\alpha-\beta) = (\cos\alpha + \boldsymbol{i}\sin\alpha)(\cos\beta - \boldsymbol{i}\sin\beta), \tag{5.2.8}$$

which comes from the division identity

$$\frac{\mathbf{cis}\,\alpha}{\mathbf{cis}\,\beta} = \mathbf{cis}\,(\alpha - \beta). \tag{5.2.9}$$

With the help of formulas (5.2.6) and (5.2.9), we can also derive the following formulas that allow us to express products and ratios of complex numbers.

Polar Forms of Multiplication and Division

Given two complex numbers presented in polar form as $z_1 = |z_1|\,\mathbf{cis}\,\theta_1$ and $z_2 = |z_1|\,\mathbf{cis}\,\theta_2$, their product and ratio can also be presented in polar form as:

$$z_1 z_2 = |z_1| \cdot |z_2| \cdot \mathbf{cis}\,(\theta_1 + \theta_2); \tag{5.2.10}$$

$$\frac{z_1}{z_2} = \frac{|z_1|}{|z_2|} \cdot \mathbf{cis}\,(\theta_1 - \theta_2), \quad (z_2 \neq 0).. \tag{5.2.11}$$

Example 5.2.2. Suppose we are given complex numbers $z_1 = 3\,\mathbf{cis}\,\frac{\pi}{3}$, $z_2 = 2\,\mathbf{cis}\,\frac{\pi}{6}$, and we want to compute their product and their ratio.

Using the formulas (5.2.10) and (5.2.11), we can directly compute:

$$z_1 z_2 = 3 \cdot 2 \cdot \mathbf{cis}\left(\frac{\pi}{3} + \frac{\pi}{6}\right) = 6\,\mathbf{cis}\,\frac{\pi}{2} = 6(\cos\frac{\pi}{2} + \boldsymbol{i}\sin\frac{\pi}{2}) = 6\,\boldsymbol{i};$$

$$\frac{z_1}{z_2} = \frac{3}{2} \cdot \mathbf{cis}\left(\frac{\pi}{3} - \frac{\pi}{6}\right) = \frac{3}{2}\,\mathbf{cis}\,\frac{\pi}{6} = \frac{3}{2}(\cos\frac{\pi}{6} + \boldsymbol{i}\sin\frac{\pi}{6}) = \frac{3}{2}\left(\frac{\sqrt{3}}{2} + \frac{1}{2}\boldsymbol{i}\right) = \frac{3\sqrt{3}}{4} + \frac{3}{4}\boldsymbol{i}$$

Of course, we can also write both our numbers algebraically as:

$$z_1 = 3\left(\cos\frac{\pi}{3} + \boldsymbol{i}\sin\frac{\pi}{3}\right) = 3\left(\frac{1}{2} + \frac{\sqrt{3}}{2}\boldsymbol{i}\right) = \frac{3}{2} + \frac{3\sqrt{3}}{2}\boldsymbol{i},$$

$$z_2 = 2\left(\cos\frac{\pi}{6} + \boldsymbol{i}\sin\frac{\pi}{6}\right) = 2\left(\frac{\sqrt{3}}{2} + \frac{1}{2}\boldsymbol{i}\right) = \sqrt{3} + \boldsymbol{i},$$

and we can compute the product and ratio algebraically. Whether we use one method or the other, it is up to us.

Exercises

In Exercises 1-8 you are asked to find all polar representations of the given complex number, and indicate the principal one. Use **exact values**

1. $z = 3\,\boldsymbol{i}$.

2. $z = -12\,\boldsymbol{i}$.

3. $z = 5$.

4. $z = -15$.

5. $z = 10 - 10\,i$.

6. $z = 7 + 7\,i$.

7. $z = 2\sqrt{3} - 2\,i$.

8. $z = 4 + 4\sqrt{3}\,i$.

In Exercises 9-13 you are express the given complex number, which is presented in polar form, in *algebraic form:* $a + b\,i$. Use **exact** values.

9. $z = 8\mathbf{cis}\,\frac{\pi}{2}$.

10. $z = 4\mathbf{cis}\,\frac{3\pi}{2}$.

11. $z = 2\mathbf{cis}\,\frac{3\pi}{4}$.

12. $z = 6\mathbf{cis}\,\frac{5\pi}{6}$.

13. $z = 2\mathbf{cis}\,\frac{7\pi}{3}$.

In Exercises 14-18 you are asked to compute a product, or a ratio of two complex numbers in **two** ways:

(i) Algebraically.

(ii) Using polar representations, that is, by presenting each one of the numbers, as well as the result, in polar form.

You can use the two calculations to verify you got the right answer.

14. $(4\sqrt{3} - 4\,i)(2 - 2\sqrt{3}\,i)$.

15. $(4 - 4\,i)(2 - 2\,i)$.

16. $\dfrac{4\sqrt{3} - 4\,i}{2 + 2\sqrt{3}\,i}$.

17. $\dfrac{8\,i}{2\sqrt{3} - 2\,i}$.

18. $\dfrac{-6\,i}{3 + 3\,i}$.

5.3 Powers and Roots

In this section we learn how to use polar representations of complex numbers in order to deal with powers and power equations.

De Moivre's Formula

As we learned in Section 5.2, polar representations are particularly nice, with respect to the multiplication operation, as seen in formula (5.2.10), which reads:

$$(r_1 \operatorname{cis} \theta_1)(r_2 \operatorname{cis} \theta_2) = (r_1 r_2) \operatorname{cis} (\theta_1 + \theta_2).$$

We can, of course apply this rule repeatedly (for a product of more than two factors). In particular, when we multiply one number with itself several times we obtain the following simple rule.

De Moivre's Formula

$$[r(\cos\theta + i\sin\theta)]^N = r^N(\cos N\theta + i\sin N\theta), \quad N \text{ positive integer} \tag{5.3.1}$$

CLARIFICATIONS. De Moivre's Formula can also be expanded to allow *negative* exponents, in which case we can also write

$$[r(\cos\theta + i\sin\theta)]^{-N} = r^{-N}(\cos N\theta - i\sin N\theta), \quad N \text{ positive integer}.$$

Example 5.3.1. Suppose we want to compute a large power, for instance: $(1-i)^{10}$ Although we can work this calculation algebraically, using De Moivre's Formula, our work will be greatly simplified.

We start off by finding a polar representation of the complex number $z = 1-i$, which has $\operatorname{Re} z = 1$ and $\operatorname{Im} z = -1$

We begin by computing the modulus:

$$r = |z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2} = \sqrt{1^2 + (-1)^2} = \sqrt{2}.$$

As for an argument, since $\operatorname{Im} z$ is *negative*, the principal argument is:

$$\tau = -\arccos\left(\frac{\operatorname{Re} z}{|z|}\right) = -\arccos\left(\frac{1}{\sqrt{2}}\right) = -\frac{\pi}{4}.$$

Of course, we could have also found the principal polar representation of $1-i$ geometrically, by inspecting this picture:

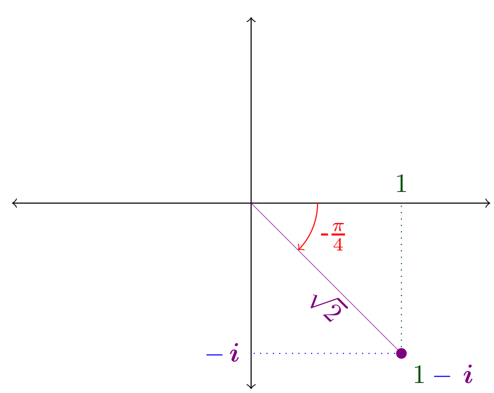


Figure 5.3.1

With all these preparations, we now have $1-\boldsymbol{i} = \sqrt{2}\,\mathbf{cis}\left(-\frac{\pi}{4}\right) = \sqrt{2}\left(\cos\left(-\frac{\pi}{4}\right) + \boldsymbol{i}\sin\left(-\frac{\pi}{4}\right)\right)$, so using De Moivre's Formula we get:

$$
\begin{aligned}
(1-\boldsymbol{i})^{10} &= (\sqrt{2})^{10}\left(\cos\left(10\left(-\frac{\pi}{4}\right)\right) + \boldsymbol{i}\sin\left(10\left(-\frac{\pi}{4}\right)\right)\right) = \\
&= 32\left(\cos\left(-\frac{5\pi}{2}\right) + \boldsymbol{i}\sin\left(-\frac{5\pi}{2}\right)\right) = 32\,(0+(-1)\,\boldsymbol{i}) = -32\,\boldsymbol{i}.
\end{aligned}
$$

Application to Fourier Expansions

In Section 4.3 we learned how to use Product-to-Sum Formulas to obtain the so-called *Fourier expansions* for various functions, which are sums of the form:

$$
\begin{aligned}
\mathsf{F}(t) = a_0 + a_1\cos t + a_2\cos 2t + a_3\cos 3t + \ldots + a_N\cos Nt + \\
+ b_1\sin t + b_2\sin 2t + b_3\sin 3t + \ldots + b_N\sin Nt,
\end{aligned}
\tag{5.3.2}
$$

As it turns out, another way to obtain these expansions is to use De Moivre's Formula, for complex numbers of the form

$$\boldsymbol{u} = \mathbf{cis}\,t = \cos t + \boldsymbol{i}\sin t. \tag{5.3.3}$$

A complex number of this form is called *unimodular*, because the fact that $\boldsymbol{u}$ can be presented as in (5.3.3) is equivalent to the condition $|\boldsymbol{u}| = 1$. The main features of unimodular numbers are collected in the following formula package.

Unimodular Number Identities

Given $\boldsymbol{u}$ and t satisfying (5.3.3), then for any integer N, the following identities hold:

$$\boldsymbol{u}^N = \cos Nt + \boldsymbol{i}\sin Nt \tag{5.3.4}$$

$$\cos Nt = \frac{1}{2}\left(\boldsymbol{u}^N + \boldsymbol{u}^{-N}\right) \tag{5.3.5}$$

$$\sin Nt = \frac{1}{2i}\left(\boldsymbol{u}^N - \boldsymbol{u}^{-N}\right) \tag{5.3.6}$$

CLARIFICATION. Using unimodular numbers allows us to switch back and forth between *Fourier sums* like (5.3.2) and the so-called *Laurent sums*, which are sums of expressions of the form $c_k\boldsymbol{u}^k$, with complex coefficients c_k and integer (possibly negative) exponents k. This method makes it very easy to compute Fourier expansions, and also makes it clear why they are *unique*.

Example 5.3.2. Suppose we want to find the Fourier expansion of $f(t) = 4\cos t\cos 2t\sin 3t$. (Compare with Exercise 17 from Section 4.3.)

We start off by setting up the Laurent sum, using the numbers $\boldsymbol{u} = \cos t + \boldsymbol{i}\sin t$, so by the above identities we can present our function as:

$$\begin{aligned}
f(t) &= 4\cos t \cos 2t \sin 3t = 4\left(\frac{1}{2}\left(\boldsymbol{u}+\boldsymbol{u}^{-1}\right)\right)\left(\frac{1}{2}\left(\boldsymbol{u}^2+\boldsymbol{u}^{-2}\right)\right)\left(\frac{1}{2i}\left(\boldsymbol{u}^3-\boldsymbol{u}^{-3}\right)\right) = \\
&= \frac{4}{8i}\left(\boldsymbol{u}+\boldsymbol{u}^{-1}\right)\left(\boldsymbol{u}^2+\boldsymbol{u}^{-2}\right)\left(\boldsymbol{u}^3-\boldsymbol{u}^{-3}\right) = \frac{1}{2i}\left(\boldsymbol{u}^3+\boldsymbol{u}+\boldsymbol{u}^{-1}+\boldsymbol{u}^{-3}\right)\left(\boldsymbol{u}^3-\boldsymbol{u}^{-3}\right) = \\
&= \frac{1}{2i}\left(\boldsymbol{u}^6+\boldsymbol{u}^4+\boldsymbol{u}^2+1-1-\boldsymbol{u}^{-2}-\boldsymbol{u}^{-4}-\boldsymbol{u}^{-6}\right) = \\
&= \frac{1}{2i}\left(\boldsymbol{u}^6+\boldsymbol{u}^4+\boldsymbol{u}^2-\boldsymbol{u}^{-2}-\boldsymbol{u}^{-4}-\boldsymbol{u}^{-6}\right).
\end{aligned}$$

But now, by grouping the matching powers of $\boldsymbol{u}$, we can also write

$$\begin{aligned}
f(t) &= \frac{1}{2i}\left(\boldsymbol{u}^6+\boldsymbol{u}^4+\boldsymbol{u}^2-\boldsymbol{u}^{-2}-\boldsymbol{u}^{-4}-\boldsymbol{u}^{-6}\right) = \\
&= \frac{1}{2i}\left(\boldsymbol{u}^6-\boldsymbol{u}^{-6}\right)+\frac{1}{2i}\left(\boldsymbol{u}^4-\boldsymbol{u}^{-4}\right)+\frac{1}{2i}\left(\boldsymbol{u}^2-\boldsymbol{u}^{-2}\right) = \\
&= \sin 6t \;+\; \sin 4t \;+\; \sin 2t.
\end{aligned}$$

Roots

Using De Moivre's Formula, we can also tackle *power equations*, which are those of the form

$$x^N = \boldsymbol{w}, \tag{5.3.7}$$

where $N \geq 2$ is some integer, and $\boldsymbol{w}$ is some (fixed) complex number. The unknown quantity is, of course x, which in most cases will be complex too. The solutions of the power equation (5.3.7) are called the *N^{th} roots of $\boldsymbol{w}$*.

When we want to understand the algebraic features of power equations, it is helpful to start with the special case when $\boldsymbol{w} = 1$, when the solutions will be referred to as the *N^{th} roots of unity*.

The Roots of Unity Theorem

For any integer $N \geq 2$, there are exactly N distinct roots of unity, listed as $\boldsymbol{u}_0, \boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{N-1}$, which are given by the formula:

$$\boldsymbol{u}_k = \mathbf{cis}\,\frac{2k\pi}{N} = \cos\frac{2k\pi}{N} + \boldsymbol{i}\sin\frac{2k\pi}{N}, \quad k = 0, 1, 2, \dots, N-1. \tag{5.3.8}$$

Proof. If we consider the angles $\theta_0, \theta_1, \dots, \theta_{N-1}$, given by $\theta_k = \dfrac{2k\pi}{N}$, $k = 0, 1, 2, \dots, N-1$, then it is pretty clear that we have the inequalities

$$0 = \theta_0 < \theta_1 < \theta_2 < \ \dots\ < \theta_{N-1} < 2\pi,$$

so all numbers given by (5.3.8), which are of the form $\boldsymbol{u}_k = \mathbf{cis}\,\theta_k$, are *distinct*.

Since for each such angle, we have $N\theta_k = 2k\pi$, using the Unimodular Number Formulas (a special case of De Moivre's Formula) it follows that

$$(\boldsymbol{u}_k)^N = \cos 2k\pi + \boldsymbol{i}\sin 2k\pi \;=\; 1, \quad k = 0, 1, 2, \dots, N-1,$$

so *the numbers $\boldsymbol{u}_0, \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N-1}$ are roots of unity*.

Putting everything together, we have found N *distinct solutions* of the equation $x^N = 1$, which means that we have found N *distinct zeros* of the polynomial $x^N - 1$. By the Fundamental Theorem of Algebra, it follows that this polynomial *factors* as:

$$x^N - 1 = (x - \boldsymbol{u}_0)(x - \boldsymbol{u}_1)(x - \boldsymbol{u}_2) \cdots (x - \boldsymbol{u}_{N-1}),$$

thus $\boldsymbol{u}_0, \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N-1}$ are *the only solutions* of $x^N - 1 = 0$. □

CLARIFICATIONS. Roots of unity are always *unimodular*. No matter what N is, we see the the first number $\boldsymbol{u}_0$ produced by formula (5.3.8) is *always equal to* 1.

For small N, the roots of unity can be easily found directly, by solving the power equation $x^N = 1$ *algebraically*. For instance, the *square roots of unity* (the case $N = 2$) are simply ± 1. For larger values of N, although it might still be possible to solve the power equation algebraically, the trigonometric approach might work a little faster.

Example 5.3.3. The *cubic roots of unity*, which correspond to $N = 3$, are:

$$\begin{aligned}
\boldsymbol{u}_0 &= \cos\frac{2 \cdot 0 \cdot \pi}{3} + \boldsymbol{i}\sin\frac{2 \cdot 0 \cdot \pi}{3} = \cos 0 + \boldsymbol{i}\sin 0 = 1; \\
\boldsymbol{u}_1 &= \cos\frac{2 \cdot 1 \cdot \pi}{3} + \boldsymbol{i}\sin\frac{2 \cdot 1 \cdot \pi}{3} = \cos\frac{2\pi}{3} + \boldsymbol{i}\sin\frac{2\pi}{3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}\boldsymbol{i}; \\
\boldsymbol{u}_2 &= \cos\frac{2 \cdot 2 \cdot \pi}{3} + \boldsymbol{i}\sin\frac{2 \cdot 2 \cdot \pi}{3} = \cos\frac{4\pi}{3} + \boldsymbol{i}\sin\frac{4\pi}{3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}\boldsymbol{i}.
\end{aligned}$$

These same numbers can also be obtained by solving the equation $x^3 - 1 = 0$ using the factoring formula for difference of cubes:

$$(x - 1)(x^2 + x + 1) = 0.$$

Example 5.3.4. The *quartic (4^{th}) roots of unity* are:

$$\begin{aligned}
\boldsymbol{u}_0 &= \cos\frac{2 \cdot 0 \cdot \pi}{4} + \boldsymbol{i}\sin\frac{2 \cdot 0 \cdot \pi}{4} = \cos 0 + \boldsymbol{i}\sin 0 = 1; \\
\boldsymbol{u}_1 &= \cos\frac{2 \cdot 1 \cdot \pi}{4} + \boldsymbol{i}\sin\frac{2 \cdot 1 \cdot \pi}{4} = \cos\frac{\pi}{2} + \boldsymbol{i}\sin\frac{\pi}{2} = \boldsymbol{i}; \\
\boldsymbol{u}_2 &= \cos\frac{2 \cdot 2 \cdot \pi}{4} + \boldsymbol{i}\sin\frac{2 \cdot 2 \cdot \pi}{4} = \cos\pi + \boldsymbol{i}\sin\pi = -1 \\
\boldsymbol{u}_3 &= \cos\frac{2 \cdot 3 \cdot \pi}{4} + \boldsymbol{i}\sin\frac{2 \cdot 3 \cdot \pi}{4} = \cos\frac{3\pi}{2} + \boldsymbol{i}\sin\frac{3\pi}{2} = -\boldsymbol{i}.
\end{aligned}$$

The General Roots Theorem

If $N \geq 2$ is an integer, and $\boldsymbol{w}$ is some non-zero complex number, then there are exactly N distinct roots of $\boldsymbol{w}$. Furthermore:

I. *If $\boldsymbol{z}$ is one particular root, then all roots are of the form*

$$\boldsymbol{z}\boldsymbol{u}_0, \boldsymbol{z}\boldsymbol{u}_1, \ldots, \boldsymbol{z}\boldsymbol{u}_{N-1}, \tag{5.3.9}$$

where $\boldsymbol{u}_0, \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N-1}$ are the N^{th} roots of unity.

II. *If θ is an argument for $\boldsymbol{w}$, then:*

(A) *one particular N^{th} root of $\boldsymbol{w}$ is the number:* $\sqrt[N]{|\boldsymbol{w}|}\left(\cos\frac{\theta}{N} + \boldsymbol{i}\sin\frac{\theta}{N}\right)$;

(B) *all the N distinct N^{th} roots of $\boldsymbol{w}$ can be listed as $\boldsymbol{z}_0, \boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{N-1}$, with the numbers given by the formula:*

$$\boldsymbol{z}_k = \sqrt[N]{|\boldsymbol{w}|}\left(\cos\frac{\theta + 2k\pi}{N} + \boldsymbol{i}\sin\frac{\theta + 2k\pi}{N}\right), \quad k = 0, 1, 2, \ldots, N-1. \tag{5.3.10}$$

Proof. Statement I is pretty clear, because, as we argued in the Roots of Unity Theorem, the polynomial $x^N - \boldsymbol{w}$ can only have at most N distinct zeros, and if $\boldsymbol{z}$ is one of them, then so are $\boldsymbol{z}\boldsymbol{u}_0, \boldsymbol{z}\boldsymbol{u}_1, \ldots, \boldsymbol{z}\boldsymbol{u}_{N-1}$, simply because

$$(\boldsymbol{z}\boldsymbol{u}_k)^N = \boldsymbol{z}^N (\boldsymbol{u}_k)^N = \boldsymbol{w} \cdot 1 = \boldsymbol{w}.$$

(Of course, since the $\boldsymbol{u}$'s are all distinct, so will be the numbers $\boldsymbol{z}\boldsymbol{u}_0, \boldsymbol{z}\boldsymbol{u}_1, \ldots, \boldsymbol{z}\boldsymbol{u}_{N-1}$.)

For statement II.(i), we simply use De Moivre's Formula:

$$\left[\sqrt[N]{|\boldsymbol{w}|}\left(\cos\frac{\theta}{N} + \boldsymbol{i}\sin\frac{\theta}{N}\right)\right]^N = \left(\sqrt[N]{|\boldsymbol{w}|}\right)^N\left(\cos N\frac{\theta}{N} + \boldsymbol{i}\sin N\frac{\theta}{N}\right) =$$
$$= |\boldsymbol{w}|\,(\cos\theta + \boldsymbol{i}\sin\theta) = \boldsymbol{w}.$$

Statement II.(ii) follows now from statement I, because we know that every root is of the form $\boldsymbol{z}\boldsymbol{u}_k$ (where $\boldsymbol{z}\boldsymbol{u}_k$ is the particular one constructed above), and by the Product Formula (5.2.10) we can obviously write

$$\boldsymbol{z}\boldsymbol{u}_k = \left[\sqrt[N]{|\boldsymbol{w}|}\,\mathbf{cis}\,\frac{\theta}{N}\right] \cdot \left[\mathbf{cis}\,\frac{2k\pi}{N}\right] = \sqrt[N]{|\boldsymbol{w}|}\,\mathbf{cis}\left(\frac{\theta}{N} + \frac{2k\pi}{N}\right) = \boldsymbol{z}_k$$ □

In choosing an argument θ for $\boldsymbol{w}$, we have quite a bit of freedom. For example, we can choose τ, the *principal argument* for $\boldsymbol{w}$, but we can also choose θ to be of the form "τ plus some multiple of 2π." If we choose two different arguments, the formula (5.3.10) will work slightly differently, but in the end will produce the *exact same numbers*, possibly listed differently. Example 5.3.5 below illustrates this phenomenon.

Example 5.3.5. Suppose we want to compute the *cubic roots* of $\boldsymbol{w} = -8\boldsymbol{i}$.

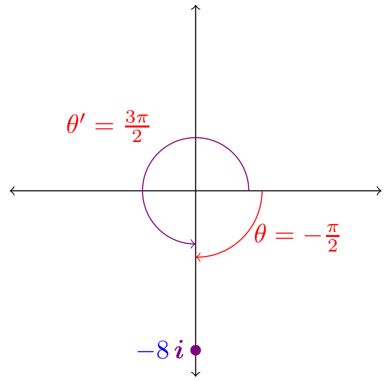


Figure 5.3.2

Pictorially, we can easily find the modulus $|w| = 8$ and principal argument $\theta = -\dfrac{\pi}{2}$. The above picture shows another valid argument, in the form $\theta' = \dfrac{3\pi}{2}$.

Using formula (5.3.10) with $\theta = -\dfrac{\pi}{2}$, the three cubic roots of $-8i$ are:

$$z_0 = \sqrt[3]{8}\left(\cos\frac{-\frac{\pi}{2} + 2\cdot 0\cdot\pi}{3} + i\sin\frac{-\frac{\pi}{2} + 2\cdot 0\cdot\pi}{3}\right) = 2\left(\cos\left(-\frac{\pi}{6}\right) + i\sin\left(-\frac{\pi}{6}\right)\right) =$$
$$= 2\left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right) = \sqrt{3} - i;$$
$$z_1 = \sqrt[3]{8}\left(\cos\frac{-\frac{\pi}{2} + 2\cdot 1\cdot\pi}{3} + i\sin\frac{-\frac{\pi}{2} + 2\cdot 1\cdot\pi}{3}\right) = 2\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right) = 2i;$$
$$z_2 = \sqrt[3]{8}\left(\cos\frac{-\frac{\pi}{2} + 2\cdot 2\cdot\pi}{3} + i\sin\frac{-\frac{\pi}{2} + 2\cdot 2\cdot\pi}{3}\right) = 2\left(\cos\frac{7\pi}{6} + i\sin\frac{7\pi}{6}\right) =$$
$$= 2\left(-\frac{\sqrt{3}}{2} - \frac{1}{2}i\right) = -\sqrt{3} - i.$$

What happens if instead of θ we use $\theta' = \dfrac{3\pi}{2}$? Using this value, formula (5.3.10) produces the following values:

$$z_0' = \sqrt[3]{8}\left(\cos\frac{\frac{3\pi}{2} + 2\cdot 0\cdot\pi}{3} + i\sin\frac{\frac{3\pi}{2} + 2\cdot 0\cdot\pi}{3}\right) = 2\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right) = 2i;$$
$$z_1' = \sqrt[3]{8}\left(\cos\frac{\frac{3\pi}{2} + 2\cdot 1\cdot\pi}{3} + i\sin\frac{\frac{3\pi}{2} + 2\cdot 1\cdot\pi}{3}\right) = 2\left(\cos\frac{7\pi}{6} + i\sin\frac{7\pi}{6}\right) =$$
$$= 2\left(-\frac{\sqrt{3}}{2} - \frac{1}{2}i\right) = -\sqrt{3} - i;$$
$$z_2' = \sqrt[3]{8}\left(\cos\frac{\frac{3\pi}{2} + 2\cdot 2\cdot\pi}{3} + i\sin\frac{\frac{3\pi}{2} + 2\cdot 2\cdot\pi}{3}\right) = 2\left(\cos\frac{11\pi}{6} + i\sin\frac{11\pi}{6}\right) =$$
$$= 2\left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right) = \sqrt{3} - i.$$

As we see, we get the exact same three numbers: $\sqrt{3} - \boldsymbol{i}$, $-\sqrt{3} - \boldsymbol{i}$, and $2\,\boldsymbol{i}$, except that they appear in different order.

Finally, there is yet a third way to find our roots, using one particular root $\boldsymbol{z} = 2\,\boldsymbol{i}$, by combining statement II.(i) from the General Roots Theorem, with what we already know the cubic roots of unity which we found in Example 5.3.3. So our roots can also be computed as:

$$\begin{aligned} \boldsymbol{z}\boldsymbol{u}_0 &= (2\,\boldsymbol{i}) \cdot 1 = 2\,\boldsymbol{i}; \\ \boldsymbol{z}\boldsymbol{u}_1 &= (2\,\boldsymbol{i}) \cdot \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}\,\boldsymbol{i} \right) = -\sqrt{3} - \boldsymbol{i}; \\ \boldsymbol{z}\boldsymbol{u}_2 &= (2\,\boldsymbol{i}) \cdot \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}\,\boldsymbol{i} \right) = \sqrt{3} - \boldsymbol{i}. \end{aligned}$$

Example 5.3.6. Suppose we want to find all (real and complex) solutions of the equation

$$x^8 + 15x^4 - 16 = 0. \tag{5.3.11}$$

We start off by making the substitution $x^4 = \boldsymbol{w}$, after which our equation becomes:

$$\boldsymbol{w}^2 + 15\boldsymbol{w} - 16 = 0. \tag{5.3.12}$$

Using the Quadratic Formula, the above can be easily solved:

$$\boldsymbol{w} = \frac{-15 \pm \sqrt{15^2 - 4 \cdot 1 \cdot (-16)}}{2 \cdot 1} = \frac{-15 \pm \sqrt{289}}{2} = \frac{-15 \pm 17}{2},$$

so we have two solutions

$$\boldsymbol{w}_1 = \frac{-15 - 17}{2} = -\frac{32}{2} = -16; \qquad \boldsymbol{w}_2 = \frac{-15 + 17}{2} = \frac{2}{2} = 1.$$

With each one of these two solutions we will go back to our substitution.

The first solution of (5.3.12), which is $\boldsymbol{w}_1 = -16$, yields (by going back to the substitution) the equation $x^4 = -16$.

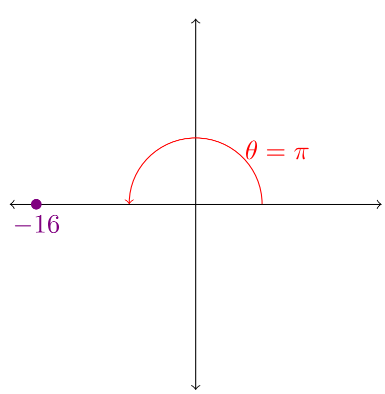


Figure 5.3.3

Pictorially, we can easily find the modulus $|\boldsymbol{w}_1| = 16$ and principal argument $\theta = \pi$. Using formula (5.3.10), the four quartic roots of -16 are:

$$\begin{aligned}
\boldsymbol{z}_0 &= \sqrt[4]{16}\left(\cos\frac{\pi+2\cdot 0\cdot\pi}{4} + \boldsymbol{i}\sin\frac{\pi+2\cdot 0\cdot\pi}{4}\right) = 2\left(\cos\left(\frac{\pi}{4}\right) + \boldsymbol{i}\sin\left(\frac{\pi}{4}\right)\right) = \\
&= 2\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\boldsymbol{i}\right) = \sqrt{2} + \sqrt{2}\boldsymbol{i}; \\
\boldsymbol{z}_1 &= \sqrt[4]{16}\left(\cos\frac{\pi+2\cdot 1\cdot\pi}{4} + \boldsymbol{i}\sin\frac{\pi+2\cdot 1\cdot\pi}{4}\right) = 2\left(\cos\left(\frac{3\pi}{4}\right) + \boldsymbol{i}\sin\left(\frac{3\pi}{4}\right)\right) = \\
&= 2\left(-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\boldsymbol{i}\right) = -\sqrt{2} + \sqrt{2}\boldsymbol{i}; \\
\boldsymbol{z}_2 &= \sqrt[4]{16}\left(\cos\frac{\pi+2\cdot 2\cdot\pi}{4} + \boldsymbol{i}\sin\frac{\pi+2\cdot 2\cdot\pi}{4}\right) = 2\left(\cos\left(\frac{5\pi}{4}\right) + \boldsymbol{i}\sin\left(\frac{5\pi}{4}\right)\right) = \\
&= 2\left(-\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}\boldsymbol{i}\right) = -\sqrt{2} - \sqrt{2}\boldsymbol{i}; \\
\boldsymbol{z}_3 &= \sqrt[4]{16}\left(\cos\frac{\pi+2\cdot 3\cdot\pi}{4} + \boldsymbol{i}\sin\frac{\pi+2\cdot 3\cdot\pi}{4}\right) = 2\left(\cos\left(\frac{7\pi}{4}\right) + \boldsymbol{i}\sin\left(\frac{7\pi}{4}\right)\right) = \\
&= 2\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}\boldsymbol{i}\right) = \sqrt{2} - \sqrt{2}\boldsymbol{i}.
\end{aligned}$$

The second solution of (5.3.12) is $\boldsymbol{w}_2 = 1$, which (by going back to our substitution) yields the equation $x^4 = 1$, which has as solutions the quartic roots of unity, so based on what we found out in Example 5.3.4, these solutions will be ± 1 and $\pm\boldsymbol{i}$.

CONCLUSION: Our equation (5.3.11) has *eight solutions*:

(i) The solutions of $x^4 = -16$, which are: $\boldsymbol{z}_0 = \sqrt{2} + \sqrt{2}\boldsymbol{i}$, $\boldsymbol{z}_1 = -\sqrt{2} + \sqrt{2}\boldsymbol{i}$, $\boldsymbol{z}_2 = -\sqrt{2} - \sqrt{2}\boldsymbol{i}$, $\boldsymbol{z}_3 = \sqrt{2} - \sqrt{2}\boldsymbol{i}$, together with

(ii) The solutions of $x^4 = 1$, which are: $\boldsymbol{u}_0 = 1$, $\boldsymbol{u}_1 = \boldsymbol{i}$, $\boldsymbol{u}_2 = -1$, $\boldsymbol{u}_3 = -\boldsymbol{i}$.

