Unit 2: Transformations and the Ruler Postulate

Math 330B Spring 2006 (Barsamian)

1. Introduction

In the lists of Euclidean geometry postulates found in most high school geometry textbooks, there are two postulates having to do with measuring distance. One postulate has to do with measuring distance between any two points in the plane. The other postulate has to do with "coordinate functions" on lines, the so-called "ruler postulate". Being aimed at high school students, the wording of the postulates does not use specialized mathematical terminology. But to really understand the meaning and significance of the postulates, it helps to have some facility with more specialized terminology and to be familiar with some related mathematical concepts. In this unit, we will study some of the basic concepts and terminology of "transformations of the plane", and then go on to see how these tools will allow us to better understand the distance and ruler postulates.

2. Transformations of the Plane

2.1. Introduction: Points, Vectors, and Transformations

When discussing transformations, we will use the concepts of points and vectors. The two have a lot in common—they both involve pairs of real numbers, for example, and can be represented on a grid—and we will often switch from talking about one to talking about the other with no explicit indication. And in fact, we will use somewhat casual notation, a shorthand that blurs the distinction between the two. All mathematicians use this casual notation: completely precise notation would be cumbersome to write and tiring to read. But it is important that everyone understand what the shorthand is really short for. In this section, we will define points and vectors, make precise their relationship and some of the ways in which we will mix and switch the two, and introduce the casual notation that will be used throughout the rest of the unit.

Definition 1 the cartesian plane, points

- Words: the *cartesian plane*
- Symbol: $\mathbb{R}^2$
- Meaning: the set of ordered pairs of real numbers, $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$
- Additional notation: elements of the cartesian plane are called *points*. That is, a *point* is an ordered pair of real numbers.

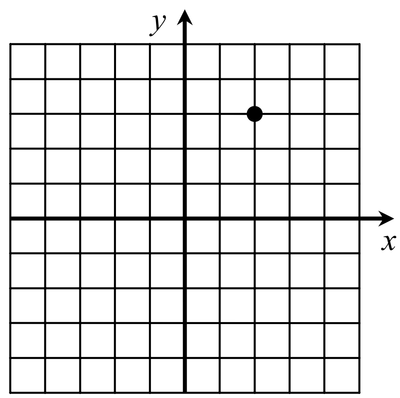
Remark: There is already some subtlety here, in what has *not* been said. Namely, there are no algebraic operations associated with the cartesian plane. No addition, no scalar multiplication, no multiplication of any sort. If you think of points as locations, then this is no big deal. For instance, it would not make any sense to speak of "Cleveland plus Toledo", or "five times Nelsonville". The plane is just a set. But later we will do things that seem to look like addition of points, or scalar multiplication of points. It will be important to realize that in those situations, the algebraic operations really involve vectors

Definition 2 the vector space $\mathbb{R}^2$

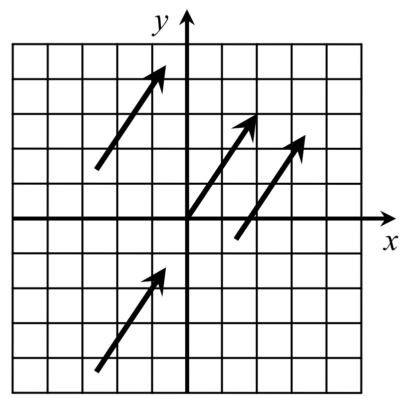
- Symbol: The symbol $\mathbb{R}^2$ is common, but a more correct symbol is $[\mathbb{R}^2, +, \cdot]$, or something similar, that indicates that there are algebraic operations associated with the set.
- Meaning: the set of ordered pairs of real numbers, along with the operations of vector addition and scalar multiplication

- Additional notation: elements of the vector space $\left[\mathbb{R}^2,+,\cdot\right]$ are called *vectors*, and are denoted by the symbols $\langle v_1,v_2\rangle$ or $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, the choice depending on typesetting concerns, mainly. Either way, note that the brackets are brackets, *not* parentheses.
- Vector addition is defined by $\langle v_1,v_2\rangle+\langle w_1,w_2\rangle=\langle v_1+w_1,v_2+w_2\rangle$.
- Scalar multiplication is defined by $c\langle v_1,v_2\rangle=\langle cv_1,cv_2\rangle$, for $c\in\mathbb{R}$.