Application: Complex Square Roots Formula

Using the General Roots Theorem, we can now devise a method for computing *square roots* of complex numbers. To be more specific, suppose we start with some complex number $\boldsymbol{w}$, and we want to solve the power equation: $x^2 = \boldsymbol{w}$. Assume, for simplicity, that $\boldsymbol{w}$ *is not real*. (The case when $\boldsymbol{w}$ *is real* will be treated separately.) According to the recipe set forth by statement II.(i) from the General Roots Theorem, we only need to find *one* root, for which we only need one argument for $\boldsymbol{w}$. The easiest choice is the *principal* argument, which is given as:

$$\theta = (\text{sign of Im}\,\boldsymbol{w})\arccos\left(\frac{\text{Re}\,\boldsymbol{w}}{|\boldsymbol{w}|}\right),$$

which by construction satisfies

$$\cos\theta = \frac{\text{Re}\,\boldsymbol{w}}{|\boldsymbol{w}|}. \tag{5.3.13}$$

By construction (recall that w is assumed to be not real), we also know that the principal argument θ sits in the interval $(-\pi, \pi)$, so $\theta/2$ sits in the interval $(-\pi/2, \pi/2)$. In particular, by the Half-Angle Formulas, the principal argument θ will satisfy:

$$\cos(\theta/2) = \sqrt{\frac{1}{2}(1+\cos\theta)} = \sqrt{\frac{1}{2}\left(1+\frac{\text{Re}\,w}{|w|}\right)}.$$

As for $\sin(\theta/2)$, again by the Half-Angle ormulas, its value will depend on the *sign of* Im w, as follows:

- If $\text{Im}\,w < 0$, then $\theta/2$ sits in the interval $(-\pi/2, 0)$, thus:

$$\sin(\theta/2) = -\sqrt{\frac{1}{2}(1-\cos\theta)} = -\sqrt{\frac{1}{2}\left(1-\frac{\text{Re}\,w}{|w|}\right)}$$

- If $\text{Im}\,w > 0$, then $\theta/2$ sits in the interval $(0, \pi/2)$, thus:

$$\sin(\theta/2) = \sqrt{\frac{1}{2}(1-\cos\theta)} = \sqrt{\frac{1}{2}\left(1-\frac{\text{Re}\,w}{|w|}\right)}$$

Therefore, the unified formula for $\sin(\theta/2)$ is:

$$\sin(\theta/2) = (\text{sign of Im}\,w)\sqrt{\frac{1}{2}\left(1-\frac{\text{Re}\,w}{|w|}\right)}.$$

Using the General Roots Theorem, we now obtain the following.

Complex Square Roots Formula

If w is a complex number which is *non-real*, then the two square roots of w are $\pm z$, where

$$z = \sqrt{\frac{|w|+\text{Re}\,w}{2}} + (\text{sign of Im}\,w)\,i\sqrt{\frac{|w|-\text{Re}\,w}{2}}. \tag{5.3.14}$$

If w is a complex number which is *real*, then its square roots are:
(i) $\pm\sqrt{w}$, if $w \geq 0$;
(ii) $\pm i\sqrt{|w|}$, if $w < 0$.

CLARIFICATION. The real case is pretty obvious. As for the non-real case, all we have to use is the General Roots Theorem, which provides us with a particular square root:

$$\begin{aligned} z &= \sqrt{|w|}\left[\cos(\theta/2) + i\sin(\theta/2)\right] = \\ &= \sqrt{|w|}\left[\sqrt{\frac{1}{2}\left(1+\frac{\text{Re}\,w}{|w|}\right)} + (\text{sign of Im}\,w)\,i\sqrt{\frac{1}{2}\left(1-\frac{\text{Re}\,w}{|w|}\right)}\right] = \\ &= \sqrt{\frac{|w|+\text{Re}\,w}{2}} + (\text{sign of Im}\,w)\,i\sqrt{\frac{|w|-\text{Re}\,w}{2}}. \end{aligned}$$

Example 5.3.7. Suppose we want to find the two square roots of $w = -12 - 16\,i$. Our number has $\operatorname{Re} w = -12$ and $\operatorname{Im} w = -16$, so its modulus is:

$$|w| = \sqrt{(\operatorname{Re} w)^2 + (\operatorname{Im} w)^2} = \sqrt{(-12)^2 + (-16)^2} = \sqrt{144 + 256} = \sqrt{400} = 20.$$

Since $\operatorname{Im} w$ is *negative*, using (5.3.14), it follows that *one* of the square roots of w is:

$$z = \sqrt{\frac{20 - 12}{2}} - i\sqrt{\frac{20 - (-12)}{2}} = \sqrt{4} - i\sqrt{16} = 2 - 4\,i.$$

Of course, the other square root of w is:

$$-z = -(2 - 4\,i) = -2 + 4\,i.$$

It is worth pointing out here that these roots were also computed in Example 5.1.4 from Section 5.1, with considerable effort. The calculation shown here demonstrates the efficiency of the General Root Theorem.

Exercises

In Exercises 1-6 you are asked to use De Moivre's Theorem to carry one certain power calculations.

1. Compute $(\sqrt{3} + i)^8$.
2. Compute $(\sqrt{3} - i)^{10}$.
3. Compute $(1 + \sqrt{3}i)^6$.
4. Compute $(1 - \sqrt{3}i)^8$.
5. Compute $(1 + i)^6$.
6. Compute $(1 - i)^{12}$.

In Exercises 7-14 you are asked to find all the specified root of certain real or complex numbers.

7. Find all 6^{th} roots of unity.
8. Find all 8^{th} roots of unity.
9. Find all cubic roots 27.
10. Find all cubic roots $27i$.
11. Find all quartic roots -256.
12. Find all 6^{th} roots of -64.
13. Find all square roots $-4i$.
14. Find all square roots $3 - 4i$.

In Exercises 15-19 you are asked to find all (real and complex) solutions of the given equation.

15. $x^4 - 3x^2 - 4 = 0$.

16. $x^6 + 7x^3 - 8 = 0$.

17. $x^6 - 26x^3 - 27 = 0$.

18. $x^8 - 97x^4 + 1296 = 0$.

19. $x^{12} - 9x^6 + 8 = 0$.

Chapter 6

Applications of Trigonometry in Analytic Geometry

In this Chapter we discuss applications of Trigonometry to Analytic Geometry, specifically to the classification of the curves represented by *quadratic equations* in two variables, which are those of the form:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \tag{6.0.1}$$

where A, B, C, D, E, F are constants.

As it turns out, all curves that are presented by such equations can be presented geometrically as *conic sections*, which are obtained by "slicing" a (double) cone

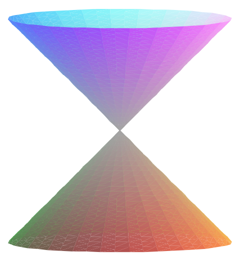
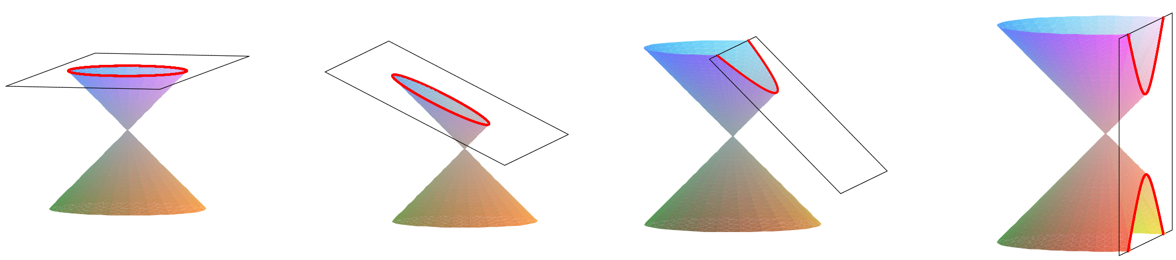


Figure 6.0.1

with a plane. When the "slicing" plane *does not contain the tip of the double cone*, the possible shapes of the section can be as follows:



(A) *circle* (B) *ellipse* (C) *parabola* (D) *hyperbola*

Figure 6.0.2

In cases (A) and (B) the section is a *closed curve*. Except for case (A), when the "slicing" plane is *perpendicular to the cone axis*, the resulting curve that arises in case (B) is what we call an *ellipse*. In case (C), which occurs when the "slicing" plane is *parallel to one of the cone generators*, the section consists of a *non-closed single piece curve*, which we call a *parabola*. In the remaining case (D), we see a *two-piece* section, which we call a *hyperbola*. The four types of curves obtained this way are called *non-degenerate conics*.

The exceptional cases, in which the set of points given by equation (6.0.1) is of neither one of the above forms, are as follows:

(i) *no points whatsoever*, for example $x^2 + 1 = 0$;
(ii) *a single point*, for example $x^2 + y^2 = 0$;
(iii) *one line*, for example $x^2 = 0$;
(iv) *two intersecting lines*, for example $x^2 - y^2 = 0$;
(v) *two parallel lines*, for example $x^2 - 1 = 0$.

All of these exceptional curves are called *degenerate conics*. (Some of them, such as (ii), (iii) and (v), can in fact be presented as honest conic sections, by using "slicing" planes that pass through the tip of the cone. However, cases (i) and (v) cannot be presented this way.)

Proving that all conics are represented by quadratic equations of the form (6.0.1) is quite technical, thus will be omitted it from this text, as we will not need it. Instead, our focus will be to understand the *geometry* of curves given by quadratic equations, particularly their *shapes*. In the first three Sections in this Chapter we will analyze the three types of curves (which correspond to the *non-degenerate* cases (A)-(C) above) directly, without any reference to their presentations as conic sections. For awhile, we will only deal with *special* quadratic equations, which are those that have $B = 0$. For this type of analysis we will not need any Trigonometry. However, Trigonometry will play a key role when we will consider *general* quadratic equations, as explained in Section 6.4, as well as in Section 6.5, where we will learn a bit about polar equations.

Since we are dealing quite a bit with quadratic equations, the reader is suggested to review the *easy square completion identity*

$$\boldsymbol{a}\boldsymbol{t}^{\mathbf{2}} + \boldsymbol{b}\boldsymbol{t} + \boldsymbol{c} = \boldsymbol{a}\left(\boldsymbol{t} + \frac{\boldsymbol{b}}{2\boldsymbol{a}}\right)^{\mathbf{2}} - \frac{\boldsymbol{D}}{4\boldsymbol{a}}. \tag{6.0.2}$$

(In the above formula, the symbol $\boldsymbol{t}$ designates some *variable*, and $\boldsymbol{D} = \boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}$ is the *discriminant*.)

The best way to introduce a "nice" curve is to define it as a *geometric locus*, that is, as a *set of points that share a common geometric property*. One such example of a geometric locus is for instance a *circle*.

6.1 Parabolas

Most of the readers have seen the word "parabola" as used to describe graphs of quadratic functions, thus given by equations of the form $y = ax^2 + bx + c$, where a, b, c are constants, with $a \neq 0$. In this section we will learn a lot more about such curves, in particular about their geometry.

Geometric Definition of Parabolas

As a *geometric locus*, a parabola is defined as follows.

> Given a *line* $\mathscr{D}$, and some *point* F *not sitting on* $\mathscr{D}$, the **parabola with focus F and directrix $\mathscr{D}$** is the *set of all points P in the plane, that are at equal distance from F and $\mathscr{D}$*, that is,
>
> $$\text{dist}(P, F) = \text{dist}(P, \mathscr{D}). \tag{6.1.1}$$

CLARIFICATIONS. As we shall soon see, if we know that a given curve $\mathscr{P}$ *is* a parabola, then its focus and directrix are uniquely determined.

If we where to draw a parabola, the complete figure, which includes both the focus and the directrix, should look as shown in the figure below.

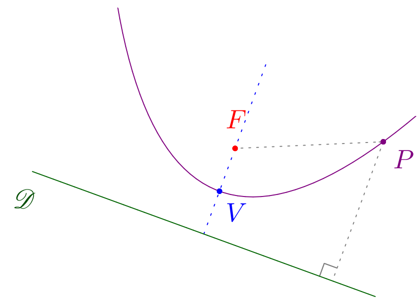


Figure 6.1.1

In the figure above P marks a typical point on the parabola. Among all points on the parabola, there is distinguished one, called the **vertex**, which is *the point on the parabola, which is closest to the directix*. In Figure 6.1.1 above, the vertex is marked as V. By definition, *the vertex is half-way between the focus and the directrix*. The *line that passes through both the focus and the vertex*, which is also the *line that passes through the focus and is perpendicular to the directrix*, is referred to as the **focal axis** of the parabola. It is more-or-less obvious that *the focal axis is a line of symmetry for the parabola*.

(Optional) Tangent Lines and the Reflective Properties

Using the geometric definition, we can neatly describe the *tangent line* to a parabola, which goes as follows.

Suppose we have a parabola $\mathscr{P}$ with focus F and directrix $\mathscr{D}$, and we fix some point P on the parabola. By definition, if we take Q to be the projection of P on $\mathscr{D}$, we know that

$$\text{dist}(P, F) = \text{dist}(P, \mathscr{D}) = \text{dist}(P, Q). \tag{6.1.2}$$

Tangent Line Theorem

With the set-up as above, the perpendicular bisector of $\overline{FQ}$ is tangent to the parabola $\mathscr{P}$ at the point P.

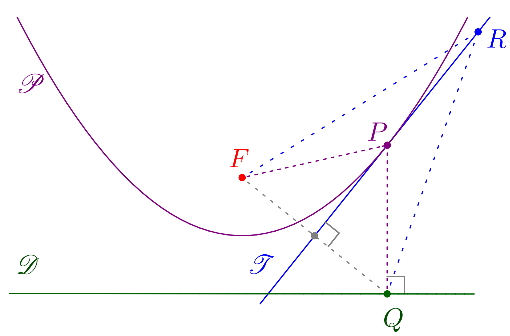


Figure 6.1.2

Proof. Let $\mathscr{T}$ denote perpendicular bisector of $\overline{FQ}$. First of all, since F does not sit on $\mathscr{D}$, it follows that *$\mathscr{T}$ is never perpendicular to $\mathscr{D}$*. Secondly, by (6.1.2), the point P clearly sits on $\mathscr{T}$, therefore $\mathscr{T}$ intersects $\mathscr{P}$ at least once: at the point P. All that remains to be proved is the fact that $\mathscr{T}$ *does not intersect* $\mathscr{P}$ *at any other point distinct from* P. Pick an arbitrary point R on $\mathscr{T}$, which is distinct from P. By construction (of the perpendicular bisector), we know that

$$\text{dist}(R, F) = \text{dist}(R, Q) \geq \text{dist}(R, \mathscr{D}). \tag{6.1.3}$$

Note that the last inequality in (6.1.3) is *strict*, unless Q *coincides with the projection of* R *on* $\mathscr{D}$. However, if this coincidence takes place, then it would force the line $\mathscr{L}$ to be perpendicular to $\mathscr{D}$, which is impossible. Since $\text{dist}(R, F) > \text{dist}(R, \mathscr{D})$, it follows that R cannot sit on $\mathscr{P}$, and the proof is complete. □

CLARIFICATION. With the same set-up as above, the three angles marked β, β' and α, formed by the tangent line $\mathscr{T}$ with the perpendicular line to $\mathscr{D}$ passing through P, and with the line passing through F and P, are all congruent.

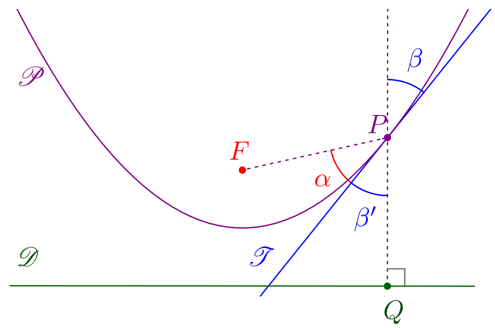


Figure 6.1.3

The congruence between β and β' is obvious, because they represent opposite angles. The fact that the angles marked β' and α are congruent, follows from the Tangent Line Theorem, which implies that $\mathscr{T}$ is the angle bisector of the angle $\angle FPQ$.

One way to interpret the congruence between α and β is to state it as a pair of *reflective* properties.

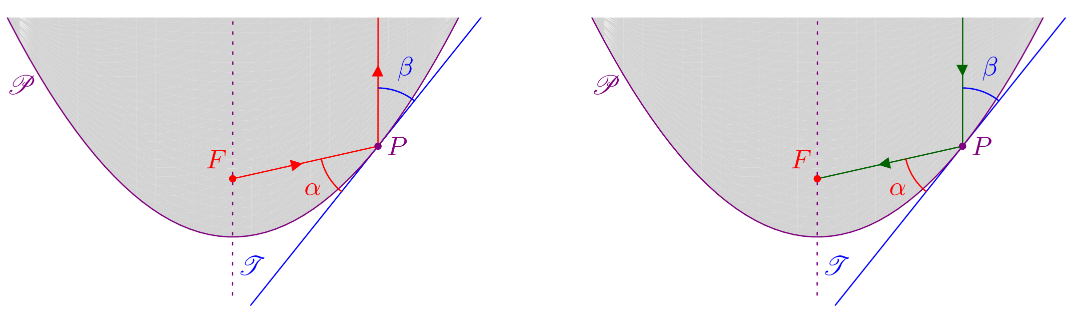


Figure 6.1.4

Reflective Properties of Parabolas

If we think a parabola as a (reflective) wall, then it has the following features.

I. Once a *ray that emanates from the focus* hits the parabola, its reflection is a *ray inside the parabola, which is parallel to the focal axis*.

II. Once a *ray inside the parabola, which is parallel to the focal axis* hits the parabola, its reflection is a *ray that passes through the focus*.

The reflective properties are used in many applications. Property I is used for designing automobile headlights, sound speakers, and concert halls. Property II is used for designing satellite antennas.

Parabolas in Standard Position and their Equations

We say that a parabola $\mathscr{P}$ is in *standard position*, if either

(A) $\mathscr{P}$ has *horizontal directrix* – which is the same as saying that $\mathscr{P}$ has a *vertical focal axis*, or

(B) $\mathscr{P}$ has *vertical directrix* – which is the same as saying that $\mathscr{P}$ has a *horizontal focal axis*.

By easy geometric considerations, it is pretty clear that *any parabola can be obtained by rotating a standard position parabola*. We will clarify this issue later, in Section 6.4, so for now we will only be concerned with parabolas that are in standard position. It is also clear the *any standard position parabola is obtained by translating a standard position parabola which has the origin as vertex*. We refer to this special type of parabolas as the "easy" parabolas, because their equations are not difficult to obtain.

Assume, for instance, we have an "easy" parabola, with *horizontal directrix*. In particular, it follows that the focal axis (which is perpendicular to the directrix) *coincides with the y-axis*, so the focus of such a parabola is of the form $F(0, p)$, for some real number $p \neq 0$. Also, since the directrix $\mathscr{D}$ is horizontal, and the vertex (in this special case, the origin) is *half-way between F and $\mathscr{D}$*, if follows that the directrix has equation $\mathscr{D}: \quad y = -p$. With this set-up, for any point $P(x, y)$, the distances to F and $\mathscr{D}$ are very easy to compute:

$$\begin{aligned} \text{dist}(P, F) &= \sqrt{(x-0)^2 + (y-p)^2} = \sqrt{x^2 + (y-p)^2}; \\ \text{dist}(P, \mathscr{D}) &= \sqrt{(x-x)^2 + (y-(-p))^2} = \sqrt{(y+p)^2} = |y+p|. \end{aligned}$$

(The second formula follows from the fact that, for any point $P(x, y)$, its projection Q on $\mathscr{D}$ – which helps us compute $\text{dist}(P, \mathscr{D}) = \text{dist}(P, Q)$, has coordinates $Q(x, -p)$.) So now the geometric definition of the parabola (with has focus F and directrix $\mathscr{D}$) will make its equation be:

$$\sqrt{x^2 + (y-p)^2} = |y+p|,$$

which by taking squares is the same as:

$$x^2 + (y-p)^2 = (y+p)^2.$$

Of course, by computing the squares on both sides, the above equation is same as:

$$x^2 + y^2 - 2py + p^2 = y^2 + 2py + p^2,$$

which is equivalent to:

$$x^2 = 4py. \tag{6.1.4}$$

A similar calculation can be done in the case of an "easy" parabola with *vertical directrix*, and the equation will have the form:

$$y^2 = 4px, \tag{6.1.5}$$

where now the focus is $F(p, 0)$ and the directrix has equation $\mathscr{D} : \quad x = -p$.

As for the standard position parabolas, which have the vertex at an arbitrary location, say $V(x_V, y_V)$, the equations will be either (6.1.4) or (6.1.5), with

$(\star)$ x *replaced by* $x - x_V$, and y *replaced by* $y - y_V$.

These findings are summarized as follows.

Equations of Parabolas in Standard Position

Assume a parabola $\mathscr{P}$ is in *standard position*, and has vertex $V(x_V, y_V)$ and focus $F(x_F, y_F)$. Then, depending on the directrix (or focal axis) orientation, the equation of $\mathscr{P}$ has one of the following forms.

(A) If the parabola $\mathscr{P}$ has *horizontal directrix* $\mathscr{D} : \quad y = y_{\mathscr{D}}$ (or equivalently, *vertical focal axis*), then $\mathscr{P}$ can be presented by an equation of the form

$$(x - x_V)^2 = 4p(y - y_V), \tag{6.1.6}$$

where:
- the focus has coordinates $x_F = x_V$ and $y_F = y_V + p$;
- the significant coordinate of the directrix is $y_{\mathscr{D}} = y_V - p$.

(B) If the parabola $\mathscr{P}$ has *vertical directrix* $\mathscr{D} : \quad x = x_{\mathscr{D}}$ (or equivalently, *horizontal focal axis*), then $\mathscr{P}$ can be presented by an equation of the form

$$(y - y_V)^2 = 4p(x - x_V), \tag{6.1.7}$$

where:
- the focus has coordinates $y_F = y_V$ and $x_F = x_V + p$;
- the significant coordinate of the directrix is $x_{\mathscr{D}} = x_V - p$.

The coefficient p that appears in the above equations will be called the **focal parameter**. Its absolute value is of course

$$|p| = \text{dist}(V, F) = \text{dist}(V, \mathscr{D}) = \tfrac{1}{2}\text{dist}(F, \mathscr{D})$$

is what we call the **focal distance**.

Finding Equations of Parabolas from Geometric Data

Based on the information we have concerning the form of the equations of a standard position parabola, we can find these equations solely based on geometric data. In other words, we are able to solve the following type of problem.

Geometric-to-Analytic Problem

Given (enough) geometric information about a parabola, find its equation.

In all instances of this problem, we will be given *two* of the following elements: vertex, directrix, focus. As for the methodology of solving the above Geometric-to-Analytic Problem, the key steps are:

I. Determine the *orientation*, that is, figure out if the directrix (or focal axis) is either horizontal or vertical. If the *directrix* is already given, this is easy to figure out. If the vertex and focus are given, then we can clearly determine to orientation of the *focal axis*.

II. Find the focal parameter p, which is needed in either form of the equation. The easy way to figure this out is to think of p as having to do with the *displacements* needed to travel from $\mathscr{D}$ to V, and from V to F. Both of these moves take place along the *focal axis*, so only one coordinate will change:
- If the focal axis is *horizontal*, only the x-coordinate will change, thus

$$p = x_V - x_{\mathscr{D}} = x_F - x_V = \tfrac{1}{2}\left(x_F - x_{\mathscr{D}}\right).$$

- If the focal axis is *vertical*, only the y-coordinate will change, thus

$$p = y_V - y_{\mathscr{D}} = y_F - y_V = \tfrac{1}{2}\left(y_F - y_{\mathscr{D}}\right).$$

III. Compute the coordinates of the *vertex*. If the vertex is not given already, all we have to do here is to use the fact that *the vertex is half-way between the focus and the directrix*.

Example 6.1.1. Suppose we want to find the equation of a parabola with directrix $\mathscr{D}$: $x = 2$ and vertex $V(4, 1)$.

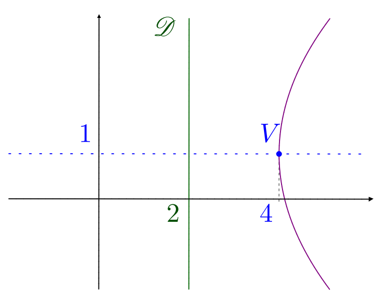


Figure 6.1.5

Since the directrix $\mathscr{D}$ is clearly vertical, the equation will have to match (6.1.7). Also, since we are already given the vertex, we know that $x_V = 4$ and $y_V = 1$, so if we match this information with (6.1.7), our provisional equation is:

$$(y - 1)^2 = 4p(x - 4). \tag{6.1.8}$$

As for the focal parameter p, we simply use the coordinates that change on the focal axis, when we move from $\mathscr{D}$ to x_V, so:

$$p = x_V - x_{\mathscr{D}} = 4 - 2 = 2,$$

so when we go back to (6.1.1), our equation becomes $(y-1)^2 = 4 \cdot 2(x-4)$, which is the same as:

$$(y-1)^2 = 8(x-4).$$

Example 6.1.2. Suppose we want to find the equation of a parabola with directrix $\mathscr{D}$: $y=1$ and focus $F(1,0)$.

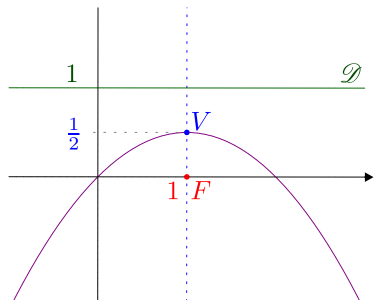


Figure 6.1.6

Since the directrix $\mathscr{D}$ is clearly horizontal, the equation will have to match (6.1.6). Clearly, as the vertex is half-way between the focus $V(1,0)$ and the directrix $\mathscr{D}$: $y=1$, it is pretty clear that

- the vertex is $V\left(1, \dfrac{1}{2}\right)$, and
- $p = y_V - y_{\mathscr{D}} = \dfrac{1}{2} - 1 = -\dfrac{1}{2}$.

When we plug all this information in (6.1.6), our equation becomes $(x-1)^2 = 4\left(-\dfrac{1}{2}\right)\left(y-\dfrac{1}{2}\right)$, which is the same as:

$$(x-1)^2 = -2y+1.$$

From Equations to Geometry

The Geometric-to-Analytic Problem discussed earlier can be reversed, so we can also consider the following.

Analytic-to-Geometric Problem

Given the equation of a parabola, obtain all geometric information about the parabola.

By "playing" a little bit with the equations of standard position parabolas, it is pretty clear that each one of these equations can be presented in one of the following specialized forms of the general quadratic equation (6.0.1):

$$Ax^2 + Dx + Ey + F = 0, \quad \text{with } A, E \neq 0$$

or

$$Cy^2 + Dx + Ey + F = 0, \quad \text{with } C, D \neq 0$$

Upon dividing by the coefficient to the variable that is not squared, it follows that these equations can be reduced to nicer ones, which are of one of the following two forms:

$$y = ax^2 + bx + c \tag{6.1.9}$$

or

$$x = ay^2 + by + c, \tag{6.1.10}$$

with $a \neq 0$. Of course, by *square completion*, the equation (6.1.9) can be rewritten as $y = a(x - h)^2 + k$, which will match (6.1.6). Likewise, (6.1.10) can be rewritten as $x = a(y - k)^2 + h$, which will match (6.1.7). These findings are summarized as follows.

Parabolas Given by Function-Like Equations

I. An equation of the form

$$y = ax^2 + bx + c \tag{6.1.11}$$

always represents a a parabola with *vertical focal axis* – equivalently, with *horizontal directrix*. Moreover, if after completing the squares, we re-write this equation in the form

$$y = a(x - h)^2 + k, \tag{6.1.12}$$

then the coordinates of the *vertex* are precisely $x_V = h$ and $y_V = k$.

II. An equation of the form

$$x = ay^2 + by + c \tag{6.1.13}$$

always represents a a parabola with *horizontal focal axis* – equivalently, with *vertical directrix*. Moreover, if after completing the squares, we re-write this equation in the form

$$x = a(y - k)^2 + h, \tag{6.1.14}$$

then again, the coordinates of the *vertex* are precisely $x_V = h$ and $y_V = k$.

In either case, the focal parameter p and the coefficient a (hereafter referred to as the **shape parameter**) are related by the identities

$$a = \frac{1}{4p}; \qquad p = \frac{1}{4a}. \tag{6.1.15}$$

Example 6.1.3. Consider the equation

$$x = 2y^2 + 16y - 4$$

and let us find all geometric information of the parabola it represents: focus, vertex and directrix.

We start off by completing squares in the right-hand side. When we concentrate on the terms that contain y, using the easy square completion identity we can write

$$2y^2 + 16y = 2(y + 4)^2 - 32,$$

so if we go back to our original equation, the right-hand side can be written as $2(y+4)^2 - 32 - 4 = 2(y + 4)^2 - 36$, so now our equation is:

$$x = 2(y + 4)^2 - 36.$$

Of course, this equation matches case II above, so based on the above fact it follows that:

(i) Our parabola has *horizontal focal axis*.

(ii) The vertex of the parabola is $V(-36,-4)$. (Be careful here! The term $y+4$ matches $y-k$, so $k=-4$. The term -36 matches h, so $h=-36$.)

(iii) Using (6.1.15) the focal coefficient is

$$p=\frac{1}{4a}=\frac{1}{4\cdot 2}=\frac{1}{8}.$$

(iv) Since the focal axis is horizontal, the coordinates of the focus satisfy $x_F=x_V+p=-36+\frac{1}{8}=-\frac{287}{8}$, and $y_F=y_V=-4$. In other words, the focus is: $F\left(-\frac{287}{8},-4\right)$.

(v) The directrix (which is vertical) has significant coordinate $x_{\mathscr{D}}=x_V-p=-36-\frac{1}{8}=-\frac{289}{8}$, so the directrix has equation

$$\mathscr{D}:\ x=-\frac{289}{8}.$$

Exercises

In Exercises 1-13 you are asked to find the equation of the parabola based on the given information.

1. Vertex $V(2,2)$; focus $F(-4,2)$.
2. Vertex $V(-2,2)$; focus $F(-4,2)$.
3. Vertex $V(1,-1)$; focus $F(1,7)$.
4. Vertex $V(-2,-3)$; focus $F(-2,-5)$.
5. Focus $F(2,3)$; directrix: $y=1$.
6. Focus $F(2,-2)$; directrix: $y=1$.
7. Focus $F(1,1)$; directrix: $x=-3$.
8. Focus $F(-5,2)$; directrix: $x=1$.
9. Vertex $V(-4,3)$; directrix: $y=4$.
10. Vertex $V(6,1)$; directrix: $x=1$.
11. Vertex $V(2,-3)$; directrix: $x=5$.
12. Vertex $V(2,-3)$; vertical focal axis; parabola passes through $P(4,4)$.
13. Vertex $V(2,-3)$; horizontal focal axis; parabola passes through $P(4,4)$.

In Exercises 14-19 you are given the equation of a parabola, and are asked to

(i) indicate focal axis orientation;

(ii) find the vertex;
(iii) find the focus;
(iv) find the directrix.

14. $y = x^2 - 6x + 5$.

15. $y = -\frac{1}{2}x^2 + 3x + 1$.

16. $y = -\frac{1}{4}x^2 + 6x - 2$.

17. $x = \frac{1}{4}y^2 - y$.

18. $x = -\frac{3}{2}y^2 + 3y - 1$.

19. $x = -y^2 + 2y$.

6.2 Ellipses

Ellipses are certain sets of points which, in many aspects, resemble circles.

Geometric Definition of Ellipses

As a *geometric locus*, an ellipse is defined as follows.

Given *two distinct points* F, F' and some real number $\Delta > \text{dist}(F, F')$, the **ellipse with foci F, F' and width Δ** is the *set of all points P in the plane, that satisfy the equality*

$$\text{dist}(P, F) + \text{dist}(P, F') = \Delta. \tag{6.2.1}$$

CLARIFICATIONS. If we where to draw an ellipse, we could devise the following method. Attach the ends of a string of length Δ to the two points F and F', then using a pencil we stretch the string to be completely tight, and we move the pencil by keeping the string tight at all times.

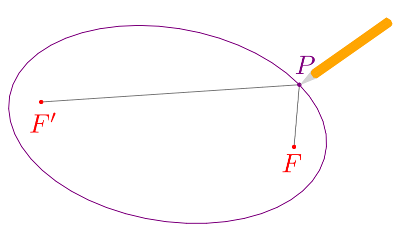


Figure 6.2.1

Besides the foci, a complete depiction of an ellipse should include several other important points, which help us determine all its geometric features. Such a complete picture is shown in Figure 6.2.2 below, which includes several lines and points, which are determined from the *symmetries* of the ellipse, as follows.

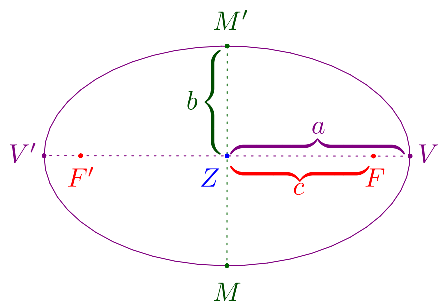


Figure 6.2.2

(i) The *line that passes through the foci F, F' of the ellipse* is a *line of symmetry* for the ellipse, which we will refer to as the *major symmetry line*. The points where *the major symmetry line intersects the ellipse* are called the **vertices** (or the **major points**) of the ellipse. In Figure 6.2.2, the vertices are marked V and V'. The line segment $\overline{VV'}$ formed by the vertices is called the **major axis** (or the **focal axis**) of the ellipse.

(ii) The *the perpendicular bisector of the major axis* is another *line of symmetry* for the ellipse, which we will refer to as the *minor symmetry line*. The points where *the minor symmetry line intersects the ellipse* are called the **minor points**) of the ellipse. In Figure 6.2.2, the minor points are marked M and M'. The line segment $\overline{MM'}$ formed by the minor points is called the **minor axis** of the ellipse.

(iii) The *point where the major and minor axes intersect* is a *center of symmetry* for the ellipse, which we will refer to as the **center** of the ellipse. In Figure 6.2.2, the center is marked Z. It is pretty clear (by symmetry) that *the center of the ellipse is the midpoint of each one of the axes $\overline{VV'}$, $\overline{MM'}$, as well as the midpoint of the segment $\overline{FF'}$ determined by the foci*.

Once all (or some of the) points Z, F, F', V, V', M, M' are available, we can measure the following **shape parameters**.

- the **major radius**: $a = \text{dist}(Z, V) = \text{dist}(Z, V') = \frac{1}{2}\text{dist}(V, V')$;
- the **minor radius**: $b = \text{dist}(Z, M) = \text{dist}(Z, M') = \frac{1}{2}\text{dist}(M, M')$;
- the **focal distance**: $c = \text{dist}(Z, F) = \text{dist}(Z, F') = \frac{1}{2}\text{dist}(F, F')$.

All geometric features of an ellipse are encoded in the shape parameters, as indicated in the following formula package.

Shape Parameter Formulas for Ellipses

If $\mathscr{E}$ is an ellipse with shape parameters a (major radius), b (minor radius) and c (focal distance), then the following identities hold.

I. **The Width Formula:** *the width Δ of the ellipse is equal to the length of the major axis*, that is, $\Delta = 2a$;

II. **The Focal Distance Formula:** $c^2 = a^2 - b^2$.

CLARIFICATION. To obtain the Width Formula, all we have to use is the fact that

$$\Delta = \text{dist}(P, F) + \text{dist}(P, F'), \tag{6.2.2}$$

for all points P on the ellipse. If we specialize (6.2.2) to a vertex, say V, it follows that

$$\Delta = \text{dist}(V, F) + \text{dist}(V, F'),$$

and everything is pretty clear (see Figure 6.2.2), because we clearly have $\text{dist}(V, F) = a - c$ and $\text{dist}(V, F') = a + c$.

To obtain the Focal Distance Formula, we specialize (6.2.2) to a minor point, say M, we obtain[15]

$$2a = \Delta = \text{dist}(M, F) + \text{dist}(M, F') = 2\,\text{dist}(M, F),$$

thus $\text{dist}(M, F) = a$. The desired formula now follows immediately from Pythagoras' Theorem applied to the right triangle $\triangle ZMF$ (with hypotenuse $MF = a$ and legs $MZ = b$ and $ZF = c$), which yields: $a^2 = b^2 + c^2$.

It is pretty clear that the shape parameters completely determine all geometric features of an ellipse. In particular, if two ellipses have *identical shape parameters*, then they are *congruent*. (Of course, by the Focal Distance Formula, it suffices to match only two of the shape parameters.)

Another way to look at an ellipse is to look at various ratios between its shape parameters. Again, by the Focal Distance Formula, one ratio completely determines all other ratios, as shown in formulas (6.2.6) below. The preferred ratio that one uses is:

$$\boldsymbol{e} = \frac{\textit{focal distance}}{\textit{major radius}} = \frac{c}{a}, \tag{6.2.3}$$

which is referred to as the **eccentricity** of the ellipse. It is pretty clear that the eccentricity of an ellipse always satisfies the inequality: $0 < \boldsymbol{e} < 1$. Also, straight from the definition, we have

$$c = a\boldsymbol{e}, \tag{6.2.4}$$

so when we replace this in the Focal Distance Formula, we get $a^2\boldsymbol{e}^2 = a^2 - b^2$, which immediately yields

$$b = a\sqrt{1 - \boldsymbol{e}^2}. \tag{6.2.5}$$

In particular, the eccentricity determines the other two ratios involving the shape parameters:

$$\frac{b}{a} = \sqrt{1 - \boldsymbol{e}^2}; \quad \frac{c}{b} = \frac{\boldsymbol{e}}{\sqrt{1 - \boldsymbol{e}^2}}. \tag{6.2.6}$$

Using these calculations, it is pretty clear that the eccentricity determines all geometric features of an ellipse, *up to a dilation*, so if two ellipses have *equal eccentricities*, then they are *similar*.

At the beginning of this section we mentioned that ellipses resemble *circles*. One can in fact think circles as *exceptional* ellipses, which are those with *zero eccentricity*, in which the two foci coincide(!) with the center. The shape parameters for such exceptional ellipses satisfy $a = b$ (= *radius of the circle*) and $c = 0$. The only trouble with circles, thought as exceptional ellipses, is that they do not have clearly defined major and minor axes, since every line passing through the center is a line of symmetry. Other than this awkward anomaly, it is safe to think circles as ellipses.

(Optional) Tangent Lines and the Reflective Properties

Using the geometric definition, we can neatly describe the *tangent line* to an ellipse, which goes as follows.

[15] The last equality is due to the fact that M sits on the perpendicular bisector of $\overline{FF'}$, thus $\text{dist}(M, F) = \text{dist}(M, F')$.

Suppose we have an ellipse $\mathscr{E}$ of width Δ, with foci F, F', and we fix some point P on the ellipse. Construct a point Q on the line passing through F and P, so that $FQ = \Delta$, and P lies on the segment $\overline{FQ}$, as shown in Figure 6.2.3 below.

Tangent Line Theorem

With the set-up as above, the perpendicular bisector of $\overline{F'Q}$ is tangent to the ellipse $\mathscr{E}$ at the point P.

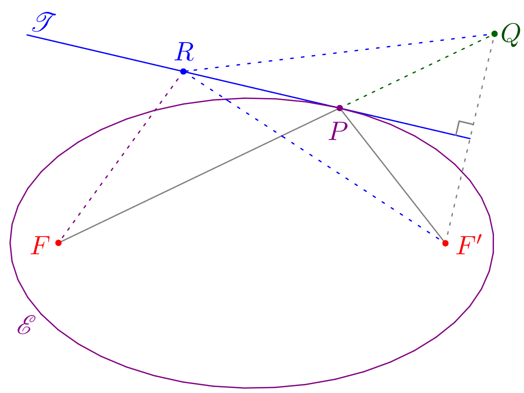


Figure 6.2.3

Proof. Let $\mathscr{T}$ denote perpendicular bisector of $\overline{F'Q}$. Since by construction we know that

$$\text{dist}(P, F) + \text{dist}(P, Q) = \text{dist}(F, Q) = \Delta, \tag{6.2.7}$$

using the definition of the ellipse as in (6.2.1), it follows that

$$\text{dist}(P, F') = \text{dist}(P, Q). \tag{6.2.8}$$

so the point P clearly sits on $\mathscr{T}$, therefore $\mathscr{T}$ intersects $\mathscr{E}$ at least once: at the point P. All that remains to be proved is the fact that $\mathscr{T}$ *does not intersect $\mathscr{E}$ at any other point distinct from P*. Pick an arbitrary point R on $\mathscr{T}$, which is distinct from P. On the one hand, by construction (of the perpendicular bisector), we know that

$$\text{dist}(R, Q) = \text{dist}(R, F'). \tag{6.2.9}$$

On the other hand, by the triangle inequality, applied to $\triangle\, RFQ$ it follows that

$$\text{dist}(R, F) + \text{dist}(R, Q) > \text{dist}(F, Q). \tag{6.2.10}$$

Using (6.2.9) and (6.2.7), the above inequality simply reads

$$\text{dist}(R, F) + \text{dist}(R, F') > \Delta, \tag{6.2.11}$$

which clearly shows that R cannot sit on $\mathscr{E}$, thus completing the proof. $\square$

CLARIFICATION. With the same set-up as above, the three angles marked β, β' and α, formed by the tangent line $\mathscr{T}$ with the segments $\overline{PF}$, $\overline{PQ}$ and $\overline{PF'}$, are all congruent.

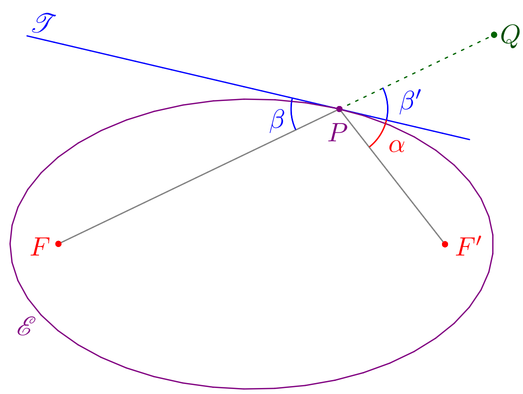


Figure 6.2.4

The congruence between β and β' is obvious, because they represent opposite angles. The fact that the angles marked β' and α are congruent, follows from the Tangent Line Theorem, which implies that $\mathscr{T}$ is the angle bisector of the angle $\angle F'PQ$.

One way to interpret the congruence between α and β is to state it as a *reflective* property.

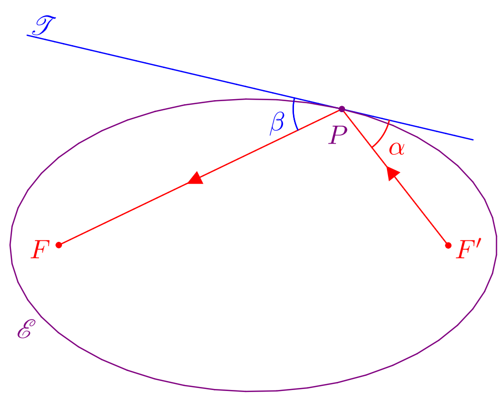


Figure 6.2.5

Reflective Property of Ellipses

If we think an ellipse as the (reflective) wall of a chamber, then, once a *ray that emanates from one focus* hits the ellipse, its reflection is a *ray that passes through the other focus*.

The reflective property of ellipses has many applications in ultrasound therapy. If the ultrasound source is placed at a focus inside an elliptical chamber, then all ultrasound waves, which bounce off the chamber wall, will concentrate at the other focus, where the particular spot that needs treatment is placed. The same principle is used for radiation treatment of cancer tumors.

Ellipses in Standard Position and their Equations

We say that an ellipse $\mathscr{E}$ is in *standard position*, if either

(A) $\mathscr{E}$ has *horizontal major axis* – which is the same as saying that $\mathscr{E}$ has a *vertical minor axis*, or

(B) $\mathscr{E}$ has *vertical major axis* – which is the same as saying that $\mathscr{E}$ has a *horizontal minor axis*.

By easy geometric considerations, it is pretty clear that *any ellipse can be obtained by rotating a standard position ellipse*. We will clarify this issue later, in Section 6.4, so for now we will only be concerned with ellipses that are in standard position. It is also clear the *any standard position ellipse is obtained by translating a standard position ellipse which has its center at the origin*. We refer to this special type of ellipses as the "easy" ellipses, because their equations are not difficult to obtain.

Assume, for instance, we have an "easy" ellipse, with *horizontal major axis*. In particular, it follows that *the major symmetry line is the x-axis*, so the foci will have coordinates $F(c,0)$ and $F'(-c,0)$, while the vertices will have coordinates $V(a,0)$ and $V'(-a,0)$. With this set-up, for any point $P(x,y)$, the distances to F and F' are:

$$\begin{aligned} \text{dist}(P,F) &= \sqrt{(x-c)^2+(y-0)^2} = \sqrt{x^2+c^2-2cx+y^2}; \\ \text{dist}(P,F') &= \sqrt{(x-(-c))^2+(y-0)^2} = \sqrt{x^2+c^2+2cx+y^2}. \end{aligned}$$

So now the geometric definition of the ellipse reads:

$$\sqrt{x^2+2cx+c^2+y^2}+\sqrt{x^2-2cx+c^2+y^2}=2a. \tag{6.2.12}$$

When we subtract the second term from both sides, we get:

$$\sqrt{x^2+2cx+c^2+y^2}=2a-\sqrt{x^2-2cx+c^2+y^2}, \tag{6.2.13}$$

so when we square both sides, we obtain

$$\begin{aligned} x^2+2cx+c^2+y^2 &= \left(2a-\sqrt{x^2-2cx+c^2+y^2}\right)^2 = \\ &= 4a^2-4a\sqrt{x^2-2cx+c^2+y^2}+x^2-2cx+c^2+y^2. \end{aligned}$$

Upon subtracting the left-hand side from the right-hand side and adding $4a\sqrt{x^2-2cx+c^2+y^2}$ to both sides, we obtain:

$$4a\sqrt{x^2-2cx+c^2+y^2} = \ = 4a^2+x^2-2cx+c^2+y^2-(x^2+2cx+c^2+y^2)=4a^2-4cx,$$

which by dividing everything by 4 yields:

$$a\sqrt{x^2-2cx+c^2+y^2} = \ = a^2-cx. \tag{6.2.14}$$

By squaring both sides of (6.2.14), we get $a^2(x^2-2cx+c^2+y^2)=(a^2-cx)^2$, which reads:

$$a^2x^2-2a^2cx+a^2c^2+a^2y^2=a^4-2a^2cx+c^2x^2. \tag{6.2.15}$$

By moving all terms in from right- to left-hand side, except for a^4, and moving a^2c^2 from left- to right-hand side, the terms $2a^2cx$ cancel, so by grouping we now get

$$(a^2-c^2)x^2+a^2y^2=a^4-a^2c^2=a^2(a^2-c^2). \tag{6.2.16}$$

Using the Focal Distance Formula, we know that $a^2-c^2=b^2$, so the above equation simply reads

$$b^2x^2+a^2y^2=a^2b^2, \tag{6.2.17}$$

so upon dividing everything by the right-hand side and simplifying, we finally get:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{6.2.18}$$

In the process of obtaining the equation (6.2.18), we used the *squaring* method twice. Basically we converted from an equation of the form $\sqrt{\heartsuit} = \#$ to its squared version: $\heartsuit = \#^2$, so we need to be concerned with the possibility that we may have added *extraneous solutions*. Concerning this possibility, all we have to notice here are the following two easy facts:

(i) For any fixed value of x, the equation (6.2.17), which can be re-written as

$$y^2 = \frac{a^2b^2 - b^2x^2}{a^2} = \left(\frac{b}{a}\right)^2 (a^2 - x^2),$$

has *at most two solutions*, namely $y = \pm\frac{b}{a}\sqrt{a^2 - x^2}$. Clearly solutions exist, only if $a^2 \geq x^2$, which is the same as $|x| \leq a$. Specifically, the number of solutions y of (6.2.17) is: (A) *two*, if $|x| < a$; (B) *one*, if $|x| = a$; (C) *none*, if $|x| > a$.

(ii) For any fixed value of x, there are again at most two values of y that give rise to a point $P(x, y)$ on the ellipse, and the number of possible points matches exactly the cases (A), (B), (C) above.

Since the process of deriving (6.2.18) from the original equation (6.2.12) does not increase the number of solutions y (for any given x), these two equations are indeed equivalent.

A similar calculation can be obtained for an "easy" ellipse, with *vertical major axis*, and the equation will have the form:

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1. \tag{6.2.19}$$

where now the foci will be $F(0, c)$ and $F'(0, -c)$, while the vertices will have coordinates $V(0, a)$ and $V'(0, -a)$.

As for standard position ellipses, which have the center at an arbitrary location, say $Z(x_Z, y_Z)$, the equations will be either (6.2.18) or (6.2.19), with

($\star$) x *replaced by* $x - x_Z$, and y *replaced by* $y - y_Z$.

These findings are summarized as follows.

Equations of Ellipses in Standard Position

Assume an ellipse $\mathscr{E}$ is in *standard position*, with center $Z(x_Z, y_Z)$, major radius a, minor radius b and focal distance c. Then, depending on the major (or minor) axis orientation, the equation of $\mathscr{E}$ has one of the following forms.

(A) If the ellipse $\mathscr{E}$ has *horizontal major axis* (or equivalently, *vertical minor axis*), then $\mathscr{E}$ can be presented by an equation of the form

$$\frac{(x - x_Z)^2}{a^2} + \frac{(y - y_Z)^2}{b^2} = 1, \tag{6.2.20}$$

and the ellipse has:
- vertices located at $(x_Z \pm a, y_Z)$;
- foci located at $(x_Z \pm c, y_Z)$;
- minor points located at $(x_Z, y_Z \pm b)$.

(B) If the ellipse $\mathscr{E}$ has *vertical major axis* (or equivalently, *horizontal minor axis*), then $\mathscr{E}$ can be presented by an equation of the form

$$\frac{(x - x_Z)^2}{b^2} + \frac{(y - y_Z)^2}{a^2} = 1, \tag{6.2.21}$$

and the ellipse has:
- vertices located at $(x_Z, y_Z \pm a)$;
- foci located at $(x_Z, y_Z \pm c)$;
- minor points located at $(x_Z \pm b, y_Z)$.

Finding Equations of Ellipses from Geometric Data

Based on the information we have concerning the form of the equations of a standard position ellipse, we can find these equations solely based on geometric data. In other words, we are able to solve the following type of problem.

Geometric-to-Analytic Problem

Given (enough) geometric information about an ellipse, find its equation.

Unlike what we have seen in Section 6.1, this problem has many possible instances. As for the methodology of solving the above Geometric-to-Analytic Problem, the key steps are:

I. Determine the *orientation*, that is, figure out if the given information allows you to find the orientation of one of the axes. Using this information, start building the template for the equation, which is either (6.2.20) or (6.2.21).

II. Locate the *center*. Remember that the center is the common midpoint of each of the axes, as well as the midpoint of the segment determined by the foci.

III. Compute some *shape parameters*, based on the given information. Remember that each one of the parameters a, b, c measures either a distance between two distinguished points, or half of a length of a distinguished segment:
- The major radius a is equal to the *distance from the center to either one of the vertices*; equivalently a is equal to *half the distance between the vertices*. The major radius a is also equal to the *distance between any focus and any minor point*.
- The minor radius b is equal to the *distance from the center to either one of the minor points*; equivalently b is equal to *half the distance between the minor points*.
- The focal distance c is equal to the *distance from the center to either one of the foci*; equivalently c is equal to *half the distance between the foci*.

Since for building the equation, we only need a^2 and b^2, in case one of them is missing, but we are able to find either the focal distance c, or the eccentricity $\boldsymbol{e}$, we can use the Focal Distance Formula, or the eccentricity formulas:

- $a^2 = b^2 + c^2 = \dfrac{c^2}{\boldsymbol{e}^2} = \dfrac{b^2}{1-\boldsymbol{e}^2}$;
- $b^2 = a^2 - c^2 = a^2(1-\boldsymbol{e}^2) = \dfrac{c^2(1-\boldsymbol{e}^2)}{\boldsymbol{e}^2}$.

Example 6.2.1. Suppose we want to find the equation of an ellipse with vertices $V(-2,1)$ and $V'(-2,5)$, and one focus at $F(-2,2)$.

Since the vertices, have equal x-coordinates, they lie on the vertical line $x=-2$. In particular, our ellipse will have *vertical major axis*, so its equation looks like

$$\frac{(x-x_Z)^2}{b^2} + \frac{(y-y_Z)^2}{a^2} = 1, \tag{6.2.22}$$

Since the center Z is the midpoint of the segment determined by the vertices, it will also lie on the the vertical line $x=-2$, so the x-coordinate of the center will be $x_Z = -2$. As for the y-coordinate of the center, it will be equal to the average of the y-coordinates of the vertices, so:

$$y_Z = \tfrac{1}{2}\big(1+5\big) = \tfrac{1}{2}\cdot 6 = 3.$$

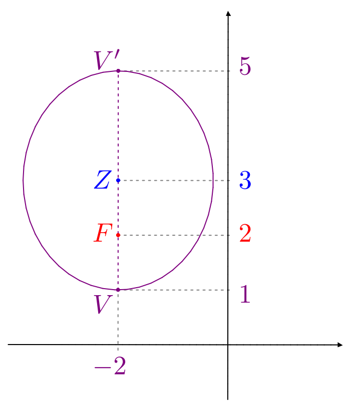


Figure 6.2.6

Using the center $Z(-2,3)$ and of one vertex, say $V(-2,5)$ it is clear that the major radius is:

$$a = \text{dist}(Z,V) = 2.$$

Using the center $Z(-2,3)$ and the given focus $F(-2,2)$ it is clear that the focal distance is:

$$c = \text{dist}(Z,F) = 1.$$

Using the Focal Distance Formula, the minor radius b satisfies

$$b^2 = a^2 - c^2 = 2^2 - 1^2 = 4 - 1 = 3.$$

Now we have all ingredients needed to fill in the template (6.2.22), so our equation becomes:

$$\frac{(x+2)^2}{3} + \frac{(y-3)^2}{4} = 1. \tag{6.2.23}$$

(The numerator of the first fraction in (6.2.22) is $(x-x_Z)^2 = (x-(-2))^2 = (x+2)^2$; the numerator of the second fraction in (6.2.22) is $(y-y_Z)^2 = (y-3)^2$.)

From Equations to Geometry

The Geometric-to-Analytic Problem discussed earlier can be reversed, so we can also consider the following.

Analytic-to-Geometric Problem

Given the equation of an ellipse, obtain all geometric information about the ellipse.

By "playing" a little bit with the equations of standard position ellipses, it is pretty clear that these equations can be presented in the following special form of the general quadratic equation (6.0.1) which looks like:

$$Ax^2 + Cy^2 + Dx + Ey + F = 0, \quad \text{with } A \text{ and } C \textit{ non-zero numbers with same sign}. \tag{6.2.24}$$

An equation of this form is called a **standard elliptic equation**. Using *square completion*, any standard elliptic equation can be studied as follows

Standard Elliptic Equation Analysis

Any standard elliptic equation (6.2.24) can be transformed by square completion into and equation of the form

$$A(x-h)^2 + C(y-k)^2 = M, \quad \text{with } A \text{ and } C \textit{ non-zero numbers with same sign}. \tag{6.2.25}$$

Depending of the sign of the right-hand side M, the equation (6.2.25) represents one of the following curves.

(A) If $M = 0$, then (6.2.25) represents *one point:* (h, k).
(B) If $M \neq 0$ and has *opposite sign to A (and C)*, then (6.2.25) represents *the empty set: there is no point satisfying it*.
(C) If $M \neq 0$ and has *same sign as A (and C)*, then (6.2.25) represents either a *circle*, or an *ellipse in standard position*, with *center* $Z(h, k)$. More specifically:
  - (C_1) If $A = C$, then (6.2.25) represents a *circle* of radius $R = \sqrt{M/A}(= \sqrt{M/C})$.
  - (C_2) If $A \neq C$, then (6.2.25) represents an *ellipse in standard position*, with major radius $a = \max\left(\sqrt{M/A}, \sqrt{M/C}\right)$, and minor radius $b = \min\left(\sqrt{M/A}, \sqrt{M/C}\right)$. In particular,
    - if $M/A > M/C$, then the ellipse has *horizontal major axis*;
    - if $M/A < M/C$, then the ellipse has *vertical major axis*.

CLARIFICATIONS. Since A and C have same sign, it follows that the left-hand side of (6.2.25) will either be

(i) zero, and this happens only when $(x, y) = (h, k)$; or
(ii) a non-zero number with same sign as A (and C), when $(x, y) \neq (h, k)$.

Based on this observation, cases (A) and (B) in the above analysis follow immediately. As for the remaining case (C), all we have to observe is that, when $M \neq 0$, the equation (6.2.25) can be easily transformed (by dividing everything by M) into

$$\frac{(x-h)^2}{M/A} + \frac{(y-k)^2}{M/C} = 1, \tag{6.2.26}$$

which clearly matches the equation of an ellipse (or circle).

In practice, when we are asked to determine the geometry of a curve represented by an elliptic equation of the form (6.2.24), it is best if we carry on the following steps.

I. Use square completion to transform (6.2.24) into (6.2.25) , then briefly analyze the equation to decide if we are in one of the extreme cases (A) or (B).

II. Assuming we are in case (C), do a second transformation, to get to the form (6.2.26), and match it with the equation of a standard position ellipse: either (6.2.20) or (6.2.21).

Example 6.2.2. Consider the standard elliptic equation

$$9x^2 + 16y^2 - 18x + 64y - 503 = 0. \tag{6.2.27}$$

and let us find all geometric information of the curve it represents: center, foci, vertices and minor points.

We start off by completing squares in the left-hand side. When we concentrate on the terms that contain x, using the easy square completion identity we can write

$$9x^2 - 18x = 9(x-1)^2 - 9. \tag{6.2.28}$$

When we concentrate on the terms that contain y, using the easy square completion identity we can write

$$16y^2 + 64y = 16(y+2)^2 - 64. \tag{6.2.29}$$

When we use (6.2.28) and (6.2.29) back in (6.2.27), our equation becomes

$$9(x-1)^2 - 9 + 16(y+2)^2 - 64 - 503 = 0,$$

which after grouping the constant terms and "moving" the result on the right, becomes:

$$9(x-1)^2 + 16(y+2)^2 = 576. \tag{6.2.30}$$

Next we divide everything by 576 and we get

$$\frac{(x-1)^2}{576/9} + \frac{(y+2)^2}{576/16} = 1.$$

which after simplifications in the denominators becomes

$$\frac{(x-1)^2}{64} + \frac{(y+2)^2}{36} = 1. \tag{6.2.31}$$

Since the larger of the denominators (that is, 64) is the one under the x-term, our ellipse has *horizontal major axis*, thus the major radius is $a = \sqrt{64} = 8$, and the minor radius is $b = \sqrt{36} = 6$. Additionally, the focal distance is

$$c = \sqrt{a^2 - b^2} = \sqrt{64 - 36} = \sqrt{28} = 2\sqrt{7},$$

so the eccentricity of our ellipse is:

$$\boldsymbol{e} = \frac{c}{a} = \frac{2\sqrt{7}}{8} = \frac{\sqrt{7}}{4}.$$

Furthermore, the center of the ellipse is $Z(1,-2)$. (Be careful here! The term $x-1$ matches $x-h$, so $h=1$. The term $y+2$ matches $y-k$, so $k=-2$.)

Since our ellipse has horizontal major axis, its distinguished points are as follows

- The vertices are located at $(x_Z \pm a, y_Z) = (1 \pm 8, -2)$, so they are the points $V(9,-2)$ and $V'(-7,-2)$.
- The minor points are located at $(x_Z, y_Z \pm b) = (1, -2 \pm 6)$, so they are the points $M(1,4)$ and $M'(1,-8)$.
- The foci are located at $(x_Z \pm c, y_Z) = (1 \pm 2\sqrt{7}, -2)$, so they are the points $F(1+2\sqrt{7},-2)$ and $F'(1-2\sqrt{7},-2)$.

Graphing Ellipses

If we are asked to graph the curve represented by a standard elliptic equation, one option is to *complete squares*, thus rewriting our equation in the form (6.2.25), and then *solve for* y, which after some easy algebraic manipulations reduces to solving an equation of the form

$$(y-k)^2 = \textit{expression in } x, \tag{6.2.32}$$

which can then be easily solved by taking square roots $y - k = \pm\sqrt{\textit{expression in } x}$, to yield

$$y = k \pm \sqrt{\textit{expression in } x}. \tag{6.2.33}$$

Strictly speaking, in order to get to an equation like (6.2.32), our square completion should be done only on the terms that involve y. In fact, to obtain the solutions (6.2.33), we do not need square completion at all: we could use the Quadratic Formula directly, on the original equation! Once we get to (6.2.33), all we need to do is to graph *both* equations given by (6.2.33), one for "+" and one for "−," on a calculator.

Example 6.2.3. Let us revisit the standard elliptic equation

$$9x^2 + 16y^2 - 18x + 64y - 503 = 0. \tag{6.2.34}$$

which we analyzed in Example 6.2.2, and let us assume we are only asked to graph it.

We can set up our equation as a quadratic equation with y as the unknown, like

$$16y^2 + 64y + (9x^2 - 18x - 503) = 0,$$

and then using the Quadratic Formula, we can solve for y:

$$\begin{aligned} y &= \frac{-64 \pm \sqrt{64^2 - 4\cdot 16\cdot(9x^2-18x-503)}}{2\cdot 16} = \frac{-64 \pm \sqrt{64\left[64-(9x^2-18x-503)\right]}}{32} = \\ &= \frac{-64 \pm 8\sqrt{64 - 9x^2 + 18x + 503}}{32} = \frac{-64}{32} \pm \frac{8\sqrt{-9x^2+18x+567}}{32} = \\ &= -2 \pm \frac{\sqrt{-9x^2+18x+567}}{4}. \end{aligned}$$

Therefore, in order to graph our ellipse on a calculator, we need to graph both of the following equations

$$\begin{aligned} y &= -2 + \frac{\sqrt{-9x^2+18x+567}}{4}, \\ y &= -2 - \frac{\sqrt{-9x^2+18x+567}}{4}, \end{aligned}$$

and the result is depicted in Figure 6.2.7 below.

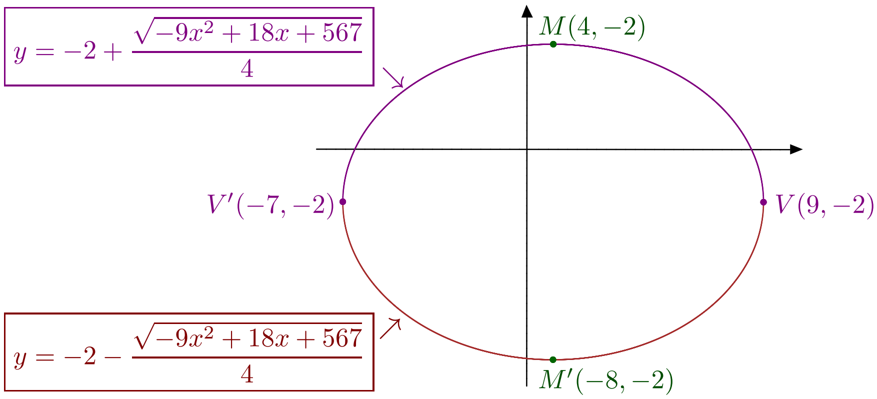


Figure 6.2.7

What happens if we do not have a calculator? We can sketch the graph of our ellipse *by hand*, using the complete analysis of our equation, as shown in Example 6.2.2. Using our findings concerning the geometry of our ellipse, we can sketch a *bounding box* for our ellipse, which is a rectangle that has the vertices and the minor points at the midpoints of the sides of the rectangle. A sketch using this bounding box is shown below:

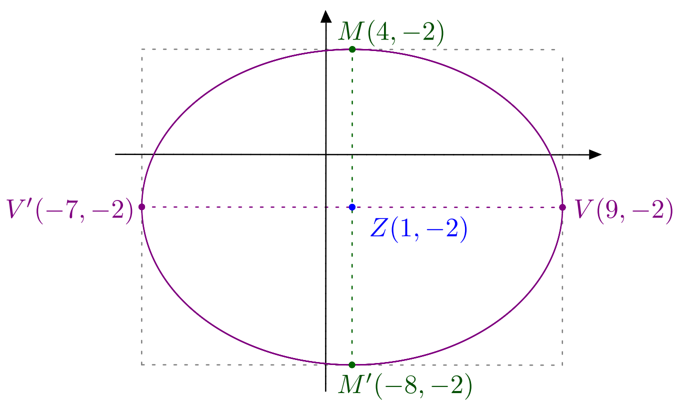


Figure 6.2.8

The software supplied in the K-STATE ONLINE HOMEWORK SYSTEM can only draw ellipses using bounding boxes. Therefore, in order to graph curves represented by elliptic equations, we must do a more-or-less complete analysis, by locating the center, at least one vertex, and at least one minor point.