Remark: Both points and vectors can be visualized on a set of axes, but in different ways. Points are visualized as dots, and are not moveable. Vectors are visualized as arrows, and float. That is, the vector $\langle v_1,v_2\rangle$ can be visualized as an arrow extending from any point (x,y) to the point $(x+v_1,y+v_1)$.



There is just one way to represent the *point* $(2,3)$.



There are many ways to represent the *vector* $\langle 2,3\rangle$.

We will often "think of" points as vectors, and vice-versa, in an obvious way. This sounds vague, but it can be made precise by saying that there is a bijection from the set of points to the set of vectors. A bijection, of course, is a 1-to-1, ONTO function. We won't give this function a name, but we could give it a symbol:

Definition 3 the obvious map from the set of points to the set of vectors

- Symbol: $\overrightarrow{(\ )}$ (The arrow reminds us that we're turning something into a vector.)
- Meaning: The function $\overrightarrow{(\ )}:\mathbb{R}^2\to\left[\mathbb{R}^2,+,\cdot\right]$ defined by the rule $\overrightarrow{(x,y)}=\langle x,y\rangle$

For example, $\overrightarrow{(2,3)}=\langle 2,3\rangle$. As another example, let $P\in\mathbb{R}^2$ be some point. Then the symbol $\vec{P}$ represents the associated vector. Note that as the above pictures show, there is just one way to represent the point $(2,3)$, but there are many ways to represent the vector $\langle 2,3\rangle$. The same goes for P and $\vec{P}$. The vector $\vec{P}$ floats; the point P does not.

The $\overrightarrow{(\ )}$ symbol just introduced is widely used, even though it has no name and nobody bothers to be explicit about exactly what it means. Notice that we have defined it using the terminology of functions. It is not hard to see that it is a one-to-one and onto function, and so it has an inverse function. But the inverse function, also frequently used, is usually given neither a name nor a symbol. We will introduce a symbol for it, just so that we can talk about it for awhile, but we will not use the symbol in practice.

Definition 4 the obvious map from the set of vectors to the set of points

- Symbol: $\bullet(\ )$ (The dot reminds us that we're turning something into a point.)
- Meaning: The function $\bullet(\ ):\left[\mathbb{R}^2,+,\cdot\right]\to\mathbb{R}^2$ defined by the rule $\bullet(\langle v_1,v_2\rangle)=(v_1,v_2)$
- Visual interpretation: Given the vector $\langle v_1,v_2\rangle$ as input to this function, place the vector so that its initial point (the point located at the tail) is at the origin. Then the terminal point (the point located at the head of the vector) is the output of the function.

For example, $\bullet(\langle 2,3\rangle)=(2,3)$, and $\bullet\left(\begin{bmatrix}-5\\7\end{bmatrix}\right)=(-5,7)$.

As I said, nobody uses this symbol. But frequently we will "think of" a vector as a point. When we do that, realize that it is possible to be completely precise about the process.

There is a common way of building a vector from *two* points. The symbol used to denote the process is the same as the symbol that we used above for the building of a vector from one point.

Definition 5 the vector defined by two points

- Symbol: $\overrightarrow{PQ}$
- Usage: P and Q are points
- Meaning: The vector with initial point P and terminal point Q.
- Meaning in symbols: If $P=(x,y)$ and $Q=(w,z)$, then $\overrightarrow{PQ}$ is the vector $\langle w-x,z-y\rangle$.
- Observation: $\overrightarrow{PQ}=\vec{P}-\vec{Q}$
- Remark: Realize that although we may visualize the vector $\overrightarrow{PQ}$ as the particular arrow with initial point P and terminal point Q, this vector floats like any other.

There is also a way of using a given point and a given vector to manufacture a new point. The process is called *translation*.

Definition 6 translation of a point by a vector

- Symbol: $P+\vec{v}$
- Spoken: The translation of P by $\vec{v}$.
- Usage: $P=(x,y)$ is a point and $\vec{v}=\langle v_1,v_2\rangle$ is a vector.
- Meaning: The point $(x+v_1,y+v_2)$
- Additional notation: the vector $\vec{v}$ is called the translation vector.