Exercises

In Exercises 1-14 you are asked to find the equation of an ellipse, based on the given information.

1. Vertices $V(2, 2)$ and $V'(-4, 2)$; one focus at $F(1, 2)$.

2. Vertices $V(2, -3)$ and $V'(2, 7)$; one focus at $F(2, -1)$.

3. Vertices $V(2, 0)$ and $V'(2, 10)$; one minor point at $M(4, 5)$.

4. Vertices $V(-3,1)$ and $V'(5,1)$; one minor point at $M(2,2)$.

5. Foci $F(-2,3)$ and $F'(-2,9)$; one vertex at $V(-2,0)$.

6. Foci $F(0,0)$ and $F'(8,0)$; one vertex at $V(-4,0)$.

7. Minor points $M(-6,5)$ and $M'(4,5)$; one focus at $F(-1,2)$.

8. Minor points $M(-1,-1)$ and $M'(5,-1)$; one focus at $F(2,0)$.

9. Foci $F(-2,0)$ and $F'(-2,6)$; one minor point at $M(-4,3)$.

10. Foci $F(-2,0)$ and $F'(8,0)$; one minor point at $M(3,-1)$.

11*. Ellipse in standard position with one vertex $V(0,3)$ and one minor point $M(1,5)$.

12. Vertices $V(-5,1)$ and $V'(7,1)$; eccentricity $\frac{1}{3}$.

13. Foci $F(1,1)$ and $F'(9,1)$; eccentricity $\frac{1}{2}$.

14. Minor points $M(-7,1)$ and $M'(1,1)$; eccentricity $\frac{3}{5}$.

In Exercises 15-20 you are given a standard elliptic equation, and are asked to

(i) indicate the major axis orientation;
(ii) find the the center;
(iii) find the vertices;
(iv) find the foci;
(v) find the minor points;
(vi) sketch the graph.

15. $4x^2 + 9y^2 - 36 = 0$.

16. $16x^2 + y^2 = 144$.

17. $4(x+2)^2 + 9(y-3)^2 = 576$.

18. $4x^2 + y^2 + 8x - 2y - 11 = 0$.

19. $4x^2 + 9y^2 - 24x + 18y - 19 = 0$.

20. $4x^2 + 9y^2 + 32x + 36y + 64 = 0$.

6.3 Hyperbolas

Hyperbolas are certain sets of points which, unlike the previous two kinds (parabolas and ellipses), are *two-branch curves*. In some respects, they resemble a *pair of lines*.

Geometric Definition of Hyperbolas

As a *geometric locus*, a hyperbola is defined as follows.

Given *two distinct points* F, F' and some real number $\Delta < \text{dist}(F, F')$, the **hyperbola with foci F, F' and width Δ** is the *set of all points P in the plane, that satisfy the equality*

$$\text{dist}(P, F) - \text{dist}(P, F') = \pm\Delta. \tag{6.3.1}$$

CLARIFICATIONS. Besides the foci, a complete depiction of a hyperbola should include several other elements, which help us determine all its geometric features. Such a complete picture is shown in Figure 6.3.1 below, which includes several lines and points, which are determined from the *symmetries* of the hyperbola, as follows.

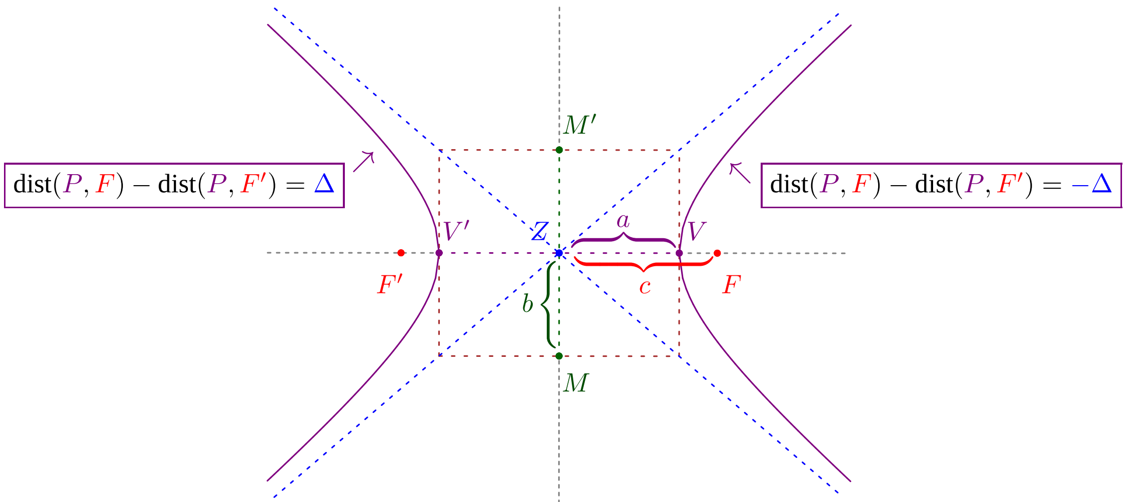


Figure 6.3.1

(i) The *line that passes through the foci F, F' of the hyperbola* is a *line of symmetry* for the hyperbola, which we will refer to as the *major symmetry line*. The points where *the major symmetry line intersects the hyperbola* are called the **vertices** (or the **major points**) of the hyperbola. In Figure 6.3.1, the vertices are marked V and V'. The line segment $\overline{VV'}$ formed by the vertices is called the **major axis** (or the **focal axis**) of the hyperbola.

(ii) The *the perpendicular bisector of the major axis* is another *line of symmetry* for the hyperbola, which we will refer to as the *minor symmetry line*.

(iii) The *point where the major and minor symmetry lines intersect* is a *center of symmetry* for the hyperbola, which we will refer to as the **center** of the hyperbola. In Figure 6.3.1, the center is marked Z.

(iv) (Far) away from the center, the hyperbola resembles *two lines that intersect at the center*. These two special lines (shown in Figure 6.3.1 in blue) are called the **asymptotes** of the hyperbola.

(v) If we consider the *tangent lines* at the vertices, they will intersect the asymptotes at four points, that form a rectangle, which we call the **bounding box** of the hyperbola, which is shown in brown in Figure 6.3.1. As it turns out, the following features are present:
- the *asymptotes* are the *diagonals of the bounding box*;
- the *vertices* are the *midpoints of two (opposite) sides of the bounding box*.

The midpoints of the other two sides of the bounding box (those that do not touch the hyperbola) are what we call the **minor points** of the hyperbola. These points are marked

M and M' in Figure 6.3.1. The line segment $\overline{MM'}$ formed by the minor points is called the **minor axis** of the hyperbola.

Once all (or some of the) points Z, F, F', V, V', M, M' are available, we can measure the following **shape parameters**.

- the **major radius**: $a = \text{dist}(Z, V) = \text{dist}(Z, V') = \frac{1}{2}\text{dist}(V, V')$;
- the **minor radius**: $b = \text{dist}(Z, M) = \text{dist}(Z, M') = \frac{1}{2}\text{dist}(M, M')$;
- the **focal distance**: $c = \text{dist}(Z, F) = \text{dist}(Z, F') = \frac{1}{2}\text{dist}(F, F')$.

All geometric features of a hyperbola are encoded in the shape parameters, as indicated in the following formula package.

Shape Parameter Formulas for Hyperbolas

If $\mathscr{H}$ is a hyperbola hyperbola with shape parameters a (major radius), b (minor radius) and c (focal distance), then the following identities hold.

I. **The Width Formula:** *the width Δ of the hyperbola is equal to the length of the major axis*, that is, $\Delta = 2a$;

II. **The Focal Distance Formula:** $c^2 = a^2 + b^2$.

CLARIFICATION. To obtain the Width Formula, all we have to use is the fact that

$$\text{dist}(P, F) - \text{dist}(P, F'), = \pm\Delta \tag{6.3.2}$$

for all points P on the hyperbola. If we specialize (6.3.2) to the vertex V', which is closest to F', which satisfies $\text{dist}(P, F) > \text{dist}(P, F')$, it follows that

$$\text{dist}(P, F) - \text{dist}(P, F'), = \Delta,$$

and everything is pretty clear (see Figure 6.3.1), because we clearly have $\text{dist}(V', F) = c + a$ and $\text{dist}(V', F') = c - a$.

The Focal Distance Formula is not easy to prove! This is due to the fact that the asymptotes and the bounding box are defined in a very complicated way. We will revisit this issue a little later, after we derive the equations of "easy" hyperbolas, which will also give us a proof for the Focal Distance Formula. .

Nevertheless, if we accept the Focal Distance Formula, it is pretty clear that the shape parameters completely determine all geometric features of a hyperbola. In particular, if two hyperbolas have *identical shape parameters*, then they are *congruent*. (Of course, by the Focal Distance Formula, it suffices to match only two of the shape parameters.)

Another way to look at a hyperbola is to look at various ratios between its shape parameters. Again, by the Focal Distance Formula, one ratio completely determines all other ratios, as shown in formulas (6.3.6) below. As was the case with hyperbolas, the preferred ratio that one uses is:

$$\boldsymbol{e} = \frac{\textit{focal distance}}{\textit{major radius}} = \frac{c}{a}, \tag{6.3.3}$$

which is referred to as the **eccentricity** of the hyperbola. It is pretty clear that the eccentricity of a hyperbola always satisfies the inequality: $\boldsymbol{e} > 1$. Also, straight from the definition, we have

$$c = a\boldsymbol{e}, \tag{6.3.4}$$

so when we replace this in the Focal Distance Formula, we get $a^2\boldsymbol{e}^2 = a^2 + b^2$, which immediately yields

$$b = a\sqrt{\boldsymbol{e}^2 - 1}. \tag{6.3.5}$$

In particular, the eccentricity determines the other two ratios involving the shape parameters:

$$\frac{b}{a} = \sqrt{\boldsymbol{e}^2 - 1}; \quad \frac{c}{b} = \frac{\boldsymbol{e}}{\sqrt{\boldsymbol{e}^2 - 1}}. \tag{6.3.6}$$

Using these calculations, it is pretty clear that the eccentricity determines all geometric features of a hyperbola, *up to a dilation*, so if two hyperbolas have *equal eccentricities*, then they are *similar*.

Hyperbolas in Standard Position and their Equations

We say that a hyperbola $\mathscr{H}$ is in *standard position*, if either

(A) $\mathscr{H}$ has *horizontal major axis* – which is the same as saying that $\mathscr{H}$ has a *vertical minor axis*, or

(B) $\mathscr{H}$ has *vertical major axis* – which is the same as saying that $\mathscr{H}$ has a *horizontal minor axis*.

By easy geometric considerations, it is pretty clear that *any hyperbola can be obtained by rotating a standard position hyperbola*. We will clarify this issue later, in Section 6.4, so for now we will only be concerned with hyperbolas that are in standard position. It is also clear the *any standard position hyperbola is obtained by translating a standard position hyperbola which has its center at the origin*. We refer to this special type of hyperbolas as the "easy" hyperbolas, because their equations are not difficult to obtain.

Assume, for instance, we have an "easy" hyperbola, with *vertical major axis*. In particular, it follows that *the major symmetry line is the y-axis*, so the foci will have coordinates $F(0, c)$ and $F'(0, -c)$, while the vertices will have coordinates $V(0, a)$ and $V'(0, -a)$. With this set-up, for any point $P(x, y)$, the distances to F and F' are:

$$\begin{aligned}
\text{dist}(P, F) &= \sqrt{(x-0)^2 + (y-c)^2} = \sqrt{x^2 + y^2 + c^2 - 2cy};\\
\text{dist}(P, F') &= \sqrt{(x-(0))^2 + (y-(-c))^2} = \sqrt{x^2 + y^2 + c^2 + 2cy}.
\end{aligned}$$

So now the geometric definition of the hyperbola reads:

$$\sqrt{x^2 + y^2 + 2cy + c^2} - \sqrt{x^2 + y^2 - 2cy + c^2} = \pm 2a. \tag{6.3.7}$$

When we add the second term from both sides, we get:

$$\sqrt{x^2 + y^2 + 2cy + c^2} = \sqrt{x^2 + y^2 - 2cy + c^2} \pm 2a, \tag{6.3.8}$$

so when we square both sides, we obtain

$$\begin{aligned}
x^2 + y^2 + 2cx + c^2 &= \left(\sqrt{x^2 + y^2 - 2cx + c^2} \pm 2a\right)^2 =\\
&= x^2 + y^2 - 2cy + c^2 + 4a^2 \pm 4a\sqrt{x^2 + y^2 - 2cy + c^2}.
\end{aligned}$$

Upon subtracting the first four terms from the left-hand side from the right-hand side, we obtain:

$$x^2 + y^2 + 2cy + c^2 - (x^2 + y^2 - 2cy + c^2 + 4a^2) = \pm 4a\sqrt{x^2 + y^2 - 2cx + c^2}.$$

By opening the parentheses, after all cancellations, we get:

$$4cy - 4a^2 = \pm 4a\sqrt{x^2 + y^2 - 2cx + c^2},$$

which by dividing everything by 4 yields:

$$cy - a^2 = \pm a\sqrt{x^2 + y^2 - 2cx + c^2}. \tag{6.3.9}$$

By squaring both sides of (6.3.9), we get $(cy - a^2)^2 = a^2(x^2 + y^2 - 2cx + c^2)$, which reads:

$$a^4 - 2a^2cy + c^2y^2 = a^2x^2 + a^2y^2 - 2a^2cy + a^2c^2. \tag{6.3.10}$$

By moving all terms in from right- to left-hand side, except for a^2c^2, and moving a^4 from left- to right-hand side, the terms $2a^2cy$ cancel, so by grouping we now get

$$(c^2 - a^2)y^2 - a^2x^2 = a^2c^2 - a^4a^2(c^2 - a^2). \tag{6.3.11}$$

At this point, we define the number $b = \sqrt{c^2 - a^2}$, so so the above equation simply reads

$$b^2y^2 - a^2x^2 = a^2b^2, \tag{6.3.12}$$

so upon dividing everything by the right-hand side and simplifying, we finally get:

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1. \tag{6.3.13}$$

In the process of obtaining the equation (6.3.13), we used the *squaring* method twice. Basically we converted from an equation of the form $\sqrt{\heartsuit} = \#$ to its squared version: $\heartsuit = \#^2$, so we need to be concerned with the possibility that we may have added *extraneous solutions*. Concerning this possibility, all we have to notice here are the following two easy facts:

(i) For any fixed value of x, the equation (6.3.12), which can be re-written as

$$y^2 = \frac{a^2b^2 + a^2x^2}{b^2} = \left(\frac{a}{b}\right)^2(b^2 + x^2),$$

has *exactly two solutions*, namely

$$y = \pm\frac{a}{b}\sqrt{b^2 + x^2}. \tag{6.3.14}$$

(ii) For any fixed value of x, there are again exactly two values of y that give rise to a point $P(x, y)$ on the hyperbola.

Since the process of deriving (6.3.13) from the original equation (6.3.7) does not increase the number of solutions y (for any given x), these two equations are indeed equivalent.

Concerning the two *asymptotes* of a hyperbola given by the equation (6.3.13), let us examine the behavior of the solutions (6.3.14) as x gets large. Consider for instance the "$+$" solution from (6.3.14), and present it as a function

$$y = f(x) = \frac{a}{b}\sqrt{b^2 + x^2},$$

and let us see what happens as x gets large (and positive). When we subtract $g(x) = \frac{a}{b}x$ from $f(x)$, a simple calculation, based on the formula $\sqrt{\heartsuit} - \spadesuit = \frac{\heartsuit - \spadesuit^2}{\sqrt{\heartsuit} + \spadesuit}$ yields

$$f(x) - g(x) = \frac{a}{b}\left(\sqrt{b^2 + x^2} - x\right) \frac{a}{b} \cdot \frac{(b^2 + x^2) - x^2}{\sqrt{b^2 + x^2} + x} = \frac{a}{b} \cdot \frac{b^2}{\sqrt{b^2 + x^2} + x} = \frac{ab}{\sqrt{b^2 + x^2} + x}.$$

Since the denominator is positive and greater than x, what we get this way is that

$$0 < f(x) - g(x) < \frac{ab}{x}, \text{ for all } x > 0.$$

As x gets larger and larger, the quantity $\frac{ab}{x}$ gets closer and closer to zero, so $f(x) - g(x)$ also gets closer and closer to zero. What this proves, is the fact that *the line* $y = g(x) = \frac{a}{b}x$ *represents a slant asymptote for the "right end" of the graph of* $y = f(x) = \frac{a}{b}\sqrt{b^2 + x^2}$*, thus this line is an asymptote for the hyperbola*. By symmetry, the line $y = -g(x) = -\frac{a}{b}x$ is also an asymptote for the hyperbola. In conclusion (by symmetry, these lines are asymptotes also for the "left ends" of the hyperbola), it follows that:

> The asymptotes of an "easy" hyperbola with *vertical major axis*, which is always given by an equation of the form (6.3.13), are the lines
> $$y = \pm\frac{a}{b}x \tag{6.3.15}$$

From this fact it follows immediately that *the inner bounding box for a hyperbola given by* (6.3.13) *has exactly sides* $2a$ *(in the major axis direction) and* $2b$ *(in the minor axis direction), where* $b = \sqrt{a^2 - c^2}$. This statement is equivalent to the Focal Distance Formula.

A similar calculation can be obtained for an "easy" hyperbola, with *horizontal major axis*, and the equation will have the form:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{6.3.16}$$

where now the foci will be $F(c, 0)$ and $F'(-c, 0)$, while the vertices will have coordinates $V(a, 0)$ and $V'(-a, 0)$, and the asymptotes will be

$$y = \pm\frac{b}{a}x \tag{6.3.17}$$

As for standard position hyperbolas, which have the center at an arbitrary location, say $Z(x_Z, y_Z)$, the equations will be either (6.3.16) or (6.3.13), with

($\star$) x *replaced by* $x - x_Z$, and y *replaced by* $y - y_Z$.

These findings are summarized as follows.

Equations of Hyperbolas in Standard Position

Assume a hyperbola $\mathscr{H}$ is in *standard position*, with center $Z(x_Z, y_Z)$, major radius a, minor radius b and focal distance c. Then, depending on the major (or minor) axis orientation, the equation of $\mathscr{H}$ has one of the following forms.

(A) If the hyperbola $\mathscr{H}$ has *horizontal major axis* (or equivalently, *vertical minor axis*), then $\mathscr{H}$ can be presented by an equation of the form

$$\frac{(x - x_Z)^2}{a^2} - \frac{(y - y_Z)^2}{b^2} = 1, \tag{6.3.18}$$

and the hyperbola has:
- vertices located at $(x_Z \pm a, y_Z)$;
- foci located at $(x_Z \pm c, y_Z)$;
- minor points located at $(x_Z, y_Z \pm b)$;
- asymptotes given by the equations $y - y_Z = \pm \frac{b}{a}(x - x_Z)$.

(B) If the hyperbola $\mathscr{H}$ has *vertical major axis* (or equivalently, *horizontal minor axis*), then $\mathscr{H}$ can be presented by an equation of the form

$$\frac{(y - y_Z)^2}{a^2} - \frac{(x - x_Z)^2}{b^2} = 1, \tag{6.3.19}$$

and the hyperbola has:
- vertices located at $(x_Z, y_Z \pm a)$;
- foci located at $(x_Z, y_Z \pm c)$;
- minor points located at $(x_Z \pm b, y_Z)$;
- asymptotes given by the equations $y - y_Z = \pm \frac{a}{b}(x - x_Z)$.

Finding Equations of Hyperbolas from Geometric Data

Based on the information we have concerning the form of the equations of a standard position hyperbola, we can find these equations solely based on geometric data. In other words, we are able to solve the following type of problem.

Geometric-to-Analytic Problem

Given (enough) geometric information about a hyperbola, find its equation.

Such a problem has the same "flavor" as the corresponding one for ellipses, which we treated in Section 6.2. In particular, this problem has many possible instances, but the methodology of solving it is very similar to the one for ellipses, thus the key steps are as follows.

I. Determine the *orientation*, that is, figure out if the given information allows you to find the orientation of one of the axes. Using this information, start building the template for the equation, which is either (6.3.18) or (6.3.19).

II. Locate the *center*. Remember that the center is the common midpoint of each of the axes, as well as the midpoint of the segment determined by the foci.

III. Compute some *shape parameters*, based on the given information. Remember that each one of the parameters a, b, c measures either a distance between two distinguished points, or half of a length of a distinguished segment:

- The major radius a is equal to the *distance from the center to either one of the vertices*; equivalently a is equal to *half the distance between the vertices*. The major radius a is also equal to the *distance between any focus and any minor point*.
- The minor radius b is equal to the *distance from the center to either one of the minor points*; equivalently b is equal to *half the distance between the minor points*.
- The focal distance c is equal to the *distance from the center to either one of the foci*; equivalently c is equal to *half the distance between the foci*.

Since for building the equation, we only need a^2 and b^2, in case one of them is missing, but we are able to find either the focal distance c, or the eccentricity $\boldsymbol{e}$, we can use the Focal Distance Formula, or the eccentricity formulas:

- $a^2 = c^2 - b^2 = \dfrac{c^2}{\boldsymbol{e}^2} = \dfrac{b^2}{\boldsymbol{e}^2 - 1}$;
- $b^2 = c^2 - a^2 = a^2(\boldsymbol{e}^2 - 1) = \dfrac{c^2(\boldsymbol{e}^2 - 1)}{\boldsymbol{e}^2}$.

Example 6.3.1. Suppose we want to find the equation of a hyperbola with vertices $V(-1,1)$ and $V'(-1,5)$, and one focus at $F(-1,7)$.

Since the vertices, have equal x-coordinates, they lie on the vertical line $x = -1$. In particular, our hyperbola will have *vertical major axis*, so its equation looks like

$$\frac{(y - y_Z)^2}{a^2} - \frac{(x - x_Z)^2}{b^2} = 1, \tag{6.3.20}$$

Since the center Z is the midpoint of the segment determined by the vertices, it will also lie on the the vertical line $x = -1$, so the x-coordinate of the center will be $x_Z = -1$. As for the y-coordinate of the center, it will be equal to the average of the y-coordinates of the vertices, so:

$$y_Z = \tfrac{1}{2}\big(1 + 5\big) = \tfrac{1}{2} \cdot 6 = 3.$$

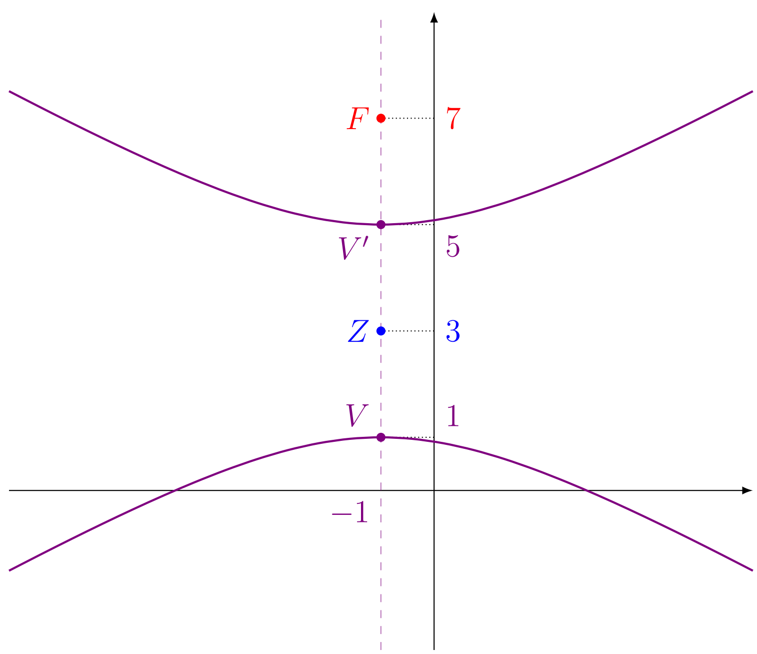


Figure 6.3.2

Using the center $Z(-1, 3)$ and of one vertex, say $V(-1, 5)$ it is clear that the major radius is:

$$a = \text{dist}(Z, V) = 2.$$

Using the center $Z(-1, 3)$ and the given focus $F(-1, 7)$ it is clear that the focal distance is:

$$c = \text{dist}(Z, F) = 4.$$

Using the Focal Distance Formula, the minor radius b satisfies

$$b^2 = c^2 - a^2 = 4^2 - 2^2 = 16 - 4 = 12.$$

Now we have all ingredients needed to fill in the template (6.3.20), so our equation becomes:

$$\frac{(y-3)^2}{4} - \frac{(x+1)^2}{12} = 1. \tag{6.3.21}$$

(The numerator of the first fraction in (6.3.20) is $(x - x_Z)^2 = (x - (-1))^2 = (x+1)^2$; the numerator of the second fraction in (6.3.20) is $(y - y_Z)^2 = (y-3)^2$.)

From Equations to Geometry

The Geometric-to-Analytic Problem discussed earlier can be reversed, so we can also consider the following.

Analytic-to-Geometric Problem

Given the equation of a hyperbola, obtain all geometric information about the hyperbola.

By "playing" a little bit with the equations of standard position hyperbolas, it is pretty clear that these equations can be presented in the following special form of the general quadratic equation (6.0.1) which looks like:

$$Ax^2 + Cy^2 + Dx + Ey + F = 0, \quad \text{with } A \text{ and } C \textit{ non-zero numbers with opposite signs}. \tag{6.3.22}$$

An equation of this form is called a **standard hyperbolic equation**. Using *square completion*, any standard hyperbolic equation can be studied as follows

Standard Hyperbolic Equation Analysis

Any standard hyperbolic equation (6.3.22) can be transformed by square completion into and equation of the form

$$A(x-h)^2 + C(y-k)^2 = M, \tag{6.3.23}$$

with A and C *non-zero numbers with opposite signs*.
Depending on whether the right-hand side M is zero or non-zero, the equation (6.3.23) represents one of the following curves.

(A) If $M = 0$, then (6.3.23) represents *two lines that intersect at* (h, k).

(B) If $M \neq 0$, then (6.3.23) represents a *hyperbola in standard position*, with *center* $Z(h, k)$. More specifically:

(B$_1$) If M has *same sign as* A, then (6.3.23) represents an *hyperbola with horizontal major axis*, with
- major radius $a = \sqrt{M/A}$, and
- minor radius $b = \sqrt{-M/C}$.

(B$_1$) If M has *same sign as* C, then (6.3.23) represents an *hyperbola with vertical major axis*, with
- major radius $a = \sqrt{M/C}$, and
- minor radius $b = \sqrt{-M/A}$.

CLARIFICATIONS. In the case when the right-hand side M of (6.3.23) is zero, then the equation can be re-written as

$$\frac{A}{C}(x-h)^2 + (y-k)^2 = 0,$$

where $\frac{A}{C}$ is some negative number, which can always be presented as $\frac{A}{C} = -m^2$, so (6.3.23) becomes

$$(y-k)^2 = m^2(x-h)^2,$$

which is equivalent to:

$$y - k = \pm m(x-h),$$

and statement (A) is now clear.

As for the remaining case (B), all we have to observe is that, when $M \neq 0$, the equation (6.3.23) can be easily transformed (by dividing everything by M) into

$$\frac{(x-h)^2}{M/A} + \frac{(y-k)^2}{M/C} = 1, \tag{6.3.24}$$

which clearly matches the equation of a hyperbola, since the denominators have *opposite signs*.

In practice, when we are asked to determine the geometry of a curve represented by a hyperbolic equation of the form (6.3.22), it is best if we carry on the following steps.

I. Use square completion to transform (6.3.22) into (6.3.23) , then briefly analyze the equation to decide if we are in the extreme case (A).

II. Assuming we are in case (B), do a second transformation, to get to the form (6.3.24), and match it with the equation of a standard position hyperbola: either (6.3.18) or (6.3.19).

Example 6.3.2. Consider the standard hyperbolic equation

$$4x^2 - 9y^2 + 32x + 18y + 91 = 0. \tag{6.3.25}$$

and let us find all geometric information of the curve it represents: center, foci, vertices, minor points and asymptotes.

We start off by completing squares in the left-hand side. When we concentrate on the terms that contain x, using the easy square completion identity we can write

$$4x^2 + 32x = 4(x+4)^2 - 64. \tag{6.3.26}$$

When we concentrate on the terms that contain y, using the easy square completion identity we can write

$$-9y^2 + 18y = -9(y-1)^2 + 9. \tag{6.3.27}$$

When we use (6.3.26) and (6.3.27) back in (6.3.25), our equation becomes

$$4(x+4)^2 - 9(y-1)^2 - 64 + 9 + 91 = 0,$$

which after grouping the constant terms and "moving" the result on the right, becomes:

$$4(x+4)^2 - 9(y-1)^2 = -36. \tag{6.3.28}$$

Next we divide everything by -36 and we get

$$\frac{(x+4)^2}{-36/4} - \frac{(y-1)^2}{-36/9} = 1.$$

which after simplifications in the denominators becomes

$$\frac{(y-1)^2}{4} - \frac{(x+4)^2}{9} = 1. \tag{6.3.29}$$

What we see here is an equation with right-hand side equal to 1 and the left-hand side is a combination of two fractions, with positive denominators, and numerators $(x+4)^2$ and $(y-1)^2$. When the left-hand side is put together,

(i) one of these fractions is *added*, namely $\dfrac{(y-1)^2}{4}$;

(ii) the other fraction gets *subtracted*, namely $\dfrac{(x+4)^2}{9}$.

Based on these observations, the correct way to match (6.3.29) with the equation of a standard position hyperbola is to match a^2 with *the denominator of fraction that gets added*, and to match b^2 with *the denominator of fraction that gets subtracted*. In our case, this matching leads to $a^2 = 4$ and $b^2 = 9$. Using the above analysis, it now follows that (6.3.29) represents a hyperbola with *vertical major axis*, with major radius $a = \sqrt{4} = 2$, and the minor radius $b = \sqrt{9} = 3$. Additionally, the focal distance is

$$c = \sqrt{a^2 + b^2} = \sqrt{4+9} = \sqrt{13},$$

so the eccentricity of our hyperbola is:

$$\boldsymbol{e} = \frac{c}{a} = \frac{\sqrt{13}}{2}.$$

Furthermore, the center of the hyperbola is $Z(-4,1)$. (Be careful here! The term $x+4$ matches $x-h$, so $h=-4$. The term $y-1$ matches $y-k$, so $k=1$.)

Since our hyperbola has vertical major axis, its geometric elements are as follows

- The vertices are located at $(x_Z, y_Z \pm a) = (-4, 1 \pm 2)$, so they are the points $V(-4, 3)$ and $V'(-4, -1)$.
- The minor points are located at $(x_Z \pm b, y_Z) = (-4 \pm 3, 1)$, so they are the points $M(-1, 1)$ and $M'(-7, 1)$.
- The foci are located at $(x_Z, y_Z \pm c) = (-4, 1 \pm \frac{\sqrt{13}}{2})$, so they are the points $F(-4, 1 + \frac{\sqrt{13}}{2})$ and $F'(-4, 1 - \frac{\sqrt{13}}{2})$.
- The asymtotes are: $\frac{y - y_Z}{a} = \pm \frac{x - x_Z}{b}$, which with our specific numbers become:

$$\frac{y-1}{2} = \pm \frac{x+4}{3}.$$

Graphing Hyperbolas

If we are asked to graph the curve represented by a standard hyperbolic equation, one option is to *complete squares*, thus rewriting our equation in the form (6.3.23), and then *solve for* y, which after some easy algebraic manipulations reduces to solving an equation of the form

$$(y-k)^2 = \textit{expression in } x, \tag{6.3.30}$$

which can then be easily solved by taking square roots $y - k = \pm\sqrt{\textit{expression in } x}$, to yield

$$y = k \pm \sqrt{\textit{expression in } x}. \tag{6.3.31}$$

Strictly speaking, in order to get to an equation like (6.3.30), our square completion should be done only on the terms that involve y. In fact, to obtain the solutions (6.3.31), we do not need square completion at all: we could use the Quadratic Formula directly, on the original equation! Once we get to (6.3.31), all we need to do is to graph *both* equations given by (6.3.31), one for "+" and one for "−," on a calculator.

Example 6.3.3. Let us revisit the standard hyperbolic equation

$$4x^2 - 9y^2 + 32x + 18y + 91 = 0. \tag{6.3.32}$$

which we analyzed in Example 6.3.2, and let us assume we are only asked to graph it.