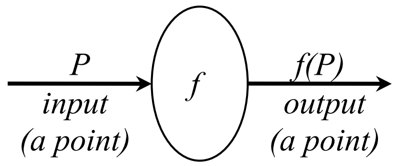
Observe that we can use our earlier notation to describe what happens when we translate a point by a vector. To accomplish the translation of P by $\vec{v}$, we must first think of point P as a vector, then add that vector to the vector $\vec{v}$, and then finally think of the resulting vector as a point. In symbols,

$P+\vec{v}=\bullet\left(\vec{P}+\vec{v}\right)$

We will spend the next few weeks discussing transformations of the plane. They are defined as follows.

Definition 7 map of the plane, transformation of the plane

- Words: "f is a map of the plane" or "f is a transformation of the plane"
- Symbol: $f:\mathbb{R}^2 \to \mathbb{R}^2$
- Meaning: f is a function from $\mathbb{R}^2$ to $\mathbb{R}^2$.
- Machine diagram:



We should think of maps (transformations) of the plane as machines that take points as input and produce points as output.

2.2. Properties of Transformations

In this section, we will define some vocabulary that is useful when discussing maps of the plane.

Definition 8 fixed point

- Words: "P is a fixed point of f", or "f fixes P"
- Usage: f is a map of the plane and P is a point of the plane
- Meaning: $f(P) = P$
- Meaning in words: To say that *P* is a fixed point of f means that when the point *P* is used as input to the function f, the resulting output is also point *P*.
- Special case: to say "f fixes the origin" means that $f(0,0) = (0,0)$.

Definition 9 bijection

- Words: "f is a bijection"
- Usage: f is some function
- Meaning: f is one-to-one and onto.

Note that transformations are not assumed to be bijections. That must be specified explicitly.

Example of a bijective map: The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y) = (3x-5, 2y+1)$ is a bijection.

Proof

step 1: Show that f is one-to-one

For any points (x_1, y_1) and (x_2, y_2), suppose that $f(x_1, y_1) = f(x_2, y_2)$. This means that $(3x_1 - 5, 2y_1 + 1) = (3x_2 - 5, 2y_2 + 1)$ as points, which is equivalent to saying that the following pair of equations is true:

$$\begin{cases} 3x_1 - 5 = 3x_2 - 5 \\ 2y_1 + 1 = 2y_2 + 1 \end{cases}$$

Using algebra, we can show that the following pair of equations is also true:

$$\begin{cases} x_1 = x_2 \\ y_1 = y_2 \end{cases}$$

But this means that $(x_1, y_1) = (x_2, y_2)$ as points.

Step 2: show that f is onto

Suppose $(a,b) \in \mathbb{R}^2$ is some desired output. Consider what happens when we use the point $(x,y) = \left(\frac{a+5}{3}, \frac{b-1}{2}\right)$ as input. The resulting output is

$$f(x,y) = f\left(\frac{a+5}{3}, \frac{b-1}{2}\right) = \left(3\left(\frac{a+5}{3}\right) - 5, 2\left(\frac{b-1}{2}\right) + 1\right) = (a,b).$$

We have shown that there exists some point $(x,y) \in \mathbb{R}^2$ such that $f(x,y) = (a,b)$.

End of Proof

Definition 10 linear map of the plane

- Words: "f is a linear map of the plane", or "f is a linear mapping of the plane",
- Usage: f is some map of the plane
- Meaning: When considered as a map of the vector space $[\mathbb{R}^2, +, \cdot]$, the map f preserves scalar multiplication and vector addition. That is, for all vectors $\vec{v}$ and $\vec{w}$ and for all real numbers *a* and *b*, $f(a\vec{v} + b\vec{w}) = af(\vec{v}) + bf(\vec{w})$.

To understand the equation in the definition of *linear map*, it is helpful to see understand what it means to say that the map preserves scalar multiplication and vector addition. We start by considering scalar multiplication.

Suppose that f is a map of the plane, and that a vector $\vec{v}$ and a number $a \in \mathbb{R}$ are given. There are two natural processes that use these items to produce a new vector.

Process 1: First multiply the vector $\vec{v}$ by the number *a* to obtain a new vector $a\vec{v}$, then use this new vector as input to the map f to obtain the output vector $f(a\vec{v})$.

Process 2: First use the vector $\vec{v}$ as input to the map f to obtain the output vector $f(\vec{v})$, then multiply this output vector by the number *a* to obtain a new vector $af(\vec{v})$.