We can set up our equation as a quadratic equation with y as the unknown, like

$$-9y^2 + 18y + (4x^2 + 32x + 91) = 0,$$

and then using the Quadratic Formula, we can solve for y:

$$\begin{aligned}
y &= \frac{-18 \pm \sqrt{(-18)^2 - 4\cdot(-9)\cdot(4x^2+32x+91)}}{2\cdot(-9)} = \frac{-18 \pm \sqrt{36\,[9 + (4x^2+32x+91)]}}{-18} = \\
&= \frac{-18 \pm \sqrt{36\,[4x^2+32x+100]}}{-18} = \frac{-18 \pm \sqrt{36\cdot 4\cdot [x^2+8x+25]}}{-18} = \\
&= \frac{-18 \pm 12\sqrt{x^2+8x+25}}{-18} = \frac{-18}{-18} \pm \frac{12\sqrt{x^2+8x+25}}{-18} = \\
&= 1 \pm \frac{2\sqrt{x^2+8x+25}}{3}.
\end{aligned}$$

Therefore, in order to graph our hyperbola on a calculator, we need to graph both of the following equations

$$y = 1 + \frac{2\sqrt{x^2+8x+25}}{3},$$
$$y = 1 - \frac{2\sqrt{x^2+8x+25}}{3},$$

and the result is depicted in Figure below.

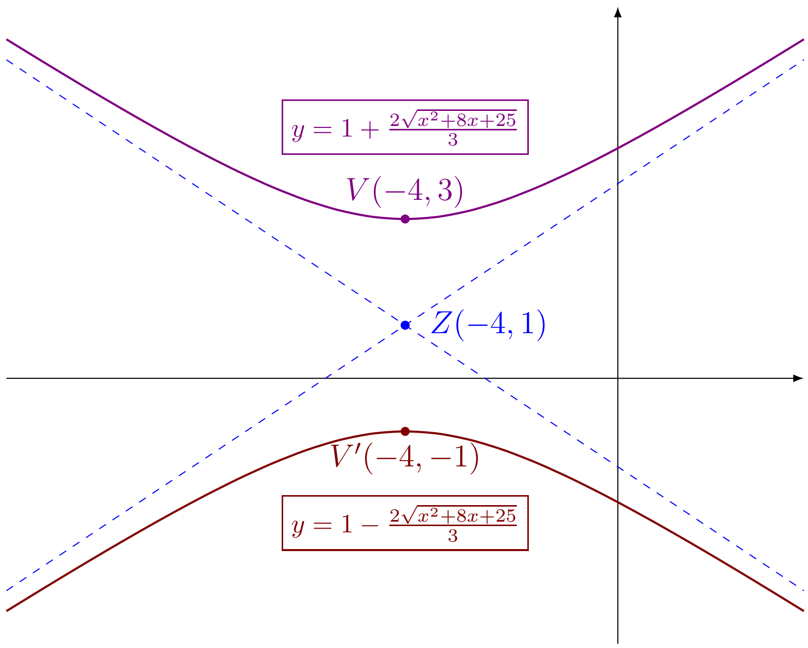


Figure 6.3.3

What happens if we do not have a calculator? We can sketch the graph of our hyperbola *by hand*, using the complete analysis of our equation, as shown in Example 6.3.2. Using our findings concerning the geometry of our hyperbola, we can sketch a *inner bounding box* for our hyperbola, which is a rectangle that has the vertices and the minor points at the midpoints of the sides of the rectangle. A sketch using this bounding box is shown in Figure 6.3.4 below, which also includes the minor points, as well as the asymptotes (which are the *diagonals of the bounding box*).

Unlike what we have seen with ellipses, the inner bounding box alone does not completely characterize a hyperbola, because the four "special" points on it – the two vertices, and the two minor points – need to be identified. Therefore, if we want to instruct a "bounding box software," (such as the one supplied on the K-STATE ONLINE HOMEWORK SYSTEM) on how to draw a hyperbola, besides identifying the bounding box itself, we must specify the *orientation of the focal axis*.

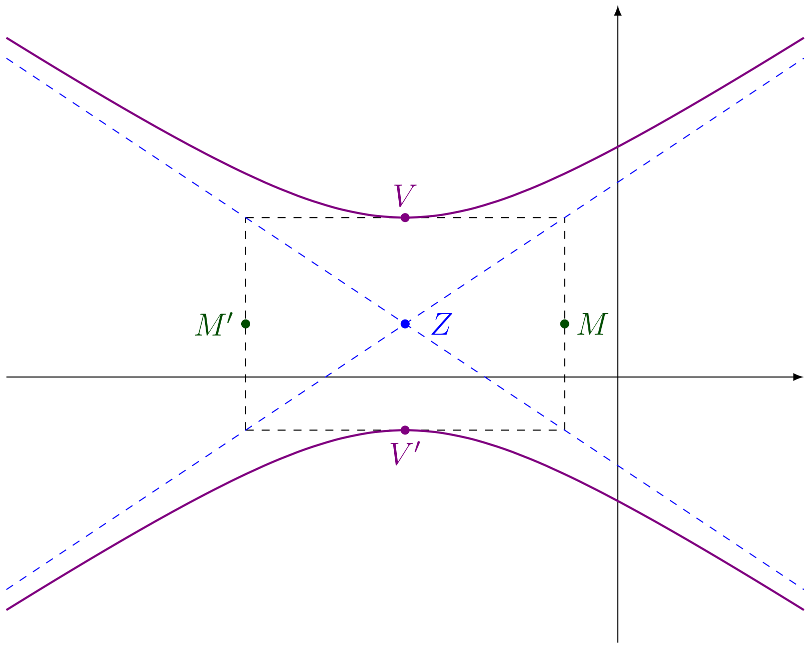


Figure 6.3.4

Exercises

In Exercises 1-14 you are asked to find the equation of a hyperbola, based on the given information.

1. Vertices $V(2,2)$ and $V'(-4,2)$; one focus at $F(3,2)$.
2. Vertices $V(2,-3)$ and $V'(2,7)$; one focus at $F(2,-5)$.
3. Vertices $V(2,0)$ and $V'(2,10)$; one minor point at $M(4,5)$.
4. Vertices $V(-3,1)$ and $V'(5,1)$; one minor point at $M(2,2)$.
5. Foci $F(-2,3)$ and $F'(-2,9)$; one vertex at $V(-2,4)$.
6. Foci $F(0,0)$ and $F'(8,0)$; one vertex at $V(2,0)$.
7. Minor points $M(-6,5)$ and $M'(4,5)$; one focus at $F(-1,8)$.
8. Minor points $M(-1,-1)$ and $M'(5,-1)$; one focus at $F(2,-6)$.
9. Foci $F(-2,0)$ and $F'(-2,6)$; one minor point at $M(-4,3)$.
10. Foci $F(-2,0)$ and $F'(8,0)$; one minor point at $M(3,-1)$.
11. Asymptotes $y+2=\pm 2(x-4)$; one vertex at $V(0,-2)$.
12. Asymptotes $y+2=\pm 3(x-4)$; one vertex at $V(4,5)$.
13. Foci $F(1,1)$ and $F'(9,1)$; eccentricity 2.

14. Vertices $V(-7,1)$ and $V'(1,1)$; eccentricity 3.

In Exercises 15-20 you are given a standard hyperbolic equation, and are asked to

(i) indicate the major axis orientation;
(ii) find the the center;
(iii) find the vertices;
(iv) find the foci;
(v) find the minor points;
(vi) find the asymptotes;
(vii) sketch the graph.

15. $4x^2 - 9y^2 - 36 = 0.$

16. $16x^2 - y^2 + 144 = 0.$

17. $4(x+2)^2 - 9(y-3)^2 = 576.$

18. $4x^2 - y^2 + 8x - 2y - 13 = 0.$

19. $4x^2 - 9y^2 - 24x - 18y + 11 = 0.$

20. $4x^2 - 9y^2 + 16x - 36y + 16 = 0.$

6.4 Quadratic Curves in General Position

In this section we will learn how to analyze the curves given by a general quadratic equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \tag{6.4.1}$$

where A, B, C, D, E, F are constants. When analyzing such an equation, the left-hand side of (6.4.1) is split into three parts, according to the degree of the terms:

(i) the *quadratic part*, which is the expression

$$\mathrm{Q}(x,y) = Ax^2 + Bxy + Cy^2; \tag{6.4.2}$$

(ii) the *linear part*, which is the expression

$$\mathrm{L}(x,y) = Dx + Ey; \tag{6.4.3}$$

(iii) a *constant term:* F.

As a matter of terminology, an expression given by (6.4.2) is called a *quadratic form*, and an expression given by (6.4.3) is called a *linear form*. Of course, our main assumption on the general equations of the form (6.4.1) that we are going to investigate will be that *the quadratic part is non-zero*, which means that, *at least one of the coefficients* A, B, C *is non-zero*.

In the preceding three sections we learned how to handle the case when *the coefficient* B *(of the mixed term* xy*) is equal to zero*, in which case our equation looks like

$$Ax^2 + Cy^2 + Dx + Ey + F = 0. \tag{6.4.4}$$

The summary of what we found out in Sections 6.1, 6.2 and 6.3 is as follows.

I. (The *parabolic* case.) If *either A or C is zero* (but not both!), then the equation (6.4.4) represents either (A) a *parabola in standard position*, or (B) *one line*, or (C) *two parallel lines*, or (D) *the empty set*.
II. (The *elliptic* case.) If *A and C are non-zero and have same sign*, then the equation (6.4.4) represents either (A_1) an *ellipse in standard position*, or (A_2) a *circle*, or (B) *one point*, or (C) *the empty set*.
III. (The *hyperbolic* case.) If *A and C are non-zero and have opposite signs*, then the equation (6.4.4) represents either (A) a *hyperbola in standard position*, or (B) *two intersecting lines*.

In this section we will learn how to handle the case when *the coefficient B (of the mixed term xy) is non-zero*

Coordinate Systems

As we learned in Section 3.1, setting up a *coordinate system* in the plane simply amounts to

- Fixing a point O in the plane – the **origin**, and
- Fixing a **perpendicular frame**, which is a pair $(\vec{\mathbf{e}}, \vec{\mathbf{f}})$ consisting of *two unit vectors, which are perpendicular*.

Once such a coordinate system is set, the coordinates (u, v) of any point P in the plane given by:

$$\begin{cases} u &= \overrightarrow{OP} \bullet \vec{\mathbf{e}} \\ v &= \overrightarrow{OP} \bullet \vec{\mathbf{f}} \end{cases} \tag{6.4.5}$$

where $\overrightarrow{OP}$ is the *position vector of P relative to the given origin O*.

According to the Orthogonal Decomposition Theorem (see Section 3.1), the pair (u, v) is the *unique ordered pair of numbers that satisfies the identity:*

$$u\vec{\mathbf{e}} + v\vec{\mathbf{f}} = \overrightarrow{OP}. \tag{6.4.6}$$

CONVENTIONS. When doing explicit calculations, we assume once and for all that we have fixed one coordinate system in the plane, which we will refer to as the *standard coordinate system*, so that its two frame vectors $\vec{\mathbf{e}}_{\text{st}}$, $\vec{\mathbf{f}}_{\text{st}}$ point in the East and North directions, respectively. The standard coordinates defined by this system will be denoted the usual way by x and y. All vectors in the plane will be denoted using matrix notation as $\vec{\mathbf{v}} = \begin{bmatrix} a \\ b \end{bmatrix}$. The actual meaning of this matrix notation is nothing else but the vector equality

$$a\vec{\mathbf{e}}_{\text{st}} + b\vec{\mathbf{f}}_{\text{st}} = \vec{\mathbf{v}}.$$

Throughout this entire section, we will limit ourselves to the case when the perpendicular frames $(\vec{\mathbf{e}}, \vec{\mathbf{f}})$, used in building up new coordinate systems, all have *positive orientation*, which means that $\vec{\mathbf{e}} \wedge \vec{\mathbf{f}} = 1$. This condition simply says that *the unit vector $\vec{\mathbf{f}}$ is obtained by rotating the unit vector $\vec{\mathbf{e}}$ by $90°$ in the counterclockwise direction*. In particular, if we specify the first vector in (standard) coordinates, $\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}$, then the condition that $\vec{\mathbf{e}}$ is a *unit* vector reads $a^2 + b^2 = 1$, and furthermore, the second vector in the frame must be $\vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}$. With all these observations in mind, the construction of (new) coordinate systems can be summarized as follows.

To give a (new) *positively oriented coordinate system* in the plane amounts to choosing two ordered pairs (h, k) and (a, b) of real numbers, with the second one satisfying the identity

$$a^2 + b^2 = 1.$$

Once such choices are made, the coordinate system is built by

- fixing the *origin* as the $O(h, k)$;
- fixing the coordinate frame vectors $(\vec{\mathbf{e}}, \vec{\mathbf{f}})$ as: $\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}$.

CLARIFICATION. Once the (new) origin O and the vectors $\vec{\mathbf{e}}$ and $\vec{\mathbf{f}}$ are set up according to the above recipe, the easy way to understand how the new coordinate system works is as follows. We place vectors $\vec{\mathbf{e}}$ and $\vec{\mathbf{f}}$ to start at O and we make them "responsible" for the positive u-axis and v-axis, respectively, and now each point P can be presented in the (new) coordinates (u, v).

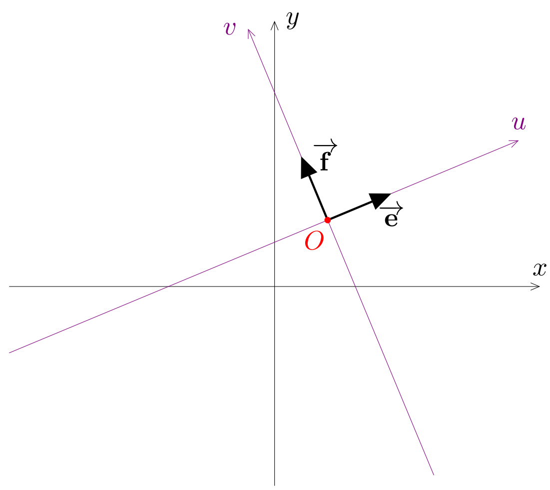


Figure 6.4.1

If our point is presented in standard coordinates as $P(x, y)$, then the new coordinates are given by the following scheme.

Standard-to-New Coordinate Change Formulas

Assume a new positively oriented coordinate system with origin $O(h, k)$, and frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}\right)$, is given, where $a^2 + b^2 = 1$. Then the new coordinates (u, v) of a point, presented in standard coordinates as $P(x, y)$, are given by:

$$\begin{cases} u &= a(x-h) + b(y-k) \\ v &= -b(x-h) + a(y-k) \end{cases} \tag{6.4.7}$$

CLARIFICATION. The formulas (6.4.7) follow immediately from (6.4.5), using the Dot Product formulas, with $\overrightarrow{OP} = \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} h \\ k \end{bmatrix} = \begin{bmatrix} x-h \\ y-k \end{bmatrix}$, and $\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}$.

Example 6.4.1. Suppose we build a (new) positively oriented coordinate system with origin $O(-1,1)$ and frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -4/5 \\ 3/5 \end{bmatrix}\right.$, and we want to find the (new) coordinates (u,v) of the point given in standard coordinates as $P(2,-2)$.

Using formulas (6.4.7), these coordinates are computed as

$$u = \frac{3}{5}(2-(-1)) + \frac{4}{5}(-2-1) = \frac{9}{5} - \frac{12}{5} = \frac{1}{5};$$
$$v = -\frac{4}{5}(2-(-1)) + \frac{3}{5}(-2-1) = -\frac{12}{5} - \frac{9}{5} = -\frac{21}{5}.$$

The formulas (6.4.7) can also be "undone," using the equation (6.4.6), thus yielding the following statement.

New-to-Standard Coordinate Change Formulas

Assume a new positively oriented coordinate system with origin $O(h,k)$, and frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}\right)$, is given, where $a^2+b^2=1$. If a point P has new coordinates (u,v), then its standard coordinates (x,y) are given by:

$$\begin{cases} x &= au - bv + h \\ y &= bu + av + k \end{cases} \tag{6.4.8}$$

The New-to-Standard-to Coordinate Change Formulas are particularly useful when *transforming equations*.

Standard-to-New Equation Change Formulas

Assume $\mathscr{C}$ is a curve given in standard coordinates (x,y) by an equation of the form

$$\mathsf{E}(x,y) = K,$$

where $\mathsf{E}(x,y)$ is some algebraic expression in x and y, and K is some contant. Assume also a (new) positively oriented coordinate system is built, with origin $O(h,k)$, frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}\right)$, is given, where $a^2+b^2=1$.
Using the (new) coordinates (u,v), the same curve $\mathscr{C}$ is represented by the equation

$$\mathsf{E}_{\text{new}}(u,v) = K,$$

where $\mathsf{E}_{\text{new}}(u,v)$ is obtained by *substituting x and y in $\mathsf{E}(x,y)$ by the expressions given by the Standard-to-New Formulas*. In other words, the new algebraic expression $\mathsf{E}_{\text{new}}(u,v)$ giving the equation of $\mathscr{C}$ in new coordinates is built as:

$$\mathsf{E}_{\text{new}}(u,v) = \mathsf{E}(au - bv + h, bu + av + k). \tag{6.4.9}$$

Example 6.4.2. Let us consider the same coordinate system we built in Example 6.4.1, and let us consider the line $\mathscr{L}$, given in standard coordinates by the equation

$$3x + 4y = 11.$$

Suppose now we want to find the equation of the same line, in the new coordinates. We start off by setting the left-hand side expression $\mathsf{E}(x,y) = 3x - 4y$, and then we "cook up" the new expression $\mathsf{E}_{\text{new}}(u,v)$ by replacing x and y using (6.4.9), so we get

$$\mathsf{E}_{\text{new}}(u,v) = \mathsf{E}\left(\frac{3}{5}u - \frac{4}{5}v + (-1), \frac{4}{5}u + \frac{3}{5}v + 1\right) = 3\left(\frac{3}{5}u - \frac{4}{5}v - 1\right) + 4\left(\frac{4}{5}u + \frac{3}{5}v + 1\right) =$$
$$= \frac{9}{5}u - \frac{12}{5}v - 3 + \frac{16}{5}u + \frac{12}{5}v + 4 = 5u + 1,$$

which means that the new equation of $\mathscr{L}$, using the new coordinates (u,v) is

$$5u + 1 = 11,$$

which is the same as: $u = 2$.

TIP. The easy way to remember the New-to-Standard-to Coordinate Change Formulas is to write them in *matrix form*:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix}, \tag{6.4.10}$$

where the matrix $\mathbf{R} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ is precisely the *rotation matrix, which represents the rotation transformation needed to turn the vector* $\vec{\mathbf{e}}_{\text{st}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *over the vector* $\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}$. Of course, if the *turning angle of* $\vec{\mathbf{e}}_{st}$ *over* $\vec{\mathbf{e}}$ is τ, then $a = \cos\tau$ and $b = \sin\tau$.

Changing Quadratic Equations Using Axes Rotations

We now apply the Standard-to-New Equation Change Formulas to general equations of the form (6.4.1), and see how the equation is transformed when we pass to a new positively coordinate system. We will simplify matters a little bit by restricting ourselves to the case when *origin is unchanged*, so $h = k = 0$.

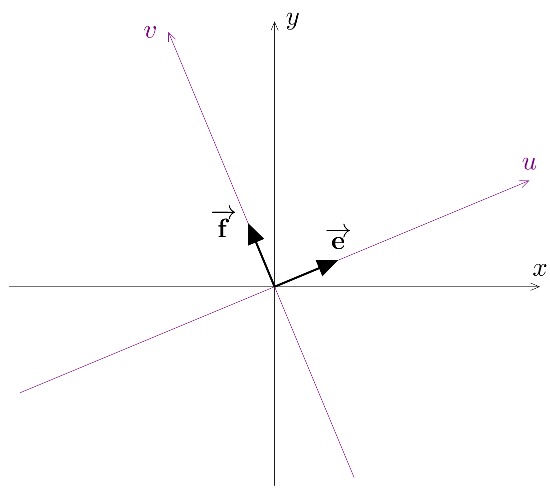


Figure 6.4.2

In other words, our new coordinate system is obtained by *rotating* the standard coordinate axes, as shown above. The new and old coordinates are now related to each other as follows.

Axes Rotation Identities

When the new coordinate system (u, v) is obtained by rotating the standard coordinate axes, so that

- the direction vector for the positive u-axis is $\vec{\mathbf{e}} = \begin{bmatrix} a \\ b \end{bmatrix}$, and
- the direction vector for the positive v-axis is $\vec{\mathbf{f}} = \begin{bmatrix} -b \\ a \end{bmatrix}$,

the new and old (standard) coordinates are related by the identities:

$$\begin{cases} u &= ax + by \\ v &= -bx + ay \end{cases} \qquad \begin{cases} x &= au - bv \\ y &= bu + av \end{cases} \tag{6.4.11}$$

Using these identities, we then know that the curve $\mathscr{C}$ represented by a general quadratic equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{6.4.12}$$

will be represented in new coordinates by the equation

$$A(au - bv)^2 + B(au - bv)(bu + av) + C(bu + av)^2 + D(au - bv) + E(bu + av) + F = 0,$$

which after doing all the simplifications and groupings will again be a new quadratic equation, of the form:

$$A_{\text{new}}u^2 + B_{\text{new}}uv + C_{\text{new}}v^2 + D_{\text{new}}u + E_{\text{new}}v + F = 0. \tag{6.4.13}$$

A careful calculation will give us the new coefficients as:

$$\begin{cases} A_{\text{new}} &= Aa^2 + Bab + Cb^2 \\ B_{\text{new}} &= -2Aab + B(a^2 - b^2) + 2Cab = B(a^2 - b^2) - 2(A - C)ab \\ C_{\text{new}} &= Ab^2 - Bab + Ca^2 \\ D_{\text{new}} &= Da + Eb \\ E_{\text{new}} &= -Db + Ea \end{cases} \tag{6.4.14}$$

TIP. A neat way produce the above formulas is to write them in matrix form as:

$$\begin{bmatrix} 2A_{\text{new}} & B_{\text{new}} \\ B_{\text{new}} & 2C_{\text{new}} \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} \cdot \begin{bmatrix} a & -b \\ b & a \end{bmatrix}; \tag{6.4.15}$$

$$\begin{bmatrix} D_{\text{new}} \\ E_{\text{new}} \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} D \\ E \end{bmatrix} \tag{6.4.16}$$

Our main task is use the formulas (6.4.14), in order to *produce a suitable coordinate system, for which the middle term B_{new} becomes zero!* Once this is accomplished, the transformed new equation (6.4.13) can be studied exactly as we explained in the beginning of this section. Based on the above formula for B_{new}, all we need to do is to find a pair (a, b), that satisfies the equations

$$\begin{cases} B(a^2 - b^2) - 2(A - C)ab &= 0 \\ a^2 + b^2 &= 1 \end{cases} \tag{6.4.17}$$

We do not need to completely solve this system of equations! Remembering that the second equation simply tells us that we can find some angle τ, such that $a = \cos\tau$ and $b = \sin\tau$, now the first equation will simply read:

$$B(\cos^2\tau - \sin^2\tau) - 2(A - C)\cos\tau\sin\tau = 0.$$

Using the formulas for Double Angles which we learned in Section 4.2, the above equation becomes

$$B\cos 2\tau - (A - C)\sin 2\tau = 0,$$

which is equivalent to: $\cot 2\tau = \dfrac{A - C}{B}$. (Of course, we need to solve (6.4.17), only when $B \neq 0$! Therefore the denominator in the above formula poses no problem.) In conclusion, we found the following method for finding a suitable new coordinate system:

General Quadratic Equation Reduction Principle

Assume $\mathscr{C}$ is a curve represented by a general quadratic equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \tag{6.4.18}$$

with $B \neq 0$, and τ is any angle that satisfies the equation

$$\cot 2\tau = \frac{A - C}{B}. \tag{6.4.19}$$

If we consider the coordinate system that has same origin as the standard system, and frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} \cos\tau \\ \sin\tau \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -\sin\tau \\ \cos\tau \end{bmatrix}\right)$, then the equation of $\mathscr{C}$ with respect to the new coordinates (u, v) is of the form

$$A_{\text{new}}u^2 + C_{\text{new}}v^2 + D_{\text{new}}u + E_{\text{new}}v + F = 0. \tag{6.4.20}$$

In particular, with our choice of τ, the curve $\mathscr{C}$ is a *(possibly degenerate) conic section, which has one symmetry line parallel or perpendicular to the vector* $\vec{\mathbf{e}}$, *so it will be in standard position, relative to the new coordinate system.*

Example 6.4.3. Consider the equation

$$xy = 1,$$

which is quite a famous one, as you may have seen it in the Algebra course.

We wish to find a suitable coordinate system, with same origin as the standard one, and with frame vectors $\left(\vec{\mathbf{e}} = \begin{bmatrix} \cos\tau \\ \sin\tau \end{bmatrix}, \vec{\mathbf{f}} = \begin{bmatrix} -\sin\tau \\ \cos\tau \end{bmatrix}\right)$, which allows us to determine the shape (and the geometry) of our curve.

Since $A = C = 0$ and $B = 1$, the equation (6.4.19) simply reads $\cot 2\tau = 0$, which has an easy solution $\tau = \dfrac{\pi}{4}$. Our coordinate frame vectors will now be:

$$\vec{\mathbf{e}} = \begin{bmatrix} \cos\dfrac{\pi}{4} \\ \\ \sin\dfrac{\pi}{4} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \end{bmatrix}; \qquad \vec{\mathbf{f}} = \begin{bmatrix} -\cos\dfrac{\pi}{4} \\ \\ \sin\dfrac{\pi}{4} \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \end{bmatrix}.$$

Using (6.4.14), the coefficients for the new equation of $\mathscr{C}$ – in the new coordinates (u, v) will be: $A_{\text{new}} = \dfrac{1}{2}$, $B_{\text{new}} = 0$, $C_{\text{new}} = -\dfrac{1}{2}$, and $D_{\text{new}} = E_{\text{new}} = 0$, so the new equation of our curve is:

$$\frac{1}{2}u^2 - \frac{1}{2}v^2 = 1,$$

which clearly represents a *hyperbola*, with center at the origin (which is the same in both coordinate systems), major and minor radii both equal to $\sqrt{2}$, and focal distance 2.

Tip. A particular solution of the system (6.4.17) can also be found *algebraically*, so we do not really need to find the angle τ. Instead, we can use the substitution $\dfrac{b}{a} = t$, so when we divide all terms in the first equation by a, that equation will become: $B(1-t^2) - 2(A-C)t = 0$, which we can also write (by taking negatives, and dividing by B) as:

$$t^2 + \frac{2(A-C)}{B}t - 1 = 0. \tag{6.4.21}$$

So one plan for finding a particular solution of (6.4.17) can be the following:

I. Solve (6.4.21) using for instance the Quadratic Formula, and retain one solution t.

II. With t found above, set $a = \dfrac{1}{\sqrt{1+t^2}}$ and $b = \dfrac{t}{\sqrt{1+t^2}}$.

(The second step uses the second equation in (6.4.17), which with our substitution simply yields $a^2(1+t^2) = 1$.)

Example 6.4.4. Consider the equation:

$$9x^2 + 16y^2 + 24xy - 4x + 3y - 10 = 0,$$

and again let us try to determine the shape of the curve represented by this equation, together with all its geometric elements.

The equation (6.4.21) becomes

$$t^2 - \frac{7}{12}t - 1 = 0,$$

which can be written equivalently as

$$12t^2 - 7t - 12 = 0.$$

Using the Quadratic Formula, we get

$$t = \frac{-(-7) \pm \sqrt{(-7)^2 - 4 \cdot 12 \cdot (-12)}}{2 \cdot 12} = \frac{7 \pm \sqrt{625}}{24} = \frac{7 \pm 25}{24}.$$

One particular solution is $t = \frac{7+25}{24} = \frac{32}{24} = \frac{4}{3}$. Using step II, we immediately get $a = \frac{3}{5}$ and $b = \frac{4}{5}$, so now using (6.4.14), the coefficients for the new equation of $\mathscr{C}$ – in the new coordinates (u, v) will be:

$$\begin{aligned}
A_{\text{new}} &= 9 \cdot \frac{9}{25} + 24 \cdot \frac{12}{25} + 16 \cdot \frac{16}{25} = \frac{81 + 288 + 256}{25} = \frac{625}{25} = 5;\\
B_{\text{new}} &= 0;\\
C_{\text{new}} &= 9 \cdot \frac{16}{25} - 24 \cdot \frac{12}{25} + 16 \cdot \frac{9}{25} = \frac{144 - 288 + 144}{25} = 0;\\
D_{\text{new}} &= -4 \cdot \frac{3}{5} + 3 \cdot \frac{4}{5} = 0;\\
E_{\text{new}} &= -(-4) \cdot \frac{4}{5} + 3 \cdot \frac{3}{5} = 5.
\end{aligned}$$

With these calculations, our new equation is:

$$5u^2 + 5v - 10 = 0,$$

which clearly represents a *parabola*. Since, after dividing everything by 5, we can re-write the new equation as

$$u^2 = (-1)(v - 2),$$

we can be a bit more precise about our parabola:

(i) The focal parameter of the parabola is $p = -\frac{1}{4}$.

(ii) The *vertex* V of the parabola has (new) coordinates $u_V = 0$ and $v_V = 2$.

(iii) The *focus* F of the parabola has (new) coordinates $u_F = 0$ and $v_F = 2 + (-\frac{1}{4}) = \frac{7}{4}$.

(iv) The *focal axis* of the parabola is the line that has (new) equation

$$u = 0. \tag{6.4.22}$$

(In other words, the focal axis is nothing else but the v-axis.)

(v) The *directrix* of the parabola is the line that has (new) equation

$$v = 2 - (-\frac{1}{4}) = \frac{9}{4}. \tag{6.4.23}$$

In order to finish this problem, we have to *revert everything to the old coordinates* (x, y). For the vertex and the focus, we are going to use the second set of formulas from the Axes Rotation Identities (6.4.11), which reads:

$$\begin{cases} x &= \frac{3}{5}u - \frac{4}{5}v \\ y &= \frac{4}{5}u + \frac{3}{5}v \end{cases}$$

In particular, the *vertex* V will have (standard) coordinates

$$\begin{cases} x_V &= \frac{3}{5} \cdot 0 - \frac{4}{5} \cdot 2 = -\frac{8}{5} \\ y_V &= \frac{4}{5} \cdot 0 + \frac{3}{5} \cdot 2 = \frac{6}{5} \end{cases}$$

Likewise, the *focus* F will have (standard) coordinates

$$\begin{cases} x_F &= \frac{3}{5}\cdot 0 - \frac{4}{5}\cdot\frac{7}{4} = -\frac{7}{5} \\ \\ y_F &= \frac{4}{5}\cdot 0 + \frac{3}{5}\cdot\frac{7}{4} = \frac{21}{20} \end{cases}$$

As for the *focal axis* and the *directrix*, we use the first set of formulas from the Axes Rotation Identities (6.4.11), which reads:

$$\begin{cases} u &= \frac{3}{5}x + \frac{4}{5}y \\ \\ v &= -\frac{4}{5}x + \frac{3}{5}y \end{cases}$$

Using these identities with (new) equation (6.4.22), it follows that the *focal axis* of our parabola has equation:

$$\frac{3}{5}x + \frac{4}{5}y = 0.$$

Likewise, using the (new) equation (6.4.23), it follows that the *directrix* of our parabola has equation:

$$-\frac{4}{5}x + \frac{3}{5}y = \frac{7}{4}.$$

The complete depiction (except for the directrix) of our parabola is shown in the picture below

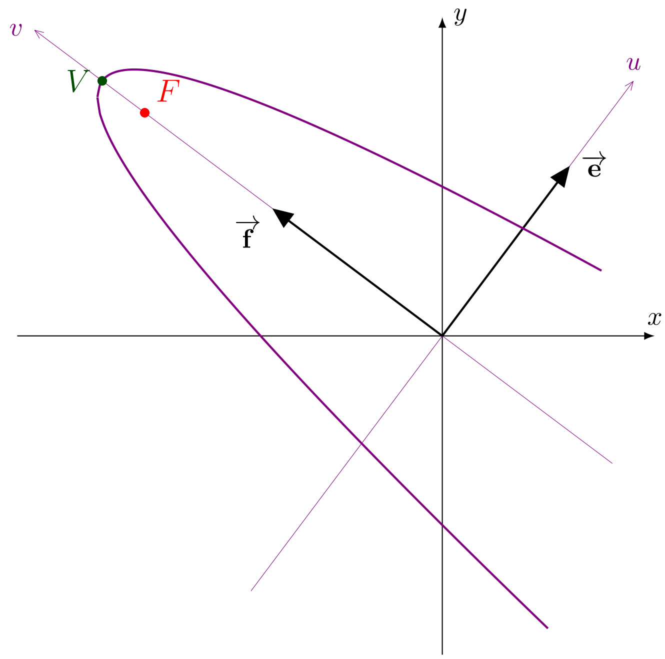


Figure 6.4.3

It is possible to graph this parabola on a graphing calculator, using the same technique as in Examples 6.2.3 and 6.3.3, which amounts to *solving the given equation for* y, *using the Quadratic Formula*. For this purpose, we regroup the terms in our equation, and re-write it as:

$$16y^2 + (24x + 3)y + 9x^2 - 4x - 10 = 0,$$

and solve it for y:

$$y = \frac{-(24x+3) \pm \sqrt{(24x+3)^2 - 4\cdot 16\cdot(9x^2-4x-10)}}{2\cdot 16}.$$

If we only want to graph these two equations (one for "+;" another for "−" sign), we can keep these formulas as they are. However, we want to work a bit more neatly, we ought to simply the quantity under the radical:

$$(24x+3)^2 - 4\cdot 16\cdot(9x^2-4x-10) = 576x^2+144x+9-576x^2+256x+640 = 400x+649,$$

so our equations will look a little simpler:

$$y = \frac{-24x-3 \pm \sqrt{400x+649}}{32}.$$

The Characteristic Identities

What we are about to discover is that, whenever we reduce a general equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{6.4.24}$$

to one of the form

$$A_{\text{new}}u^2 + C_{\text{new}}v^2 + D_{\text{new}}u + E_{\text{new}}v + F = 0, \tag{6.4.25}$$

with $B_{\text{new}} = 0$, it is also possible to find the values A_{new} and C_{new} without any need to find the new coordinate system. Let us take a closer look at the first three equalities in (6.4.14), which read:

$$\begin{cases} A_{\text{new}} &= Aa^2 + Bab + Cb^2 \\ C_{\text{new}} &= Ab^2 - Bab + Ca^2 \\ B_{\text{new}} &= -2Aab + B(a^2-b^2) + 2Cab = B(a^2-b^2) - 2(A-C)ab = 0 \end{cases} \tag{6.4.26}$$

When we add the first two equations, the terms Bab will cancel, so we get the *first characteristic identity*

$$A_{\text{new}} + C_{\text{new}} = A(a^2+b^2) + C(a^2+b^2) = A + C \tag{6.4.27}$$

(The last equality uses the equation $a^2 + b^2 = 1$.)

To obtain the other important identity, we set $a = \cos\tau$ and $b = \sin\tau$, where τ is a solution of $\cot 2\tau = \dfrac{A-C}{B}$. When we replace in (6.4.26) we can write

$$\begin{aligned} A_{\text{new}} &= A\cos^2\tau + B\sin\tau\cos\tau + C\sin^2\tau = \tfrac{1}{2}\left[A(1+\cos 2\tau) + B\sin 2\tau + C(1-\cos 2\tau)\right] = \\ &= \tfrac{1}{2}\left[A + C + (A-C)\cos 2\tau + B\sin 2\tau\right]; \\ C_{\text{new}} &= A\sin^2\tau - B\sin\tau\cos\tau + C\cos^2\tau = \tfrac{1}{2}\left[A(1-\cos 2\tau) - B\sin 2\tau + C(1+\cos 2\tau)\right] = \\ &= \tfrac{1}{2}\left[A + C - (A-C)\cos 2\tau - B\sin 2\tau\right]. \end{aligned}$$

When we multiply these two equalities and get rid of the $\frac{1}{2}$, we can use the difference of squares formula:

$$\begin{aligned} 4A_{\text{new}}C_{\text{new}} &= \left[A + C + (A-C)\cos 2\tau + B\sin 2\tau\right]\cdot\left[A + C - (A-C)\cos 2\tau - B\sin 2\tau\right] = \\ &= (A+C)^2 - \left[(A-C)\cos 2\tau + B\sin 2\tau\right]^2 = \\ &= (A+C)^2 - \left[(A-C)^2\cos^2 2\tau + 2(A-C)B\cos 2\tau\sin 2\tau + B^2\sin^2 2\tau\right]. \end{aligned} \tag{6.4.28}$$

Using the equation $\cot 2\tau = \dfrac{A-C}{B}$. we have $(A-C)\sin 2\tau = B\cos 2\tau$, so we can replace

$$(A-C)^2\cos^2 2\tau = (A-C)^2 - (A-C)^2\sin^2 2\tau = (A-C)^2 - B^2\cos^2 2\tau;$$
$$(A-C)B\cos 2\tau\sin 2\tau = B^2\cos^2 2\tau,$$

so when we go back to (6.4.28) we can write

$$\begin{aligned} 4A_{\text{new}}C_{\text{new}} &= (A+C)^2 - \left[(A-C)^2 - B^2\cos^2 2\tau + 2B^2\cos^2 2\tau + B^2\sin^2 2\tau\right] = \\ &= (A+C)^2 - \left[(A-C)^2 + B^2\cos^2 2\tau + B^2\sin^2 2\tau\right] = \\ &= (A+C)^2 - \left[(A-C)^2 + B^2\right] = (A+C)^2 - (A-C)^2 - B^2 = \\ &= A^2 + 2AC + C^2 - (A^2 - 2AC + C^2) - B^2 = 4AC - B^2. \end{aligned} \tag{6.4.29}$$

The above equation is referred to as the *second characteristic identity*.

Our findings are now summarized and enhanced as follows.

Characteristic Identities and Their Consequences

Whenever a general equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{6.4.30}$$

is reduced – with the help of a new coordinate system – to a standard equation of the form

$$A_{\text{new}}u^2 + C_{\text{new}}v^2 + D_{\text{new}}u + E_{\text{new}}v + F = 0, \tag{6.4.31}$$

the numbers A_{new} and C_{new} satisfy the identities

$$\begin{cases} A_{\text{new}} + C_{\text{new}} &= A + C \\ A_{\text{new}}C_{\text{new}} &= AC - \frac{1}{4}B^2 \end{cases} \tag{6.4.32}$$

In particular, the reduced equation (6.4.31) is:

(i) *parabolic*, if $AC - \frac{1}{4}B^2 = 0$;

(ii) *elliptic*, if $AC - \frac{1}{4}B^2 > 0$;

(iii) *hyperbolic*, if $AC - \frac{1}{4}B^2 < 0$.

The special number $\Delta = AC - \frac{1}{4}B^2$ is called the **discriminant** of the equation (6.4.30).

Example 6.4.5. Suppose we have a curve given by the equation

$$x^2 + 4xy + 3y^2 + x - 2y - 10 = 0, \tag{6.4.33}$$

and we are asked to determine the general shape of the curve.

The discriminant is $\Delta = 1\cdot 3 - \dfrac{4^2}{4} = 3 - 4 = -1$, so our curve must be a (possibly degenerate) *hyperbola*.

The Characteristic Equation

Using the characteristic identities (6.4.32) it follows that the two numbers A_{new} and C_{new} are precisely the solutions of the quadratic equation

$$\lambda^2 - (A+C)\lambda + \Delta = 0. \tag{6.4.34}$$

This follows from the obvious folding identity

$$(\lambda - A_{\text{new}})(\lambda - C_{\text{new}}) = \lambda^2 - (A_{\text{new}} + C_{\text{new}})\lambda + A_{\text{new}}C_{\text{new}},$$

so from (6.4.32), this matches exactly the left-hand side of (6.4.34).

The equation (6.4.34) is what we call the **characteristic equation** associated with the general quadratic equation (6.4.30). Of course, the mere fact that A_{new} and C_{new} are the solutions of the characteristic equation (6.4.32) does not tell us which is which. However, when dealing with general *homogeneous* equations, which are those of the form

$$Ax^2 + Bxy + Cy^2 = F = 0$$

(in which $D = E = 0$), this information sufficient for determining the *precise shape* of our curve. What happens in this special case is the fact that the reduced equation will also have $D_{\text{new}} = E_{\text{new}} = 0$, so the new equation will in fact look like

$$A_{\text{new}}u^2 + C_{\text{new}}v^2 + F = 0.$$

Example 6.4.6. Suppose we are asked to find the *precise shape* of the curve given by the equation

$$x^2 + 6xy + y^2 - 8 = 0.$$

Our coefficients are $A = 1$, $B = 6$ and $C = 1$, so the discriminant is: $\Delta = 1 \cdot 1 - \frac{6^2}{4} = 1 - 9 = -8$. The characteristic equation is then:

$$\lambda^2 - 2\lambda - 8 = 0,$$

which, using the quadratic formula, has solutions $\lambda_1 = 4$ and $\lambda_2 = -2$. Since we know that these solutions must be A_{new} and C_{new} (but we do not know which is which), this tells us the fact that, whenever we reduce our equation to a standard form, it will have one of the following forms:

$$4u^2 - 2v^2 - 8 = 0 \ \text{ (in the case when } A_{\text{new}} = 4 \text{ and } C_{\text{new}} = -2\text{);} \tag{6.4.35}$$

$$-2u^2 + 4v^2 - 8 = 0 \ \text{ (in the case when } A_{\text{new}} = -2 \text{ and } C_{\text{new}} = 4\text{).} \tag{6.4.36}$$

Upon adding 8 to both sides, then dividing everything by 8, equation (6.4.35) can be re-written as

$$\frac{u^2}{2} - \frac{v^2}{4} = 1. \tag{6.4.37}$$

Likewise, equation (6.4.36) can be re-written as

$$\frac{v^2}{2} - \frac{u^2}{4} = 1. \tag{6.4.38}$$

In either case, the curve is a *hyperbola, with major radius $\sqrt{2}$ and minor radius 2.*

Exercises

In each one of the Exercises 1-4 a new coordinate system is constructed, with a specified origin and a specified frame, and you are asked to find the new coordinates of a point $P(x, y)$ indicated in standard coordinates.

1. $O(1,1)$; $\vec{\mathbf{e}} = \begin{bmatrix} 1/2 \\ \sqrt{3}/2 \end{bmatrix}$, $\vec{\mathbf{f}} = \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}$. $P(2,-2)$.

2. Same coordinate system as in Exercise 1; $P(4,0)$.

3. $O(2,2)$; $\vec{\mathbf{e}} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$, $\vec{\mathbf{f}} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. $P(-4,5)$.

4. Same coordinate system as in Exercise 3; $P(2\sqrt{2},0)$.

In each one of the Exercises 5-8 use the given new coordinate system, and write the equation of a curve $\mathscr{C}$ which is presented in standard coordinates by the given equation.

5. Same coordinate system as in Exercise 1; curve $\mathscr{C}$ given by the equation
$$3x + 2y = 3.$$

6. Same coordinate system as in Exercise 1; curve $\mathscr{C}$ given by the equation
$$x^2 + y^2 = 4.$$

7. Same coordinate system as in Exercise 4; curve $\mathscr{C}$ given by the equation
$$(x + y - 4)^3 = 3(x - y).$$

8. Same coordinate system as in Exercise 4; curve $\mathscr{C}$ given by the equation
$$x + y = \cos(x - y).$$

In Exercises 9-12 you are asked to determine the general shape (parabola, ellipse, or hyperbola, possibly degenerate) of the curve represented by the given equation. The discriminant provides all the necessary information. (See Example 6.4.5.)

9. $x^2 + 2y^2 + 3xy = 1$.

10. $6x^2 - 10xy - 3y^2 + 6y = 7$.

11. $4x^2 + 4xy + y^2 - 7x + 8y = 2$.

12. $x^2 + 3xy = 2$.

In Exercises 13-16 you are asked to determine the precise shape of the curve represented by the given homogeneous equation. Use Example 6.4.6 as a guide.

13. $x^2 - 2y^2 + 3xy = 1$.

14. $6x^2 - 10xy - 3y^2 = 10$.

15. $4x^2 + 4xy + y^2 = 2$.

16. $x^2 + 3xy = 2$.

In Exercises 17-19 you are asked to to the following.

(i) Reduce the equation to one in the standard form (with $B_{\text{new}} = 0$) by finding a suitable coordinate system, which amounts to solving the system (6.4.17), which you can do either trigonometrically or algebraically. Once you find the new coordinate system, write down (carefully!) the reduced equation, using formulas (6.4.14).

(ii) Use square completions (one the reduced equation) to obtain all geometric information about the given curve:

(a) If the curve is degenerate, indicate so.

(b) If the curve is a parabola, find the vertex and the focal distance.

(c) If the curve is an ellipse, or a hyperbola find its center, vertices, and foci.

In cases (b) and (c), after you first locate your points in (u, v)-coordinates, you must specify them in (x, y)-coordinates.

HINT: Follow the method outlined in Example 6.4.4. In particular, as a visual aid, sketch the graph of the given equation on a calculator, by solving the given equation for y, that is, by deriving an equivalent form

$$y = \textit{expression(s) in } x,$$

using the Quadratic Formula.

17*. $11x^2 - 24xy + 4y^2 + 20 = 0$.

18*. $x^2 + 2xy + y^2 + x - y + 4 = 0$

19*. $13x^2 - 8xy + 7y^2 = 10$.

6.5 Conics in Polar Coordinates

In this section we explore the polar equations for certain conics (those that *have a focus at the pole*). The reason we limit ourselves to these special types of conics is the fact that the polar equations are particularly nice.

In preparation for the derivation of these equations, we will first derive some interesting geometric result (the Focus-Directrix Geometric Presentation), then we will use it to derive a general single type equation (the Focus-Directrix Equation), and only after that we will turn our attention to polar equations.

The derivation of the polar equation (6.5.24) is quite long and technical. A reader who only wants to know the equation (without a proof), should skip directly to the main statement from page 296, and start from there.

Focus-Directrix Geometric Presentation of Conics

We are about to begin a (long) story, which at the end will tells us that all conics, whether they are parabolas, ellipses, or hyperbolas, can be presented by a single type of an equation. In understanding how this equation comes about, let us first "play" with a conic $\mathscr{C}$ which is either an ellipse or a hyperbola with horizontal focal axis, and has has center at the origin. From Sections 6.2 and 6.3 we know that $\mathscr{C}$ is presented by an equation of the following form:

I. If $\mathscr{C}$ is an ellipse: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, with $a > b > 0$;

II. If $\mathscr{C}$ is a hyperbola: $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$, with $a, b > 0$.

Each one of these equations can in fact be written in the form

$$y^2 = \mathsf{P}(x), \tag{6.5.1}$$

where $\mathsf{P}(x)$ is a second degree polynomial. More specifically, in each of the two cases we consider we do the following:

I. If $\mathscr{C}$ is an ellipse, we multiply everything by b^2, so we get $y^2 + \dfrac{b^2}{a^2}x^2 = b^2$, so when we subtract the x-term, we get:

$$y^2 = -\frac{b^2}{a^2}x^2 + b^2. \tag{6.5.2}$$

II. If $\mathscr{C}$ is a hyperbola, we multiply everything by $-b^2$, so we get $y^2 - \dfrac{b^2}{a^2}x^2 = -b^2$, so when we subtract the x-term, we get:

$$y^2 = \frac{b^2}{a^2}x^2 - b^2. \tag{6.5.3}$$

We also know that the numbers a (the major radius) and b (the minor radius) are linked using the *eccentricity* $\boldsymbol{e}$ of $\mathscr{C}$ by the formula

$$b^2 = \begin{cases} (1 - \boldsymbol{e}^2)a^2, \text{ if } \mathscr{C} \text{ is an ellipse} \\ (\boldsymbol{e}^2 - 1)a^2, \text{ if } \mathscr{C} \text{ is a hyperbola} \end{cases}$$

and either equation (6.5.2) or (6.5.3) can be written in the form:

$$y^2 = (\boldsymbol{e}^2 - 1)(x^2 - a^2). \tag{6.5.4}$$

It turns out that the polynomial $\mathsf{P}(x) = (\boldsymbol{e}^2 - 1)(x^2 - a^2)$ that appears in the right-hand side of (6.5.4) can be written as a *difference of squares* of the form

$$\mathsf{P}(x) = \boldsymbol{e}^2(x - \delta)^2 - (x - \mu)^2 \tag{6.5.5}$$

in exactly two ways:

$$\begin{aligned} \mathsf{P}(x) &= (\boldsymbol{e}^2 - 1)(x^2 - a^2) = \\ &= \boldsymbol{e}^2\left(x - \frac{a}{\boldsymbol{e}}\right)^2 - (x - a\boldsymbol{e})^2 = \boldsymbol{e}^2\left(x - \frac{a}{\boldsymbol{e}}\right)^2 - (x - c)^2 = \qquad (6.5.6) \\ &= \boldsymbol{e}^2\left(x + \frac{a}{\boldsymbol{e}}\right)^2 - (x + a\boldsymbol{e})^2 = \boldsymbol{e}^2\left(x + \frac{a}{\boldsymbol{e}}\right)^2 - (x + c)^2, \qquad (6.5.7) \end{aligned}$$

where $c = a\boldsymbol{e}$ is the *focal distance*.

Having presented $\mathsf{P}(x)$ as in (6.5.5), we can re-write the equation of $\mathscr{C}$ in the form

$$y^2 = \boldsymbol{e}^2(x - \delta)^2 - (x - \mu)^2,$$

so after adding $(x-\mu)^2$ to both sides we can also write

$$(x-\mu)^2 + y^2 = \boldsymbol{e}^2(x-\delta)^2. \tag{6.5.8}$$

As pointed out above, there are exactly two pairs that allow $\mathscr{C}$ to be presented as in (6.5.8), namely:

$$\begin{cases} \mu_1 = c \\ \delta_1 = \dfrac{a}{\boldsymbol{e}} \end{cases} \quad \text{and} \quad \begin{cases} \mu_2 = -c \\ \delta_2 = -\dfrac{a}{\boldsymbol{e}} \end{cases}$$

In order to avoid any complication, we have to exclude the case when $\boldsymbol{e} = 0$. As we learned in Section 6.2, this case corresponds to *circles*. Even though one can think of circles as exceptional ellipses, in order to exclude them from our discussion, we will use the phrase *"honest" ellipses* to describe ellipses that are not circles.

Upon taking square roots[16] in (6.5.8), we can re-write that equation as:

$$\sqrt{(x-\mu)^2 + y^2} = \boldsymbol{e}\big|x-\delta\big|, \tag{6.5.9}$$

The nice feature of (6.5.9) is the fact that it can be interpreted geometrically. Since the μ's are the x-coordinates of the *foci* of $\mathscr{C}$, so the foci are $F_1(\mu_1, 0)$ and $F_2(\mu_2, 0)$, we can identify the left-hand side as the distance $\text{dist}(P, F)$ from $P(x, y)$ to a focus F. As is turns out, the right-hand side of (6.5.9) can also be written down using distances. After all, if we consider the vertical line $\mathscr{D}$ given by the equation $x = \delta$, then $\big|x-\delta\big| = \text{dist}(P, \mathscr{D})$. So now equation (6.5.9) looks like

$$\text{dist}(P, F) = \boldsymbol{e} \cdot \text{dist}(P, \mathscr{D}). \tag{6.5.10}$$

Since we have in fact two foci $F_1(\mu_1, 0)$ and $F_2(\mu_2, 0)$, will will have two lines $\mathscr{D}_1$: $x = \delta_1$ and $\mathscr{D}_2$: $x = \delta_2$. We call these lines the *directrices of* $\mathscr{C}$. To be more accurate, we match these driectrices with their corresponding foci, so we call

- $\mathscr{D}_1$ the *directrix of* $\mathscr{C}$ *associated with the focus* F_1, and
- $\mathscr{D}_2$ the *directrix of* $\mathscr{C}$ *associated with the focus* F_2.

Our findings up to this point are summarized as follows: *If we pick F to be one of the foci of $\mathscr{C}$, and we take $\mathscr{D}$ to be its associated directrix, then $\mathscr{C}$ is the set of all points P that satisfy the geometric condition* (6.5.10)

CLARIFICATIONS. The geometric condition (6.5.10) is very reminiscent of the *geometric definition of a parabola*(!) So it is convenient for us *define the eccentricity of a parabola to be* $\boldsymbol{e} = 1$. Using this convention, the general shape of a conic is completely determined by its eccentricity, so we can simply say that:

(A) a *hyperbola* is a conic with eccentricity $\boldsymbol{e} > 1$;
(B) a *parabola* is a conic with eccentricity $\boldsymbol{e} = 1$;
(C) an *"honest" ellipse* is a conic with eccentricity $0 < \boldsymbol{e} < 1$.
(D) a *circle* is a conic with eccentricity $\boldsymbol{e} = 0$.

When it comes to foci and directrices, the particular features of these four cases are as follows.

[16] When we take the square root in the right-hand side, we use the formula $\sqrt{\heartsuit^2} = |\heartsuit|$.

I. Hyperbolas and "honest" ellipses have *two foci and two directrices*, and each directrix is associated to a focus.
II. Parabolas have *one focus and one directrix*.
III. Circles have *one focus (the center) and no directrix*.

With this set-up, our preceding conclusion can be expanded as follows.

Focus-Directrix Geometric Presentation of Non-Degenerate Conics

If $\mathscr{C}$ is a non-degenerate conic with eccentricity $\boldsymbol{e} > 0$, then for each focus F of $\mathscr{C}$, there is a unique line $\mathscr{D}$ (the directrix of $\mathscr{C}$ associated with F), which allows one to present $\mathscr{C}$ as the set of all points P in the plane, that satisfy:

$$\mathrm{dist}(P, F) = \boldsymbol{e} \cdot \mathrm{dist}(P, \mathscr{D})$$

Furthermore, with any choice of F, the associated directrix $\mathscr{D}$ is always perpendicular to the focal axis of $\mathscr{C}$.

CLARIFICATION. Strictly speaking, in the case when $\boldsymbol{e} \neq 1$ we justified the above statements only in the case when $\mathscr{C}$ has center at the origin and has horizontal focal axis. However, if we are given an arbitrary conic $\mathscr{C}$ which is either a hyperbola or an "honest" ellipse, then using what we learned in Section 6.4, we can always devise a coordinate system (u, v) so that

(i) the origin is at the center of $\mathscr{C}$, and
(ii) the focal axis of $\mathscr{C}$ coincides with the u-axis.

When we use such a coordinate system, our conic $\mathscr{C}$ will look precisely like the particular type we treated above.

Focus-Directrix Equations of Conics

Our goal now is to turn the geometric presentation (6.5.10) into an algebraic equation. For this purpose, we are going to use the following Line Distance Formula which we proved in Section 2.1:

Given a line $\mathscr{L}$ represented by the general linear equation

$$mx + ny = q,$$

and a point $P(x_0, y_0)$, the distance from P to $\mathscr{L}$ is given by the formula:

$$\mathrm{dist}(P, \mathscr{L}) = \frac{|mx_0 + ny_0 - q|}{\sqrt{m^2 + n^2}},$$

Using this formula back in (6.5.10) and taking squares, we derive the following important unified equation for all conics.

Focus-Directrix Equations of Non-Degenerate Conics

Assume $\mathscr{C}$ is a non-degenerate conic with eccentricity $\boldsymbol{e} > 0$, and $F(h, k)$ is a focus of $\mathscr{C}$. If the directrix $\mathscr{D}$ of $\mathscr{C}$, which is associated with F, is given by the equation $mx + ny = q$, then $\mathscr{C}$ can be presented by the equation

$$(x - h)^2 + (y - k)^2 = \frac{\boldsymbol{e}^2}{m^2 + n^2}\left(mx + ny - q\right)^2, \tag{6.5.11}$$

and furthermore, the vector $\vec{\mathbf{v}} = \begin{bmatrix} m \\ n \end{bmatrix}$ *is a direction vector for the focal axis of* $\mathscr{C}$.

Example 6.5.1. Suppose a conic $\mathscr{C}$ is presented by the equation

$$(x-6)^2 + (y+8)^2 = \frac{4}{25}(3x-4y-25)^2,$$

and we want to find all its geometric elements: exact shape, vertices, foci, and center, .

We start off by matching this equation with the Focal-Directrix Equation (6.5.11), so we can match

$$\begin{cases} (x-6) \text{ matches } (x-h)\text{, thus: } h=6; \\ (y+8) \text{ matches } (y-k)\text{, thus: } k=-8; \\ (3x-4y-25) \text{ matches } (mx+ny-q)\text{, thus: } m=3,\ n=-4,\ q=25; \\ \\ \dfrac{4}{25} \text{ matches } \dfrac{\boldsymbol{e}^2}{m^2+n^2}\text{, thus } \dfrac{\boldsymbol{e}^2}{25} = \dfrac{4}{25}\text{, which yields: } \boldsymbol{e}^2 = 4\text{, so: } \boldsymbol{e} = 2 \end{cases}$$

Based on the features of the Focal-Directrix Equation, we can draw the following conclusions:

(i) The conic $\mathscr{C}$ is a *hyperbola*, with eccentricity $\boldsymbol{e} = 2$.

(ii) One focus of $\mathscr{C}$ is the point $F_1(6,-8)$.

(iii) The directrix $\mathscr{D}_1$ of $\mathscr{C}$, associated with the focus F_1 is the line given by the equation:

$$3x - 4y = 25. \tag{6.5.12}$$

(iv) The focal axis $\mathscr{F}$ of $\mathscr{C}$ is the *line* $\mathscr{L}$ *that passes through* F_1 *and is perpendicular to* $\mathscr{D}_1$. Based on what we learned in Section 2.1, this line can be presented by an equation of the form $4x + 3y =$ *number*, so if we plug in the focus ($x = 6$ and $y = -8$), we conclude that the focal axis of $\mathscr{C}$ is given by the equation:

$$4x + 3y = 0. \tag{6.5.13}$$

Using the information collected thus far, we can now find the *vertices* of $\mathscr{C}$, which are precisely the *points where* $\mathscr{C}$ *intersects the focal axis*. So in order to find the vertices, all we have to do is to solve the *system of equations made of the equation of* $\mathscr{C}$ *and the equation of the focal axis*:

$$\begin{cases} (x-6)^2 + (y+8)^2 = \dfrac{4}{25}(3x-4y-25)^2, \\ \\ 4x+3y=0. \end{cases} \tag{6.5.14}$$

Using the second equation as a substitution $y = -\dfrac{4}{3}x$, the first equation becomes:

$$(x-6)^2 + \left(-\frac{4}{3}x + 8\right)^2 = \frac{4}{25}\left(3x + \frac{16}{3}x - 25\right)^2 = \frac{4}{25}\left(\frac{25}{3}x - 25\right)^2,$$

which by expanding all squares yields:

$$x^2 - 12x + 36 + \frac{16}{9}x^2 - \frac{64}{3}x + 64 = \frac{4}{25}\left(\frac{625}{9}x^2 - \frac{1250}{3}x + 625\right) = \frac{100}{9}x^2 - \frac{200}{3}x + 100.$$

After moving all terms on one side, this equation reads:

$$x^2 - 12x + 36 + \frac{16}{9}x^2 - \frac{64}{3}x + 64 - \frac{100}{9}x^2 + \frac{200}{3}x - 100 = 0.$$

By combining the like terms, we finally get

$$-\frac{25}{3}x^2 + \frac{100}{3}x = 0,$$

which clearly has two solutions $x_1 = 4$ and $x_2 = 0$. Using our substitution $y = -\frac{4}{3}x$, it follows that the solutions of (6.5.14) are:

$$\begin{cases} x_1 = 4 \\ y_1 = -\dfrac{16}{3} \end{cases} \quad \text{and} \quad \begin{cases} x_2 = 0 \\ y_2 = 0 \end{cases}$$

Therefore, we now can conclude that the vertices of $\mathscr{C}$ are $V_1\big(4, -\frac{16}{3}\big)$ and $V_2(0,0)$. Using the coordinates of the vertices, we can locate the *center* Z of $\mathscr{C}$ as the *midpoint of the segment formed by the vertices*, so the coordinates of Z are:

$$\begin{aligned} x_Z &= \frac{1}{2}\big(x_1 + x_2\big) = \frac{1}{2}\big(4 + 0\big) = 2; \\ y_Z &= \frac{1}{2}\big(y_1 + y_2\big) = \frac{1}{2}\big(-\frac{16}{3} + 0\big) = -\frac{8}{3}. \end{aligned}$$

As for the second focus $F_2(h_2.k_2)$, we can use the given focus $F(6,-8)$ and the fact that $Z\big(2, -\frac{8}{3}\big)$ is also the *midpoint of the segment formed by the foci*, which gives us the equalities

$$\begin{cases} \dfrac{1}{2}\big(6 + h_2\big) = 2 \\ \dfrac{1}{2}\big(-8 + k_2\big) = -\dfrac{8}{3} \end{cases}$$

and we immediately get $h_2 = -2$ and $k_2 = \frac{8}{3}$.

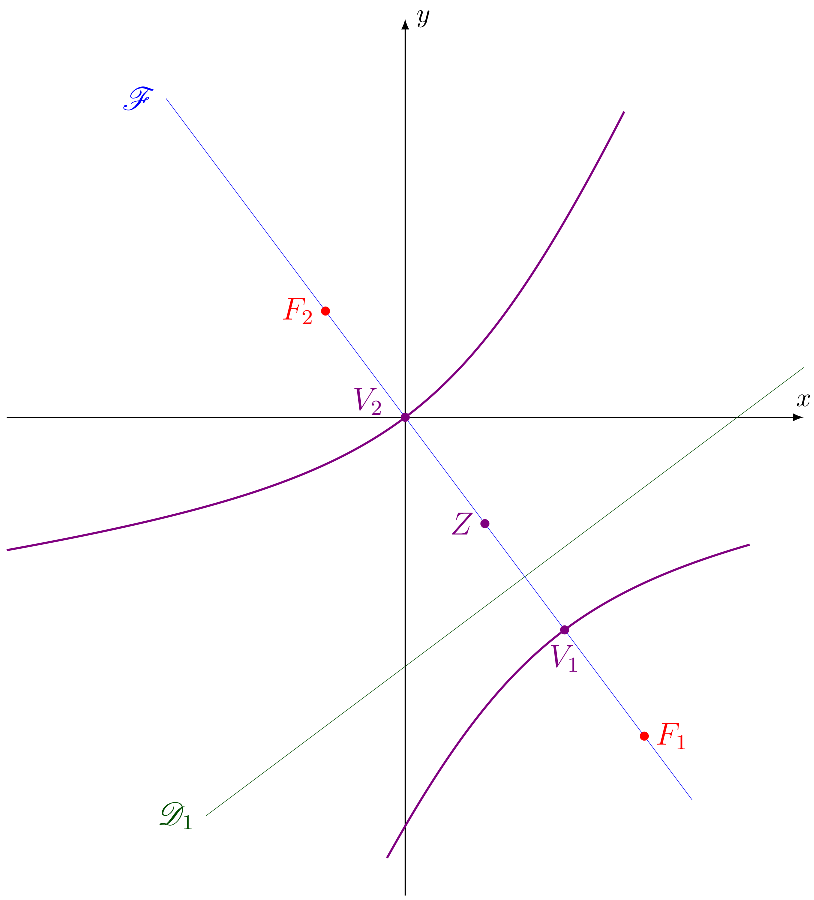


Figure 6.5.1

Finally, using the distance formula, we can compute the major radius a and the focal distance c of our hyperbola as:

$$a = \text{dist}(Z, V_1) = \sqrt{(x_Z - x_1)^2 + (y_Z - y_1)^2} = \sqrt{(2-0)^2 + \left(-\frac{8}{3} - 0\right)^2} =$$
$$= \sqrt{4 + \frac{64}{9}} = \sqrt{\frac{100}{9}} = \frac{10}{3};$$
$$c = \text{dist}(Z, F_1) = \sqrt{(x_Z - h_1)^2 + (y_Z - k_1)^2} = \sqrt{(2-6)^2 + \left(-\frac{8}{3} - (-8)\right)^2} =$$
$$= \sqrt{16 + \frac{256}{9}} = \sqrt{\frac{400}{9}} = \frac{20}{3}.$$

(We can use this calculation to confirm that our calculations are right, because we know that $\boldsymbol{e} = c/a$.) Finally, since $\mathscr{C}$ is a hyperbola, its minor radius is:

$$b = \sqrt{c^2 - a^2} = \sqrt{\left(\frac{20}{3}\right)^2 - \left(\frac{10}{3}\right)^2} = \sqrt{\frac{300}{9}} = \frac{10\sqrt{3}}{3}.$$

The Polar Equations of Conics

For the remainder of this section, we will limit ourselves to *conics which have one focus at the origin*.