If the results of these two processes are the same, that is if $f(a\vec{v}) = af(\vec{v})$, then we say that *the map f preserves scalar multiplication*.

Now we will consider vector addition. Suppose that f is a map of the plane, and that vectors $\vec{v}$ and $\vec{w}$ are given. There are two natural processes that use these items to produce a new vector.

Process 1: First add the vectors $\vec{v}$ and $\vec{w}$ to obtain a new vector $\vec{v} + \vec{w}$, then use this new vector as input to the map f to obtain the output vector $f(\vec{v} + \vec{w})$.

Process 2: First use the vector $\vec{v}$ as input to the map f to obtain the output vector $f(\vec{v})$ and use the vector $\vec{w}$ as input to the map f to obtain the output vector $f(\vec{w})$, then add these two output vectors to obtain a new vector $f(\vec{v}) + f(\vec{w})$.

If the results of these two processes are the same, that is, if $f(\vec{v} + \vec{w}) = f(\vec{v}) + f(\vec{w})$, then we say that *the map f preserves vector addition*.

Now we see that the equation $f\left(a\vec{v}+b\vec{w}\right)=af\left(\vec{v}\right)+bf\left(\vec{w}\right)$ in the definition of linear map expresses the fact that f simultaneously preserves both scalar multiplication and vector addition.

Example of a linear map
Consider the map $f:\mathbb{R}^2\to\mathbb{R}^2$ defined by $f\left(x,y\right)=\left(3x,2y\right)$. Let $\vec{v}=\langle x_1,y_1\rangle$ and $\vec{w}=\langle x_2,y_2\rangle$ be two vectors, and let $a,b\in\mathbb{R}$ be two scalars. The following calculation tells us that f is a linear map.

$f\left(a\langle x_1,y_1\rangle+b\langle x_2,y_2\rangle\right)$	$=$	$f\left(\langle ax_1,ay_1\rangle+\langle bx_2,by_2\rangle\right)$	definition of scalar multiplication
	$=$	$f\left(\langle ax_1+bx_2,ay_1+by_2\rangle\right)$	definition of vector addition
	$=$	$\langle 3\left(ax_1+bx_2\right),2\left(ay_1+by_2\right)\rangle$	definition of f
	$=$	$\langle 3ax_1+3bx_2,2ay_1+2by_2\rangle$	distributive law
	$=$	$\langle 3ax_1,2ay_1\rangle+\langle 3bx_2,2by_2\rangle$	definition of vector addition
	$=$	$a\langle 3x_1,2y_1\rangle+b\langle 3x_2,2y_2\rangle$	definition of scalar multiplication
	$=$	$af\langle x_1,y_1\rangle+bf\langle x_2,y_2\rangle$	definition of f

It is not hard to show that any linear map of the plane corresponds to left multiplication by some 2×2 matrix of real numbers. For example the map f that we just discussed corresponds to left multiplication by the matrix $M=\begin{pmatrix}3&0\\0&2\end{pmatrix}$. That is $\begin{bmatrix}3&0\\0&2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}3x\\2y\end{bmatrix}=f\left(\begin{bmatrix}x\\y\end{bmatrix}\right)$. Linear maps are usually studied by studying the corresponding matrices. Such study is the focus of linear algebra courses, but we will have a dose of it in these notes, as well.

The next batch of terminology to be introduced has to do with the relationship between maps of the plane and "distance functions". To start, we must introduce the concept of distance functions.

Definition 11 distance function

- Words: d is a distance function
- Meaning: d is a function $d:\mathbb{R}^2\times\mathbb{R}^2\to[0,\infty)$ with the following properties:
  1. (the *positive definite* property) $d\left(P,Q\right)=0$ if and only if P and Q are the same point.
  2. (the *symmetric* property) For any $P,Q\in\mathbb{R}^2$, $d\left(P,Q\right)=d\left(Q,P\right)$.
  3. (the *triangle inequality*) For any $P,Q,R\in\mathbb{R}^2$, $d\left(Q,R\right)\le d\left(P,Q\right)+d\left(Q,R\right)$.
- Additional terminology: The quantity $d\left(P,Q\right)$ is called the *distance betwen P and Q*.

Observe that the distance function takes as input a pair of points and produces as output a non-negative real number.