Assume we have such a conic $\mathscr{C}$, with eccentricity $\boldsymbol{e} > 0$, and let $\mathscr{D}$ be the directrix of $\mathscr{C}$ associated with the focus that is at the origin. If we assume $\mathscr{D}$ is given by the equation

$$mx + ny = q, \tag{6.5.15}$$

then using the Focus-Directrix Equation, we can present our conic $\mathscr{C}$ by the equation

$$x^2 + y^2 = \frac{\boldsymbol{e}^2}{m^2 + n^2}\left(mx + ny - q\right)^2. \tag{6.5.16}$$

We are almost ready to transform this into a polar equation, but before we do that, we first re-write the right-hand side as a square, so (6.5.16) becomes

$$x^2 + y^2 = \left[\frac{\boldsymbol{e}}{\sqrt{m^2 + n^2}}(mx + ny - q)\right]^2 = \left[\frac{\boldsymbol{e}}{\sqrt{m^2 + n^2}}(q - mx - ny)\right]^2$$

and then we can eliminate the square by writing:

$$\frac{\boldsymbol{e}}{\sqrt{m^2 + n^2}}(q - mx - ny) = \pm\sqrt{x^2 + y^2}. \tag{6.5.17}$$

Using now what we learned in Section 3.3, we can convert this equation to *polar coordinates*, by replacing $x = r\cos\,\theta$, $y = r\sin\,\theta$, and $\pm\sqrt{x^2 + y^2} = r$, so (6.5.17) becomes:

$$\frac{\boldsymbol{e}}{\sqrt{m^2 + n^2}}\left[q - r(m\,\cos\,\theta + n\,\sin\,\theta)\right] = r. \tag{6.5.18}$$

When we expand the left-hand side, this equation reads:

$$\frac{\boldsymbol{e}q}{\sqrt{m^2 + n^2}} - r\left(\frac{\boldsymbol{e}m}{\sqrt{m^2 + n^2}}\cos\,\theta + \frac{\boldsymbol{e}n}{\sqrt{m^2 + n^2}}\sin\,\theta\right) = r, \tag{6.5.19}$$

so when we add the term $r\left(\ \dots\ \right)$ to both sides, our equation becomes:

$$\frac{\boldsymbol{e}q}{\sqrt{m^2 + n^2}} = r\left(1 + \frac{\boldsymbol{e}m}{\sqrt{m^2 + n^2}}\cos\,\theta + \frac{\boldsymbol{e}n}{\sqrt{m^2 + n^2}}\sin\,\theta\right) \tag{6.5.20}$$

We "clean up" our calculation a little bit, as follows: denote $\dfrac{\boldsymbol{e}q}{\sqrt{m^2 + n^2}}$ by $\boldsymbol{\kappa}$; denote $\dfrac{\boldsymbol{e}m}{\sqrt{m^2 + n^2}}$ by u; and denote $\dfrac{\boldsymbol{e}n}{\sqrt{m^2 + n^2}}$ by v. With these notations, the equation (6.5.20) now reads:

$$\boldsymbol{\kappa} = r(1 + u\,\cos\,\theta + v\,\sin\,\theta) \tag{6.5.21}$$

Let us now point out two additional features of the three new numbers we introduced.

I. When we multiply all terms in the directrix equation (6.5.15) by $\dfrac{\boldsymbol{e}}{\sqrt{m^2+n^2}}$, we see that we get a new (equivalent) equation for the directrix $\mathscr{D}$, which looks exactly like:

$$ux + vy = \boldsymbol{\kappa}, \tag{6.5.22}$$

II. When we add the squares of u and v we get

$$u^2+v^2 = \left(\frac{\boldsymbol{e}m}{\sqrt{m^2+n^2}}\right)^2 + \left(\frac{\boldsymbol{e}n}{\sqrt{m^2+n^2}}\right)^2 = \frac{\boldsymbol{e}^2m^2}{m^2+n^2} + \frac{\boldsymbol{e}^2n^2}{m^2+n^2} = \\ = \frac{\boldsymbol{e}^2m^2+\boldsymbol{e}^2n^2}{m^2+n^2} = \frac{\boldsymbol{e}^2(m^2+n^2)}{m^2+n^2} = \boldsymbol{e}^2,$$

which means that the eccentricity is simply given as

$$\boldsymbol{e} = \sqrt{u^2+v^2} \tag{6.5.23}$$

When we finally look at (6.5.21), we can certainly solve it r to reach the equation (6.5.24) given in the summary below.

Polar Equation for Conics with one Focus at the Origin

Given constants $\boldsymbol{\kappa} \neq 0$ *and* u, v *arbitrary, a polar equation of the form*

$$r = \frac{\boldsymbol{\kappa}}{1 + u\cos\theta + v\sin\theta}, \tag{6.5.24}$$

represents a conic $\mathscr{C}$*, with one focus at the origin, and eccentricity* $\boldsymbol{e} = \sqrt{u^2+v^2}$*. Furthermore, if* $\boldsymbol{e} \neq 0$*, then the directrix of* $\mathscr{C}$*, that is associated to the focus at the origin, is given by the equation:*

$$ux + vy = \boldsymbol{\kappa},$$

Conversely, given a line $\mathscr{D}$ *represented by the equation* $mx + ny = q$*, with* $q \neq 0$ *(so that* $\mathscr{D}$ *does not pass through the origin), and some constant* $\boldsymbol{e} > 0$*, the conic* $\mathscr{C}$ *with eccentricity* $\boldsymbol{e}$*, with focus at the origin, and with* $\mathscr{D}$ *as the directrix associated to the focus at the origin, is given by the equation* (6.5.24)*, where* $\boldsymbol{\kappa} = \dfrac{\boldsymbol{e}q}{\sqrt{m^2+n^2}}$*,* $u = \dfrac{\boldsymbol{e}m}{\sqrt{m^2+n^2}}$*, and* $v = \dfrac{\boldsymbol{e}n}{\sqrt{m^2+n^2}}$.

ADDITIONAL CLARIFICATIONS. There is quite a bit we can add to the first statement, concerning equations of the form (6.5.24) above, especially when $\boldsymbol{e} > 0$. Using what we learned from Section 4.1, one part of the denominator from (6.5.24) can also be presented as

$$u\cos\theta + v\cos\theta = \sqrt{u^2+v^2}\cos(\theta-\phi),$$

where ϕ is any angle that satisfies the identities

$$\begin{cases} \cos\phi = \dfrac{u}{\sqrt{u^2+v^2}} \\[2ex] \sin\phi = \dfrac{v}{\sqrt{u^2+v^2}} \end{cases} \tag{6.5.25}$$

Using such an angle, and the eccentricity formula $\boldsymbol{e} = \sqrt{u^2 + v^2}$, the equation (6.5.24) can also be presented as:

$$r = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\,\cos(\theta - \phi)}. \tag{6.5.26}$$

Furthermore, since by construction we have

$$\begin{cases} u = \boldsymbol{e}\,\cos\,\phi \\ v = \boldsymbol{e}\,\cos\,\phi \end{cases} \tag{6.5.27}$$

and the equation of $\mathscr{D}$ looks like

$$(\boldsymbol{e}\,\cos\,\phi)x + (\boldsymbol{e}\,\sin\,\phi)y = \boldsymbol{\kappa},$$

it follows that the vector $\overrightarrow{\mathrm{w}} = \begin{bmatrix} \cos\,\phi \\ \sin\,\phi \end{bmatrix}$ is *perpendicular to* $\mathscr{D}$, which means that $\overrightarrow{\mathrm{w}}$ *points in the direction of the focal axis of* $\mathscr{C}$. In particular, *all points lying on the focal axis of* $\mathscr{C}$ *must have polar coordinate representations of the form* (r, ϕ) *or* $(r, \phi + \pi)$. Since the *vertices* of $\mathscr{C}$ are precisely the *point(s) where* $\mathscr{C}$ *intersects its focal axis*, it follows that the vertices V_1 and V_2 of $\mathscr{C}$ can be presented in polar coordinates by pairs (r_1, θ_1) and (r_2, θ_2), which are built using the equation (6.5.24) – or its equivalent form (6.5.26) – by plugging in the values $\theta_1 = \phi$ and $\theta_2 = \phi + \pi$. Using the equation (6.5.26), the corresponding values of r are:

$$r_1 = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\,\cos(\theta_1 - \phi)} = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\,\cos\,0} = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\cdot 1} = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}}, \tag{6.5.28}$$

$$r_2 = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\,\cos(\theta_2 - \phi)} = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\,\cos\,\pi} = \frac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}\cdot(-1)} = \frac{\boldsymbol{\kappa}}{1 - \boldsymbol{e}}, \tag{6.5.29}$$

and then our vertices $V_1(x_1, y_1)$ and $V_2(x_2, y_2)$ will have rectangular coordinates[17]

$$V_1 : \begin{cases} x_1 = r_1 \cos\,\theta_1 = \dfrac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}} \cdot \cos\,\phi = \dfrac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}} \cdot \dfrac{u}{\boldsymbol{e}} = \dfrac{\boldsymbol{\kappa} u}{(1 + \boldsymbol{e})\boldsymbol{e}} \\[2ex] y_1 = r_1 \sin\,\theta_1 = \dfrac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}} \cdot \sin\,\phi = \dfrac{\boldsymbol{\kappa}}{1 + \boldsymbol{e}} \cdot \dfrac{v}{\boldsymbol{e}} = \dfrac{\boldsymbol{\kappa} v}{(1 + \boldsymbol{e})\boldsymbol{e}} \end{cases} \tag{6.5.30}$$

$$V_2 : \begin{cases} x_2 = r_2 \cos\,\theta_2 = -r_2 \cos\,\phi = -\dfrac{\boldsymbol{\kappa}}{1 - \boldsymbol{e}} \cdot \dfrac{u}{\boldsymbol{e}} = \dfrac{\boldsymbol{\kappa} u}{(\boldsymbol{e} - 1)\boldsymbol{e}} \\[2ex] y_2 = r_2 \sin\,\theta_2 = -r_2 \sin\,\phi = -\dfrac{\boldsymbol{\kappa}}{1 - \boldsymbol{e}} \cdot \dfrac{v}{\boldsymbol{e}} = \dfrac{\boldsymbol{\kappa} v}{(\boldsymbol{e} - 1)\boldsymbol{e}} \end{cases} \tag{6.5.31}$$

Of course, in the case when $\boldsymbol{e} = 1$, the second vertex V_2 does not exist. This is not surprising, because in this case we are dealing with a *parabola*, which has only one vertex!

Another important point that can be picked up from the equation (6.5.24) is the *point* Q *where the focal axis intersects the directrix* $\mathscr{D}$. On the one hand, since this point is on the focal axis, it

[17] For the formulas (6.5.31) we use the identities $\cos(\phi + \pi) = -\cos\,\phi$ and $\sin(\phi + \pi) = -\sin\,\phi$.

follows that the vector $\overrightarrow{OQ} = \begin{bmatrix} x_Q \\ y_Q \end{bmatrix}$ must be a multiple of the vector $\vec{w} = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$, so there is some constant ρ, so that

$$\begin{cases} x_Q = \rho\cos\phi = \rho\dfrac{u}{\boldsymbol{e}} \\ \\ y_Q = \rho\sin\phi = \rho\dfrac{v}{\boldsymbol{e}} \end{cases} \tag{6.5.32}$$

On the other hand, since Q also sits on $\mathscr{D}$, it must satisfy the equation $ux_Q + uy_Q = \boldsymbol{\kappa}$, so we must have

$$\rho\left(\frac{u^2}{\boldsymbol{e}} + \frac{v^2}{\boldsymbol{e}}\right) = \boldsymbol{\kappa},$$

which reads

$$\rho \cdot \frac{u^2 + v^2}{\boldsymbol{e}} = \boldsymbol{\kappa},$$

and then using the eccentricity formula $u^2 + v^2 = \boldsymbol{e}^2$, this simplifies to

$$\rho\boldsymbol{e} = \boldsymbol{\kappa},$$

so now we get $\rho = \dfrac{\boldsymbol{\kappa}}{\boldsymbol{e}}$. Finally when we go back to (6.5.32), it follows that our special point has

$$Q: \quad \begin{cases} x_Q = \dfrac{\boldsymbol{\kappa}u}{\boldsymbol{e}^2} = \dfrac{\boldsymbol{\kappa}u}{u^2+v^2} \\ \\ y_Q = \dfrac{\boldsymbol{\kappa}v}{\boldsymbol{e}^2} = \dfrac{\boldsymbol{\kappa}v}{u^2+v^2} \end{cases}$$

Example 6.5.2. Suppose we are given the polar equation

$$r = \frac{4}{1 + 3\cos\theta},$$

and we are asked to describe the geometry of the curve $\mathscr{C}$ represented by this equation, as completely as possible, as we did in Example 6.5.1.

As it turns out, we can perfectly match the given equation with (6.5.24), by setting $\boldsymbol{\kappa} = 4$, $u = 3$ and $v = 0$. Using the eccentricity formula, we get

$$\boldsymbol{e} = \sqrt{u^2 + v^2} = \sqrt{3^2 + 0^2} = \sqrt{9} = 3,$$

so our conic $\mathscr{C}$ is a *hyperbola* with eccentricity 3, and one focus F_1 at the origin: $F_1(0,0)$. The vertices of $\mathscr{C}$ are obtained using formulas (6.5.30) and (6.5.31), which yield:

$$V_1: \quad \begin{cases} x_1 = \dfrac{\boldsymbol{\kappa}u}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{4\cdot 3}{(1+3)\cdot 3} = 1 \\ \\ y_1 = \dfrac{\boldsymbol{\kappa}v}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{4\cdot 0}{(1+3)\cdot 3} = 0 \end{cases}$$

$$V_2: \quad \begin{cases} x_2 = \dfrac{\boldsymbol{\kappa}u}{(\boldsymbol{e}-1)\boldsymbol{e}} = \dfrac{4\cdot 3}{(3-1)\cdot 3} = 2 \\ \\ y_2 = \dfrac{\boldsymbol{\kappa}u}{(\boldsymbol{e}-1)\boldsymbol{e}} = \dfrac{4\cdot 0}{(3-1)\cdot 3} = 0 \end{cases}$$

Using midpoints (exactly as in Example 6.5.1, it follows that the center has coordinates $Z\left(\frac{3}{2}, 0\right)$, and the second focus is at $F_2(3, 0)$.

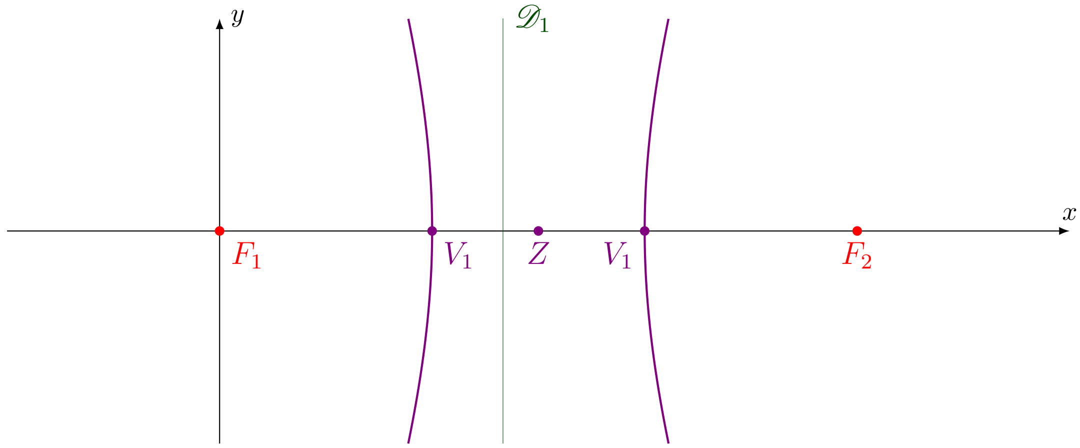


Figure 6.5.2

The focal axis of our hyperbola is the x-axis. The directrix $\mathscr{D}_1$ associated with the focus F_1 is given by the equation (6.5.22), which in our case becomes: $3x = 4$. In other words, $\mathscr{D}_1$ is the vertical line:

$$x = \frac{4}{3}.$$

The major radius a is the distance between Z and V_1, so by direct measurement we get $a = \frac{1}{2}$. The focal distance c can either be computed using the formula $c = a\boldsymbol{e}$ or by measuring the distance between Z and F_1; either method yields the same value: $c = \frac{3}{2}$. Finally, since our curve is a hyperbola, the minor radius is $b = \sqrt{c^2 - a^2} = \sqrt{\left(\frac{3}{2}\right)^2 - \left(\frac{1}{2}\right)^2} = \sqrt{\frac{9}{4} - \frac{1}{4}} = \sqrt{2}$.

Example 6.5.3. Suppose we are given the polar equation

$$r = \frac{4}{2 - \sin\theta},$$

and we are asked to describe the geometry of the curve $\mathscr{C}$ represented by this equation, as completely as possible, as we did above.

In order to match this equation with (6.5.24), we must force a factoring in the denominator, so we will re-write our equation as

$$r = \frac{4}{2\left(1 - \frac{1}{2}\sin\theta\right)} = \frac{2}{1 - \frac{1}{2}\sin\theta},$$

and now we have a perfect match with (6.5.24) by setting $\boldsymbol{\kappa} = 2$, $u = 0$ and $v = -\frac{1}{2}$. Using the eccentricity formula, we get

$$\boldsymbol{e} = \sqrt{u^2 + v^2} = \sqrt{0^2 + \left(-\frac{1}{2}\right)^2} = \sqrt{\frac{1}{4}} = \frac{1}{2},$$

so our conic $\mathscr{C}$ is an *ellipse* with eccentricity $\frac{1}{2}$, and one focus F_1 at the origin: $F_1(0,0)$. The vertices of $\mathscr{C}$ are obtained using formulas (6.5.30) and (6.5.31), which yield:

$$V_1 : \begin{cases} x_1 = \dfrac{\boldsymbol{\kappa} u}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{2\cdot 0}{\left(1+\frac{1}{2}\right)\cdot\frac{1}{2}} = 0 \\ \\ y_1 = \dfrac{\boldsymbol{\kappa} v}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{2\cdot\left(-\frac{1}{2}\right)}{\left(1+\frac{1}{2}\right)\cdot\frac{1}{2}} = \dfrac{-1}{\frac{3}{4}} = -\dfrac{4}{3} \end{cases}$$

$$V_2 : \begin{cases} x_2 = \dfrac{\boldsymbol{\kappa} u}{(\boldsymbol{e}-1)\boldsymbol{e}} = \dfrac{2\cdot 0}{\left(-\frac{1}{2}-1\right)\cdot\frac{1}{2}} = 0 \\ \\ y_2 = \dfrac{\boldsymbol{\kappa} u}{(\boldsymbol{e}-1)\boldsymbol{e}} = \dfrac{2\cdot\left(-\frac{1}{2}\right)}{\left(-\frac{1}{2}-1\right)\cdot\frac{1}{2}} = \dfrac{-1}{-\frac{1}{4}} = 4 \end{cases}$$

Using midpoints (exactly as in Example 6.5.1, it follows that the center has coordinates $Z\left(0,\frac{4}{3}\right)$, and the second focus is at $F_2\left(0,\frac{8}{3}\right)$.

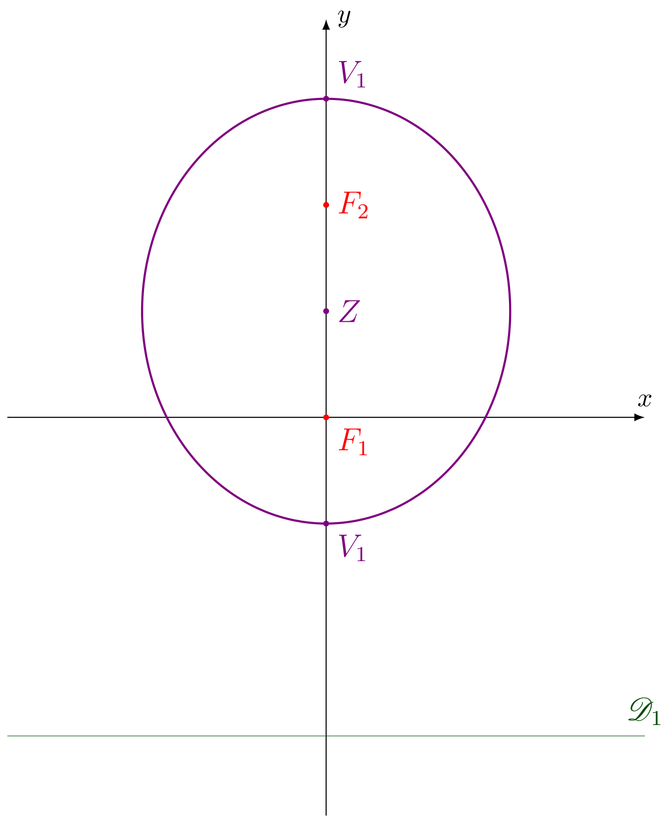


Figure 6.5.3

The focal axis of our ellipse is the y-axis. The directrix $\mathscr{D}_1$ associated with the focus F_1 is given by the equation (6.5.22), which in our case becomes: $-\frac{1}{2}y = 2$. In other words, $\mathscr{D}_1$ is the horizontal line:

$$y = -4.$$

The major radius a is the distance between Z and V_1, so by direct measurement we get $a = \dfrac{8}{3}$. The focal distance c can either be computed using the formula $c = a\boldsymbol{e}$ or by measuring the distance between Z and F_1; either method yields the same value: $c = \dfrac{4}{3}$. Finally since our curve is an ellipse, its minor radius is $b = \sqrt{a^2 - c^2} = \sqrt{\left(\dfrac{8}{3}\right)^2 - \left(\dfrac{4}{3}\right)^2} = \sqrt{\dfrac{64}{9} - \dfrac{16}{9}} = \sqrt{\dfrac{48}{9}} = \dfrac{4\sqrt{3}}{3}$.

Example 6.5.4. Suppose we are given the polar equation

$$r = \frac{10}{5 + 3\cos\theta - 4\sin\theta},$$

and we are asked to describe the geometry of this conic as completely as possible, as we did above

In order to match this equation with (6.5.24), we must force a factoring in the denominator, so we will re-write our equation as

$$r = \frac{20}{5\left(1 + \frac{3}{5}\cos\theta - \frac{4}{5}\sin\theta\right)} = \frac{4}{1 + \frac{3}{5}\cos\theta - \frac{4}{5}\sin\theta},$$

and now we have a perfect match with (6.5.24) by setting $\boldsymbol{\kappa} = 2$, $u = \dfrac{3}{5}$ and $v = -\dfrac{4}{5}$. Using the eccentricity formula, we get

$$\boldsymbol{e} = \sqrt{u^2 + v^2} = \sqrt{\left(\frac{3}{5}\right)^2 + \left(-\frac{4}{5}\right)^2} = \sqrt{\frac{9}{25} + \frac{16}{25}} = \sqrt{\frac{25}{25}} = 1,$$

so our conic $\mathscr{C}$ is a *parabola* with eccentricity 1, and the focus F at the origin: $F(0,0)$. The vertex of $\mathscr{C}$ are obtained using formula (6.5.30) which yields:

$$V: \begin{cases} x_V = \dfrac{\boldsymbol{\kappa} u}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{4 \cdot \frac{3}{5}}{(1+1)\cdot 1} = \dfrac{6}{5} \\[2ex] y_V = \dfrac{\boldsymbol{\kappa} v}{(1+\boldsymbol{e})\boldsymbol{e}} = \dfrac{4 \cdot \left(-\frac{4}{5}\right)}{(1+1)\cdot 1} = -\dfrac{8}{5} \end{cases}$$

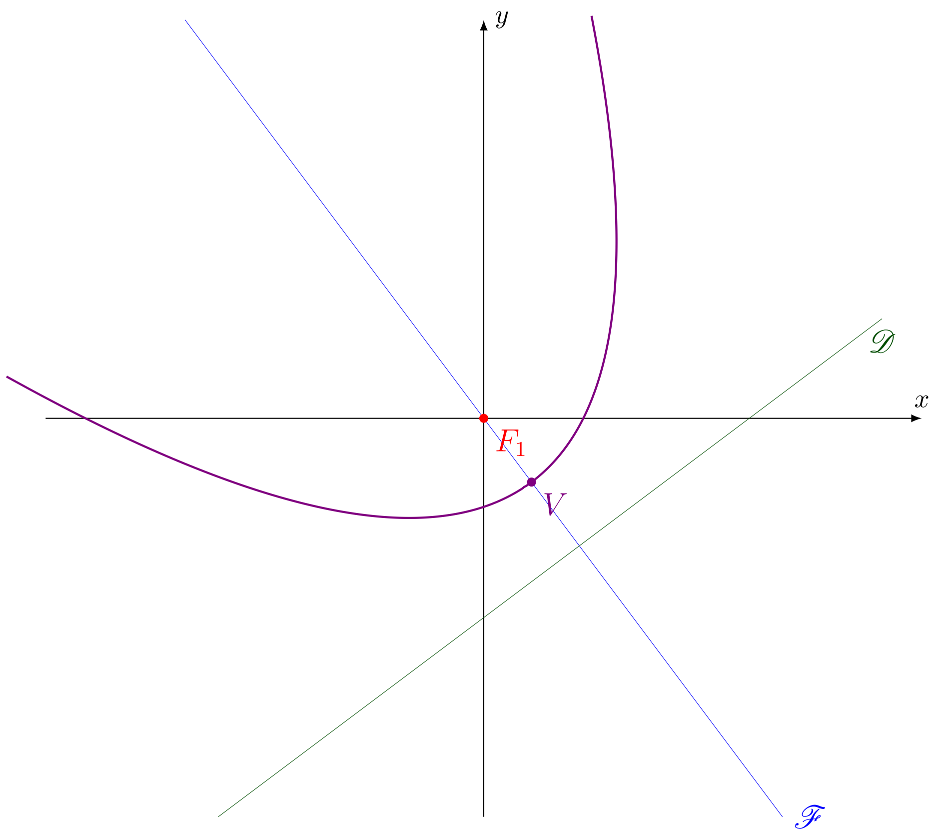


Figure 6.5.4

The directrix of our parabola is given by the equation (6.5.22), which in our case becomes:

$$\frac{3}{5}x - \frac{4}{5}y = 4,$$

which (optionally) can be written a little cleaner (by multiplying everything by 5) as:

$$3x - 4y = 20.$$

The focal axis $\mathscr{F}$ will now have the equation:

$$4x + 3y = 0.$$

Finally, the focal distance of our parabola can be computed as

$$p = \text{dist}(V, F) = \sqrt{(x_V - x_F)^2 + (y_V - y_F)^2} = \sqrt{\left(\frac{6}{5} - 0\right)^2 + \left(-\frac{8}{5} - 0\right)^2} =$$
$$= \sqrt{\frac{36}{25} + \frac{64}{25}} = \sqrt{\frac{100}{25}} = \sqrt{4} = 2.$$

Exercises

In each one of the Exercises 1-8 below, you are asked to completely describe the geometry of the conic represented by the given polar equation, that is, find the eccentricity, vertices and foci. If the curve is an ellipse or a hyperbola, also find the center, the major and minor radii. If the conic is a parabola, find its focal distance and directrix. As a visual aid, you may also sketch the graph.

1. $r = \dfrac{3}{2 - \sin\theta}$

2. $r = \dfrac{3}{2 + 3\cos\theta}$

3. $r = \dfrac{1}{3 - \cos\theta}$

4. $r = \dfrac{3}{2 - 2\cos\theta}$

5. $r = \dfrac{1}{\cos\theta - 4}$

6. $r = \dfrac{7}{3\cos\theta - 2}$

7. $r = \dfrac{5}{2\sin\theta - 1}$

8*. $r = \dfrac{1}{2 - \cos\theta} + \sin\theta$

In each one of the Exercises 9-11, find the equation (in rectangular coordinates) of a conic with the specified eccentricity, focus and associated directrix.

9*. $\boldsymbol{e} = \dfrac{1}{2}$, $F(0, -1)$, $\mathscr{D}$: $x = 2$.

10*. $\boldsymbol{e} = 1$, $F(2, -1)$, $\mathscr{D}$: $x + 2y = 0$.

11*. $\boldsymbol{e} = 2$, $F(1, -1)$, $\mathscr{D}$: $3x - 2x = 10$.

In each one of the Exercises 12-16, find the polar equation of a conic with the one focus at the origin, and specified eccentricity and directrix associated to the focus at the origin.

12. $\boldsymbol{e} = \dfrac{1}{2}$, $\mathscr{D}$: $x = 2$.

13. $\boldsymbol{e} = 1$, $\mathscr{D}$: $x + 2y = 4$.

14. $\boldsymbol{e} = \dfrac{2}{3}$, $\mathscr{D}$: $x + y = 1$.

15. $\boldsymbol{e} = 1$, $\mathscr{D}$: $5x - 2y = 10$.

16. $\boldsymbol{e} = 2$, $\mathscr{D}$: $3x + 5x = 15$.

Appendices

A Algebra Review I: Equations

This Appendix provides a quick review of the various types of equations frequently used in Trigonometry. We understand **equations** as *equalities between expressions that involve unknown quantities*. The simplest examples of equations are those the contain only *one unknown*, for example

$$x^2 + 2x = 2x^2 - 3, \tag{A.1}$$

where the unknown quantity (a.k.a. *variable*) is denoted by x. To **solve** an equation in one variable simply means to *find the values for the variable which satisfy the equality given in the equation*. Such values are called **solutions** of the equations.

▶ If you know how to solve the equation (A.1), do it now! (Otherwise, wait until we discuss *Quadratic Equations* later in this section.)

When comparing two equations, we often use the following terminology.

> Two equations are **equivalent**, if they have the same solutions.

With the above definition in mind, the "game" we play when we want to solve an equation is to *transform* them into equivalent ones, so we change the equation:

$$\text{original (old) equation} \xrightarrow{\text{transformation}} \text{transformed (new) equation.}$$

A "perfect" transformation is one that has the old and the new equations equivalent. A "safe" (but possibly "imperfect") transformation is one for which the new equation does not loose any solutions of the old one.

The two most popular "perfect" transformations are:

> I. Add or subtract a common expression from both sides of the equation.
> II. Multiply or divide both sides of the equation by a **non-zero** quantity.

Example A.1. Solve: $8x - 2 = 5x - 8$.

Solution.

original equation:	$8x - 2 = 5x - 8;$
add 2, subtract $5x$:	$8x - 2 + 2 - 5x = 5x - 8 + 2 - 5x;$
combine like terms:	$3x = -6;$
divide by $3 (\neq 0!)$:	$x = \dfrac{-6}{3} = -2.$

Example A.2. If we start with the equation $x^2 = x$ and we think of dividing by x, we are **incorrect!** We are not exactly sure if we are dividing by a non-zero number! After dividing by x the new equation looks like $x = 1$, so it is already solved. However, in the process of dividing by x we eliminated the possibility that $x = 0$, which in fact is another solution of the original equation. Thus dividing by x made us loose a solution.

Linear Equations

A **basic linear equation**, is one of the form:[18]

$$\textit{number}_1 \cdot ? = \textit{number}_2, \tag{A.2}$$

One such equation showed up in Example A.1, where we saw how to solve it. The rules for solving such equations are:

Solutions of basic linear equations

Depending on whether one or both coefficients vanish in a basic linear equation of the form (A.2), the solutions of the equation are as follows.

- If $\textit{number}_1 \neq 0$, there is only **one** solution: $? = \dfrac{\textit{number}_2}{\textit{number}_1}$, so its solution set has only one element.
- If $\textit{number}_1 = 0$ *and* $\textit{number}_2 \neq 0$, then the equation has **no solutions**, so its solution set is $\varnothing$ – the empty set.
- If $\textit{number}_1 \neq 0$ *and* $\textit{number}_2 = 0$, then the equation has **all real numbers as solutions**, so its solution set is $\mathbb{R}$ – the set of all real numbers.

A slightly larger class of equations are the so-called **general linear equations**, which look like:

$$\heartsuit \cdot ? + \spadesuit = \diamondsuit \cdot ? + \clubsuit. \tag{A.3}$$

As we have already seen in Example A.1, the following principle is clearly true in general.

Any general linear equation of the form (A.3) *can be "perfectly" transformed into a basic linear equation of the form* (A.2).

Example A.3. Let us consider the following equation in *two* variables:

$$xy = x + 2y. \tag{A.4}$$

We can think of (A.4) as a linear equation in two ways:
(i) The unknown is x, so we think y as a known quantity, thus the equation looks like:

$$\heartsuit \cdot x = x + \diamondsuit, \tag{A.5}$$

with both $\heartsuit(= y)$ and $\diamondsuit(= 2y)$ treated as numbers.
(ii) The unknown is y, so we think x as a known quantity, thus the equation looks like:

$$\clubsuit \cdot y = \spadesuit + 2y \tag{A.6}$$

with both $\clubsuit(= x)$ and $\spadesuit(= x)$ treated as numbers.