Definition 12 isometry, distance preserving

- Words: "f is an isometry of the plane", or "f is an isometry", or "f is distance preserving"
- Usage: f is a map of the plane, and there is some distance function in use, which we will call d.
- Meaning: for all points P and Q in $\mathbb{R}^2$, $d\left(P,Q\right)=d\left(f\left(P\right),f\left(Q\right)\right)$

- Meaning in words: When two points *P* and *Q* are used as input to the map f, the corresponding two outputs are denoted $f(P)$ and $f(Q)$. To say that *f* is an isometry means that the distance between two points that are used as inputs will always be the same as the distance between the two points that are the resulting outputs.
- Additional terminology: Because of the behavior just described, isometries are said to *preserve distance*.

In the exercises, you will prove the following important theorem.

Theorem 1 Given any distance function *d*, the composition of any two isometries is another isometry.

The distance function that we will be most interested in is the Euclidean distance function.

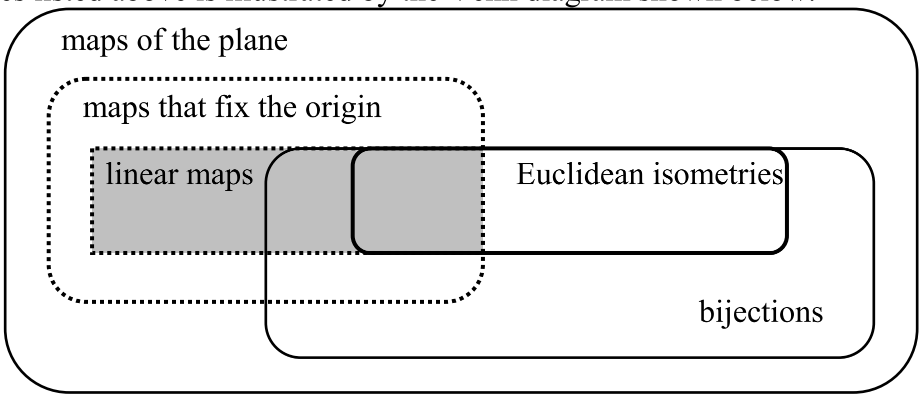
Definition 13 Euclidean distance
- Words: the Euclidean distance function
- Meaning: the function $d_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\infty)$ defined by $d_2((x,y),(w,z)) = \sqrt{(w-x)^2 + (z-y)^2}$.
- Additional terminology: If *P* and *Q* are points in $\mathbb{R}^2$, then the quantity $d_2(P,Q)$ is called the *Euclidean distance betwen P and Q*.
- Additional notation: Maps that preserve the Euclidean distance function are called *Euclidean isometries*.

Example of a Euclidean isometry
Consider the map $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y) = (x+7, y-3)$. Let $\vec{v} = \langle x_1, y_1 \rangle$ and $\vec{w} = \langle x_2, y_2 \rangle$ be two vectors. The following calculation tells us that f is a Euclidean isometry.

$d_2(f\langle x_1,y_1\rangle, f\langle x_2,y_2\rangle)$	$=$	$d_2(\langle x_1+7, y_1-3\rangle, \langle x_2+7, y_2-3\rangle)$	definition of f
	$=$	$\sqrt{((x_1+7)-(x_2+7))^2 + ((y_1-3)-(y_2-3))^2}$	definition of Euclidean distance d_2
	$=$	$\sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$	algebra
	$=$	$d_2(\langle x_1,y_1\rangle, \langle x_2,y_2\rangle)$	definition of Euclidean distance d_2

Observe that some of the terms defined above are actually *properties* that a map of the plane may or may not have. We will be particularly interested in the following four properties, and maps that do—or don't—have them.
- "*f* fixes the origin "
- "*f* is a bijection "
- "*f* is a linear mapping"
- "*f* is a Euclidean isometry"

One might expect that maps of the plane could have any combination of the four properties listed above. That would mean that there would be $2^4 = 16$ different combinations of the four properties that functions might or might not have. But in fact, the various properties are not independent of one another. For instance, it is not difficult to show that if a map *f* is a Euclidean isometry then *f* is also bijection. The converse is not true: there are certainly maps that are bijections but not Euclidean isometries. It is also not difficult to show that if f is linear then f must also fix the origin. Again, the converse is not true: there are maps that fix the origin but are not linear. But it turns out that if a map f preserves Euclidean distance and fixes the origin, then it must be linear. (This is not easy to prove.) The relationship between the four properties listed above is illustrated by the Venn diagram shown below.