[18] At times, when we don't want to use specific letters, we use "funny" symbols like $\heartsuit$, $\clubsuit$, $\spadesuit$, $\diamondsuit$, $\#$, $\sharp$, $\natural$, etc., which are place-holders for known or unknown (most often numerical) quantities.

This means that when we are asked to *solve* (A.4) *for* x, we must think as in (i); but when we are asked to *solve* (A.4) *for* y, we must think as in (ii). With these considerations in mind the solutions are as follows. (CAUTION: The solutions given in (A.8) and (A.10) are **sloppy**! We will revisit them in the note that follows this Example.)

(i). To solve for x we make a plan on how to solve (A.5): we would subtract x, then group like terms to make the equation look like a basic linear one: *number*$_1 \cdot x =$ *number*$_2$, and finally we would solve by division. With this plan in mind, our steps are as follows:

$$\begin{aligned} \text{original equation:} &\quad xy = x + 2y; \\ \text{subtract } x\text{:} &\quad xy - x = 2y; \\ \text{combine like terms:} &\quad (y-1)x = 2y; \qquad \text{(A.7)} \\ \text{divide by } (y-1)\text{:} &\quad x = \frac{2y}{y-1}. \qquad \text{(A.8)} \end{aligned}$$

(ii). To solve for y, our plan is similar: subtract $2y$, then group like terms to make the equation look like a basic linear one: *number*$_1 \cdot y =$ *number*$_2$, and finally we would solve by division.

$$\begin{aligned} \text{original equation:} &\quad xy = x + 2y; \\ \text{subtract } 2y\text{:} &\quad xy - 2y = x; \\ \text{combine like terms:} &\quad (x-2)y = x; \qquad \text{(A.9)} \\ \text{divide by } (y-1)\text{:} &\quad y = \frac{x}{x-2}. \qquad \text{(A.10)} \end{aligned}$$

Both answers given in the preceding Example are incomplete. The transformations of the original equation (A.4) into both (A.7) and (A.9) are "perfect," but the final answers given in (A.8) and (A.10) are sloppy, because we did not care that much about dividing by a number which might in fact be equal to zero: (i) in (A.8) we must account for the possibility that $y - 1 = 0$, that is, $y = 1$; (ii) in (A.10) we must account for the possibility that $x - 2 = 0$, that is, $x = 2$. Therefore the complete answers should look like:

$$x = \begin{cases} \dfrac{2y}{y-1}, \text{ if } y \neq 1 \\ \text{no solution, if } y = 1 \end{cases} \qquad y = \begin{cases} \dfrac{x}{x-2}, \text{ if } x \neq 2 \\ \text{no solution, if } x = 2 \end{cases}$$

Proportions

In elementary school we learned that a **proportion** is an *equality of fractions (ratios)*, so it is presented as

$$\frac{\text{numerator}_1}{\text{denominator}_1} = \frac{\text{numerator}_2}{\text{denominator}_2}. \qquad \text{(A.11)}$$

Of course, since we deal with fractions, *the denominators are assumed to be non-zero*. Since the division is the inverse operation to multiplication, proportions can always be transformed according to the following scheme.

Proportions in fraction-less form

I. A proportion of the form (A.11) yields an equality of two products, obtained by *cross-multiplication*

$$\text{numerator}_1 \times \text{denominator}_2 = \text{numerator}_2 \times \text{denominator}_1.$$

II. Conversely, an equality of products, in which each product has a non-zero factor

$$\spadesuit \cdot \heartsuit = \clubsuit \cdot \diamondsuit, \quad \heartsuit, \diamondsuit \neq 0,$$

yields a proportion obtained by *cross-division* by the two non-zero factors:

$$\frac{\spadesuit}{\diamondsuit} = \frac{\clubsuit}{\heartsuit}.$$

Going back and forth between two forms of a proportion is a useful technique for solving **basic proportion equations**, which are proportions in which one element is unknown, while the other three are given, so they look like $\frac{?}{\diamondsuit} = \frac{\clubsuit}{\heartsuit}$, or like $\frac{\spadesuit}{?} = \frac{\clubsuit}{\heartsuit}$. As long as the given denominators in a proportion equation are non-zero, we can "perfectly" transform the equation using cross-multiplication into a basic linear equation, which we can then solve by division. We summarize this method using the phrase "*cross-multiply, then divide*."

Example A.4. Consider the equation: $\frac{3}{t} = \frac{7}{9}$. After cross-multiplication (which is a "perfect" transformation, because the only given denominator is $9 \neq 0$), we get the basic linear equation $7t = 3 \cdot 9 = 27$, which by division (which is OK, since we divide by $7 \neq 0$) yields: $t = \frac{27}{7} = 3\frac{6}{7} = 2.857142857142\cdots = 2.\overline{857142}$.

Equations Vs. Identities

In Algebra we often use *identities* (or *formulas*), which make our life a little easier, especially when we need to solve equations, or when we want to simplify certain algebraic expressions. When comparing identities with equations, the bottom line is:

An **identity** is a "trivial" equation, than is, an *equality which is satisfied by **all** possible values of the variables*.

Example A.5. On the one hand, the equation $(x+1)^2 = x^2 + 2x + 1$ *is* an identity: no matter what value of x we plug in, the equality will be satisfied.

On the other hand, the equation $(x+1)^2 = x^2 + 1$ *is not* an identity: there are values for x, for instance $x = 1$, for which the equality fails.

Among the many of identities used in Algebra, the package below is the most popular one:

Quadratic Product Identities

$$(\heartsuit + \spadesuit)^2 = \heartsuit^2 + 2\heartsuit\spadesuit + \spadesuit^2;$$
$$(\heartsuit - \spadesuit)^2 = \heartsuit^2 - 2\heartsuit\spadesuit + \spadesuit^2;$$
$$(\heartsuit + \spadesuit)(\heartsuit - \spadesuit) = \heartsuit^2 - \spadesuit^2.$$

Equations which are "almost trivial," so they hold for instance for all values in an interval (or a union of intervals) are referred to as *conditional identities*. We write them using the format "*equation, condition on the variable(s)*," for example

$$\frac{x^2-1}{x-1} = x+1, \quad x \neq 1.$$

Power Equations with Positive Integer Exponents

Besides linear equations, the *power* equations are the next easiest to solve. A **power equation** is one of the form:

$$?^{\heartsuit} = \textit{number}. \tag{A.12}$$

with *number* and $\heartsuit$ real numbers. The number $\heartsuit$ is called the **exponent** of the power equation. At this point we only discuss the easiest case, when the exponent is a *positive integer*[19], so the left-hand of (A.12) is the product:

$$?^{\heartsuit} = \underbrace{? \cdot ? \cdots ?}_{\heartsuit \text{ factors}}.$$

Of course, when the exponent is equal to 1, the equation is already solved, so we can in fact assume that it is at least 2.

Unless we are in some very special cases (as in Example A.6) the way we solve power equations is not an honest one. What we are about to do is to make a statement and cook up a notation that will help us cheat a little bit.

Power Equation Facts and Radical Notation

Assume the exponent $\heartsuit$ is an integer ≥ 2.

FACT 1. *Whenever the equation* (A.12) *has real solutions, exactly* ***one*** *of its real solutions has the* ***same sign*** *as number (the right-hand side). This special solution is denoted by* $\sqrt[\heartsuit]{\textit{number}}$. *However, if* (A.12) *has* ***no real solution*** *(see Fact 2 below), then the quantity* $\sqrt[\heartsuit]{\textit{number}}$ *is* ***not defined!***

FACT 2. *Depending on the parity of of the exponent* $\heartsuit$*, the complete solutions of the power equation* (A.12) *are as follows.*

I. *If* $\heartsuit$ *is* ***odd****, then the equation has exactly* ***one real solution****:* $? = \sqrt[\heartsuit]{\textit{number}}$.

II. *If* $\heartsuit$ *is* ***even****, then according to the* ***sign*** *of number (the right-hand side), we have the following three possibilities.*

   (a) *If number* > 0*, the equation has* ***two real solutions:***

   - *one positive solution:* $? = \sqrt[\heartsuit]{\textit{number}}$*, and*

[19] The case when the exponent is an *arbitrary real number* will be treated in Section 0.2.

- *one negative solution:* $? = -\sqrt[\heartsuit]{number}$;

(b) *If number* $= 0$, *the equation has exactly* ***one real solution:*** $? = \sqrt[\heartsuit]{\mathbf{0}} = 0$.
(c) *If number* < 0, *the equation has* ***no real solution***.

For cases (a) *and* (b), *we can also use the "lazy" notation:* $? = \pm\sqrt[\heartsuit]{number}$.

CONVENTION. A special case of interest, particularly in Geometry and Trigonometry, is when the exponent is $\heartsuit = 2$. In this case $\sqrt[2]{\ }$ is simply denoted by $\sqrt{\ }$, and we call it the *square root operation*.

Example A.6. $\sqrt{16} = ?$

According to the above definition, we must look at the equation $?^2 = 16$, solve it, and pick the solution which is positive. It is clear that the particular solution we are looking for is 4, so we have $\sqrt{16} = 4$. As for the equation $?^2 = 16$, it has two solutions: $? = \pm 4$.

Example A.7. Solve the equation $x^2 = 2$.

According to Fact 2.II above, the solutions of this equation are: $x = \pm\sqrt{2}$. On the one hand, our use the square root notation is clearly a way to "cheat" this problem, because we still don't know how big or small $\sqrt{2}$ really is. On the other hand, one can also make a claim that, in fact we do know what $\sqrt{2}$ quite well: *$\sqrt{2}$ is the unique positive real number whose square is equal to 2*. When doing calculations, this implicit description is good enough, for example

$$(\sqrt{2})^{10} = 32,$$

and this is why we say that "$\sqrt{2}$ is an *exact* value" notation. If we want to get an idea how $\sqrt{2}$ "feels" as a number, then we will have to use a calculator (or do a hand computation as shown in Appendix A), which reveals that: $\sqrt{2} = 1.4142135623730950488016887242\dots$. Unlike what we saw in Example A.4, there is no repeating pattern [20] of decimals here, and this is evidence that $\sqrt{2}$ is *irrational*. One can actually *prove* that $\sqrt{2}$ is indeed irrational.

As we understand them, radicals are *quantities described implicitly*, so our preceding definition can be restated as follows.

For any real *number*, and any integer $\heartsuit \geq 2$,

$$\sqrt[\heartsuit]{number} = ? \text{ means: } \begin{cases} ?^{\heartsuit} = number, \\ \qquad \textbf{and} \\ ? \text{ is a real number which has } \textbf{same sign} \text{ as } number \end{cases}$$

If such a quantity exists, *it is* ***unique***. If such a quantity does not exist, then $\sqrt[\heartsuit]{number}$ *is* ***not defined***.

The following sign features are always helpful when dealing with radicals:

[20] In Example A.4 we encountered the rational number $\frac{27}{7}$, whose decimal expansion was $2.\overline{857142}$, meaning that after the decimal point, the string "857142" repeats itself. This happens with *any rational number*.

Sign Rules for Radicals

$$\sqrt[\text{any}]{\mathbf{0}} = 0; \qquad \sqrt[\text{any}]{\text{positive}} = \text{positive};$$

$$\sqrt[\text{odd}]{\text{negative}} = \text{negative}; \qquad \sqrt[\text{even}]{\text{negative}} = \text{undefined}.$$

All identities included in the formula package shown below are *conditional!* (Besides the main condition, other restrictions are shown next to the identity.)

Radical Identities

*Assume **all radicals below are defined**:*

$$\left(\sqrt[\heartsuit]{number}\right)^{\heartsuit} = number. \qquad \text{(Undoing Identity I)}$$

$$\sqrt[\heartsuit]{number^{\heartsuit}} = \begin{cases} number, & \text{if } \heartsuit \text{ is odd;} \\ |number|, & \text{if } \heartsuit \text{ is even.} \end{cases} \qquad \text{(Undoing Identity II)}$$

$$\sqrt[\heartsuit]{\sqrt[\spadesuit]{number}} = \sqrt[\heartsuit\cdot\spadesuit]{number} \qquad \text{(“Cascading” Formula)}$$

$$\sqrt[\heartsuit]{number_1 \cdot number_2} = \sqrt[\heartsuit]{number_1} \cdot \sqrt[\heartsuit]{number_2}. \qquad \text{(Product Formula)}$$

$$\sqrt[\heartsuit]{\frac{numerator}{denominator}} = \frac{\sqrt[\heartsuit]{numerator}}{\sqrt[\heartsuit]{denominator}}, \text{ if } denominator \neq \mathbf{0}. \qquad \text{(Quotient Formula)}$$

CLARIFICATION. In the even case of undoing formula II, one uses the **absolute value** of *number*, which is defined as

$$|number| = \begin{cases} number, & \text{if } number \geq 0 \\ -number, & \text{if } number < 0 \end{cases}$$

In other words $|number|$ is pretty much the same as *number*, *except that it removes the "$-$" sign, if number happens to be negative*. For instance $|3| = |-3| = 3$, so when we only care about the absolute values, the numbers 3 and -3 become indistinguishable.

Undoing even radicals without the use of absolute value is incorrect! In particular, an equality like "$\sqrt{x^2} = x$" is not a true identity. In works for $x \geq 0$, but fails for $x < 0$, for instance: $\sqrt{(-2)^2} = \sqrt{4} = 2 \neq -2$.

Example A.8. We can use the Product and Quotient Formulas to pull out squares from under square roots, for instance:

$$\sqrt{\frac{18}{175}} = \frac{\sqrt{18}}{\sqrt{175}} = \frac{\sqrt{9\cdot 2}}{\sqrt{25\cdot 7}} = \frac{\sqrt{9}\sqrt{2}}{\sqrt{25}\sqrt{7}} = \frac{3\sqrt{2}}{5\sqrt{7}}.$$

The final answer in the above calculation can be simplified to “look a little nicer” in two ways:

$$\frac{3\sqrt{2}}{5\sqrt{7}} = \frac{3(\sqrt{2})^2}{5\sqrt{7}\sqrt{2}} = \frac{3\cdot 2}{5\sqrt{7\cdot 2}} = \frac{6}{5\sqrt{14}}; \text{ or} \tag{A.13}$$

$$\frac{3\sqrt{2}}{5\sqrt{7}} = \frac{3\sqrt{2}\sqrt{7}}{5(\sqrt{7})^2} = \frac{3\sqrt{2\cdot 7}}{5\cdot 7} = \frac{3\sqrt{14}}{35}. \tag{A.14}$$

Simplifying as we did in (A.13) is referred to as *rationalizing the numerator*, whereas what we did in (A.14) is referred to as *rationalizing the denominator*.

Quadratic Expressions

Besides linear and power equations, the so-called *quadratic* equations are the next easiest to solve. Concerning the algebraic expressions involved in such equations, we use the following terminology.

A **quadratic expression** in a variable $\boldsymbol{t}$ is an algebraic expression of the form

$$\boldsymbol{Q}(\boldsymbol{t}) = \boldsymbol{a}\boldsymbol{t}^{\mathbf{2}} + \boldsymbol{b}\boldsymbol{t} + \boldsymbol{c}, \tag{A.15}$$

with $\boldsymbol{a}$ (the so-called *leading coefficient*) always assumed to be ***non-zero***. For any quadratic expression as above, three important quantities are also relevant:

- the *discriminant*: $\boldsymbol{D} = \boldsymbol{b}^{\mathbf{2}} - 4\boldsymbol{a}\boldsymbol{c}$;
- the *critical number*: $\boldsymbol{t}_{\text{critical}} = -\dfrac{\boldsymbol{b}}{\mathbf{2}\boldsymbol{a}}$;
- the *critical value*: $\boldsymbol{Q}(\boldsymbol{t}_{\text{critical}}) = -\dfrac{\boldsymbol{D}}{\mathbf{4}\boldsymbol{a}}$.

COMMENT. In the Algebra course, we learned that for a quadratic function $f(x) = ax^2+bx+c$, the critical number x_{critical} and the critical value $f(x_{\text{critical}})$ are precisely the *coordinates of the vertex* of the graph of $y = f(x)$. Alternatively, the critical value can also be computed with the help of the *discriminant*, using the second formula: $-\dfrac{\boldsymbol{D}}{\mathbf{4}\boldsymbol{a}}$.

Completed Square Identity

$$\boldsymbol{Q}(\boldsymbol{t}) = \boldsymbol{a}(\boldsymbol{t} - \boldsymbol{t}_{\text{critical}})^{\mathbf{2}} + \boldsymbol{Q}(\boldsymbol{t}_{\text{critical}}) = \boldsymbol{a}\left(\boldsymbol{t} + \frac{\boldsymbol{b}}{\mathbf{2}\boldsymbol{a}}\right)^{\mathbf{2}} - \frac{\boldsymbol{D}}{\mathbf{4}\boldsymbol{a}} \tag{A.16}$$

Example A.9. Complete the square in: $f(x) = 2x^2 - 12x - 5$.

Solution. The leading coefficient is $a = 2$ and the middle coefficient is $b = -12$. Thus the critical number is $x_{\text{critical}} = -\dfrac{-12}{2\cdot 2} = 3$, and the critical value is $f(3) = 2(3)^2 - 12(3) - 5 = -23$. Thus, after completing the square, our expression is:

$$f(x) = 2(x-3)^2 - 23.$$

Example A.10. Complete the squares in: $x^2 + 14x - 4y^2 + 8y - 7$.

Solution. This is a quadratic expression in *two* variables. It will be better in this case to split our expression as a sum of two separate expressions:

$$x^2 + 14x - 4y^2 + 8y - 7 = \big[\underbrace{x^2 + 14x}_{f(x)}\big] + \big[\underbrace{-4y^2 + 8y - 7}_{g(y)}\big], \tag{A.17}$$

and to complete the squares in each expression.[21]

In $f(x)$ the leading coefficient is $a = 1$ and the middle coefficient is $b = 14$. Thus the critical number is $x_{\text{critical}} = -\dfrac{14}{2 \cdot 1} = -7$, and the critical value is $f(-7) = (-7)^2 + 14(-7) = -49$. Thus, after completing the square, $f(x)$ becomes:

$$f(x) = (x+7)^2 - 28. \tag{A.18}$$

In $g(y)$ the leading coefficient is $a = -4$ and the middle coefficient is $b = 8$. Thus the critical number is $y_{\text{critical}} = -\dfrac{8}{2(-4)} = 1$, and the critical value is $g(1) = -4(1)^2 + 8(1) - 7 = -3$. Thus, after completing the square, $g(y)$ becomes:

$$g(y) = -4(y-1)^2 + 4. \tag{A.19}$$

Now we go back to (A.17) and replace $f(x)$ and $g(y)$ using the above two identities:

$$x^2 + 14x - 4y^2 + 8y - 7 = \big[\underbrace{(x+7)^2 - 49}_{f(x)}\big] + \big[\underbrace{-4(y-1)^2 - 3}_{g(y)}\big] = (x+7)^2 - 4(y-1)^2 - 52.$$

Quadratic Equations

A *basic quadratic equation* (with unknown $\boldsymbol{v}$) is one of the form;

$$\boldsymbol{a}\boldsymbol{t}^2 + \boldsymbol{b}\boldsymbol{t} + \boldsymbol{c} = 0, \tag{A.20}$$

where the leading coefficient $\boldsymbol{a}$ is *not equal to zero*.

We (should) always solve these equations using the following.

[21] We could have included the constant term in the first expression, so we could have split into $f(x) = x^2 + 14x - 7$ and $g(y) = -4y^2 + 8y$. In the end, the result will be the same.

Quadratic Formula

Assuming we have a quadratic equation (A.20), *with real coefficients, the real solutions are determined by the sign of the discriminant* $D = b^2 - 4ac$, *as follows:*

(A) *If* $D < 0$, *the equation has no real solutions.*

(B) *If* $D = 0$, *the equation has exactly one real solution, which is equal to the critical number* $t = -\dfrac{b}{2a}$.

(C) *If* $D > 0$, *the equation has two distinct real solutions:*

$$t = \frac{-b \pm \sqrt{D}}{2a}. \tag{A.21}$$

CLARIFICATION. Using square completion, we can re-write our equation (A.20) as

$$a\left(t + \frac{b}{2a}\right)^2 - \frac{D}{4a} = 0,$$

which upon multiplying everything by 4ℓ, can be re-written as:

$$4a^2\left(t + \frac{c}{2a}\right)^2 - D = 0.$$

Since the first term can be simplified as $4a^2\left(t + \dfrac{b}{2a}\right)^2 = \left(2a\left(t + \dfrac{b}{2a}\right)\right)^2 = (2at + b)^2$, upon adding D, our equation now reads:

$$(2at + b)^2 = D.$$

We can treat this as a power equation $?^2 = D$, with unknown $? = 2at + b$, so when we solve it we get $? = \pm\sqrt{D}$, so we now have

$$2at + b = \pm\sqrt{D},$$

and the desired formula (A.21) follows immediately, by subtracting b, then by dividing by $2a$.

It should be pointed out here that the Quadratic Formula (A.21) also works in case (B), but in that case we will have $\pm\sqrt{D} = 0$.

B Algebra Review II: Matrix Arithmetic

In this Appendix we review the basic facts concerning matrix arithmetic. We think a **matrix** simply as a *rectangular table filled with numbers*. The *size* of a matrix is always be specified in the form

number of rows $\times$ *number of columns*,

so for example a 2×2 matrix will be presented as

$$\mathbf{A} = \begin{bmatrix} \clubsuit & \diamondsuit \\ \heartsuit & \spadesuit \end{bmatrix}.$$

We "navigate" the entries (a.k.a. *coefficients*) in our matrix using notations like a_{ij}, where the first index (i) denotes the row number, and the second index (j) denotes the column number. For example, the coefficients of the matrix A given above are: $a_{11} = \clubsuit$, $a_{12} = \diamondsuit$, $a_{21} = \heartsuit$, and $a_{22} = \spadesuit$. Two matrices are **equal**, if *they have same size, and their coefficients match according to their position*.

Matrix Addition, Subtraction, and Scalar Multiplication

The main limitation of addition and subtraction of matrices is the fact that *matrices can be added/subtracted, only if they have the same size*.

Within this limitation, *addition and subtraction of matrices of same size is carried on entry-by-entry, so the result has the same size as the terms in the sum/difference*. Specifically, if we look at two $m \times n$ matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{13} & \dots & a_{mn} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & b_{13} & \dots & b_{mn} \end{bmatrix},$$

their sum/difference will be again two $m \times n$ matrices:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & a_{13} + b_{m3} & \dots & a_{mn} + b_{mn} \end{bmatrix};$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & a_{13} - b_{13} & \dots & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & a_{23} - b_{23} & \dots & a_{2n} - b_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} - b_{m1} & a_{m2} - b_{m2} & a_{13} - b_{m3} & \dots & a_{mn} - b_{mn} \end{bmatrix}$$

Example B.1.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} -2 & 1 & -3 \\ 7 & -1 & 4 \end{bmatrix} = \begin{bmatrix} 1+(-2) & 2+1 & 3+(-3) \\ 4+7 & 5+(-1) & 6+4 \end{bmatrix} = \begin{bmatrix} 11 & 3 & 0 \\ 11 & 4 & 10 \end{bmatrix}.$$

Another easy operation involving matrices is *scalar multiplication*, which is carried on by *multiplying each entry in the matrix by the given number*. Specifically, if we start with some real number t and an $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{13} & \dots & a_{mn} \end{bmatrix},$$

then the scalar product $t\mathbf{A}$ is again an $m \times n$ matrix

$$t\mathbf{A} = \begin{bmatrix} ta_{11} & ta_{12} & ta_{13} & \dots & ta_{1n} \\ ta_{21} & ta_{22} & ta_{23} & \dots & ta_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ ta_{m1} & ta_{m2} & ta_{13} & \dots & ta_{mn} \end{bmatrix}.$$

Example B.2.

$$(-2)\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} (-2)1 & (-2)2 \\ (-2)3 & (-2)4 \\ (-2)5 & (-2)6 \end{bmatrix} = \begin{bmatrix} -2 & -4 \\ -6 & -8 \\ -10 & -12 \end{bmatrix}$$

It is fairly obvious that matrix subtraction can also be understood using addition and scalar multiplication by -1, that is,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}.$$

In many respects, addition and scalar multiplication for matrices behaves very much like real number arithmetic.

If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices of same size $m \times n$, then the following identities hold.

I. *Associativity of Addition:* $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.

II. *Commutativity of Addition:* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

III. *Opposite Property:* $(-\mathbf{A}) + \mathbf{A} = \mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m\times n}$; here $-\mathbf{A} = (-1)\mathbf{A}$ and $\mathbf{0}_{m\times n}$ is the $m \times n$ *zero matrix*, which has all entries equal to zero.

IV. *Easy Scalar Multiplications:* $0\mathbf{A} = \mathbf{0}_{m\times n}$; $1\mathbf{A} = \mathbf{A}$; $t\mathbf{0}_{m\times n} = \mathbf{0}_{m\times n}$.

V. *Associativity of Scalar Multiplication:* $s(t\mathbf{A}) = t(s\mathbf{A}) = (st)\mathbf{A}$.

VI. *Distributivity over Scalar Addition:* $(s + t)\mathbf{A} = s\mathbf{A} + t\mathbf{A}$.

VII. *Distributivity over Matrix Addition:* $t(\mathbf{A} + \mathbf{B}) = t\mathbf{A} + t\mathbf{B}$.

Matrix Multiplication

The product of matrices is built up starting with its simplest version, which is described as follows. Given a *row matrix* $\mathbf{R}$ and a *column matrix* $\mathbf{C}$

$$\mathbf{R} = [x_1 \ x_2 \ \ldots \ x_n], \quad \mathbf{C} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

both having the *same number of entries*, the row-times-column product $\mathbf{RC}$ is the number:

$$\mathbf{R} \cdot \mathbf{C} = x_1y_1 + x_2y_2 + \ldots + x_ny_n.$$

More generally, given an $m \times n$ matrix $\mathbf{A}$, and and $n \times k$ matrix $\mathbf{B}$, the matrix product $\mathbf{P} = \mathbf{AB}$ is an $m \times k$ matrix, constructed by the the following rules:

(i) $\mathbf{P}$ has as many rows as $\mathbf{A}$ has, and as many columns as $\mathbf{B}$ has. Thus $\mathbf{A}$ is an $m \times k$ matrix.

(ii) The entry p_{ij} in $\mathbf{P}$ (that sits in row i and column j) is the row-times-column product

$$p_{ij} = [i^{\text{th}} \text{ row of } \mathbf{A}] \cdot [j^{\text{th}} \text{ column of } \mathbf{B}].$$

The matrix product $\mathbf{P} = \mathbf{AB}$ is defined only if the following match occurs:

$$\text{number of columns in } \mathbf{A} \;=\; \text{number of rows in } \mathbf{B}.$$

In general, the matrix product is not commutative! This means that, even in the cases when both the matrix products $\mathbf{AB}$ and $\mathbf{BA}$ are defined, they may be different. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 8 \end{bmatrix}, \text{ but } \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix}.$$

Example B.3. In Trigonometry, we only use 2×2 and 2×1 matrices, for which the only possible products we are dealing with are covered by the following formulas

$$\begin{bmatrix} \clubsuit & \diamondsuit \\ \heartsuit & \spadesuit \end{bmatrix} \cdot \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} \clubsuit s + \diamondsuit t \\ \heartsuit s + \spadesuit t \end{bmatrix}; \tag{B.1}$$

$$\begin{bmatrix} \clubsuit & \diamondsuit \\ \heartsuit & \spadesuit \end{bmatrix} \cdot \begin{bmatrix} s & u \\ t & v \end{bmatrix} = \begin{bmatrix} \clubsuit s + \diamondsuit t & \clubsuit u + \diamondsuit v \\ \heartsuit s + \spadesuit t & \heartsuit u + \spadesuit v \end{bmatrix}. \tag{B.2}$$

Even though matrix multiplication is not commutative, it still has some nice properties which again resemble some from real number arithmetic.

I. *Associativity of Matrix Multiplication:* If $\mathbf{AB}$ and $\mathbf{BC}$ are defined then the following products are defined and are equal: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

II. *Distributivity on the Right over Matrix Addition:* If $\mathbf{AB}$ and $\mathbf{AC}$ are defined, then so is $\mathbf{A}(\mathbf{B}+\mathbf{C})$, and furthermore: $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$.

III. *Distributivity on the Left over Matrix Addition:* If $\mathbf{AC}$ and $\mathbf{BC}$ are defined, then so is $(\mathbf{A}+\mathbf{B})\mathbf{C}$, and furthermore: $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$.

One special type of matrices, which appear very naturally when dealing with matrix products, are the **identity** matrices $\mathbf{I}_n$, which are $n \times n$ (thus *square*) matrices, which have *1's on the diagonal, and zeros everywhere else*. For example,

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \qquad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The most important property of such matrices is the following "easy" multiplication rule:

If $\mathbf{A}$ is an $m \times n$ matrix, then:

$$\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}. \tag{B.3}$$

Invertible Matrices

When dealing with matrix multiplications, one can also ask if there is any notion of *division*. After all, when doing number arithmetic, we always understand the division operation as the "undoing" of a multiplication, or equivalently, multiplication by *reciprocals*. With these considerations in mind, one introduces the following definition.

A *square* $n \times n$ matrix $\mathbf{A}$ is said to be **invertible**, if one can find another square $n \times n$ matrix $\mathbf{B}$, such that

$$\mathbf{AB} = \mathbf{I}_n.$$

CLARIFICATIONS. If a matrix $\mathbf{B}$ as above exists, then it is in fact *unique*, and it also satisfies the equality

$$\mathbf{BA} = \mathbf{I}_n.$$

For this, and many other reasons, the matrix $\mathbf{B}$ is denoted by $\mathbf{A}^{-1}$, and is referred to as *the inverse of* $\mathbf{A}$.

The inverse matrices (if they exist) are very useful for solving *systems of linear equations*. Any system with n equations and n unknowns, of the form

$$\left\{\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= c_n \end{aligned}\right. \tag{B.4}$$

can also be presented in matrix form as a multiplication equation $\mathbf{AX} = \mathbf{C}$, where:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}; \quad \mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

The point is that, *if* $\mathbf{A}$ *is invertible*, then the equation $\mathbf{AX} = \mathbf{C}$ can be easily solved by multiplying both sides on the left by' $\mathbf{A}^{-1}$, to give $\mathbf{X} = \mathbf{A}^{-1}\mathbf{C}$. In turn, this means that, if we have a way to compute the inverse matrix, by finding its coefficients, say

$$\mathbf{A}^{-1} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix},$$

then the given system (B.4) is immediately solved in matrix form:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix},$$

which when we write everything down reads:

$$\left\{\begin{aligned} x_1 &= b_{11}c_1 + b_{12}c_2 + \cdots + b_{1n}c_n \\ x_2 &= b_{21}c_1 + b_{22}c_2 + \cdots + b_{2n}c_n \\ &\vdots \\ x_n &= b_{n1}c_1 + b_{n2}c_2 + \cdots + b_{nn}c_n \end{aligned}\right.$$

In the College Algebra course you learned how to decide if a matrix is invertible and (if it is) how to compute its inverse, both "by hand" (using reduced row-echelon forms) and with the help of a scientific calculator. Fortunately, in the 2×2 case (the only one needed in Trigonometry), this problem is quite easy to solve, using the following scheme.

Invertibility Test & Formula for 2 × 2 Matrices

A matrix $\mathbf{A} = \begin{bmatrix} \clubsuit & \diamondsuit \\ \heartsuit & \spadesuit \end{bmatrix}$ is *invertible*, if and only if the quantity $\det(\mathbf{A}) = \clubsuit\spadesuit - \diamondsuit\heartsuit$ (which is referred to as the *determinant of* $\mathbf{A}$) is *non-zero*. Moreover, if this is the case, then:

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} \spadesuit & -\diamondsuit \\ -\heartsuit & \clubsuit \end{bmatrix}. \tag{B.5}$$

Example B.4. Consider the matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, and let us find its inverse (if it has one).

By definition of the determinant, we have $\det(\mathbf{A}) = 1 \cdot 4 - 2 \cdot 3 = -2 \neq 0$, so $\mathbf{A}$ is indeed invertible, and using (B.5) its inverse is:

$$\mathbf{A}^{-1} = \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}.$$

POLAR GRAPH PAPER