You will explore this diagram in the exercises.

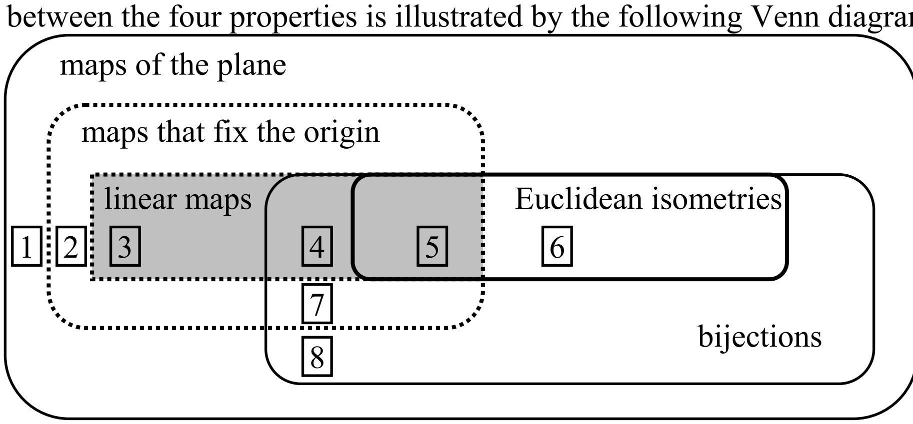
2.3. Exercises

[1] Give a short proof, with no algebra, of Theorem 1. Notice the placement of the theorem in the notes: it comes immediately after the definitions of distance and isometry, before any examples of either. This is no accident. The theorem is true for all distance functions and all isometries. Your proof should not assume anything specific about the distance functions and isometries involved.

[2] In the reading, we were introduced to four properties that mappings of $\mathbb{R}^2$ may or may not have:

- "f fixes the origin "
- "f is a bijection "
- "f is a linear mapping"
- "f is a Euclidean isometry"

The relationship between the four properties is illustrated by the following Venn diagram.



Decide where each of the following maps belongs on the Venn diagram. That is, determine which properties each map has and then determine which one of the eight regions in the diagram the map belongs in. Give the number of that region. Show your work and explain your answers.

a) The map $\pi_1 : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\pi_1(x, y) = (x, 0)$ (This map is called *projection onto the first coordinate*. The symbol pi is used because of the letter p.)

b) The map $\pi_2 : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\pi_2(x, y) = (0, y)$ (This map is called *projection onto the second coordinate*.)

c) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,y+1)$

d) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x+a,y+b)$, where a and b are constants.

e) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,2y)$

f) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,2y+1)$

g) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,y^2)$

h) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,y^2+1)$

i) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,y^3)$

j) The map $f:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y)=(x,y^3+1)$

[3] In the reading, it was mentioned that any linear map of the plane can be represented by a 2×2 matrix. In exercise [2], you should have found that some of the maps are linear. For each of the maps that are linear, give the corresponding matrix.

3. Three important families of transformations

3.1. Translations, Reflections, and Rotations

Three transformations that we will be studying this week are called translations, reflections, and rotations. Here are the definitions.

Definition 14 translation of the plane by a vector

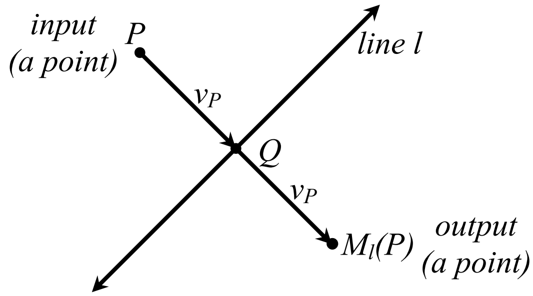
- Symbol: $T_{\vec{v}}$
- Spoken: The translation by $\vec{v}$.
- Usage: $\vec{v}=\langle v_1,v_2\rangle$ is a vector.
- Meaning: The function $T_{\vec{v}}:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $T_{\vec{v}}(P)=P+\vec{v}$
- Additional notation: the vector $\vec{v}$ is called the translation vector.

In other words, translation of the plane works by translating every point in the plane by the same translation vector.

Definition 15 reflection of the plane across a line

- Symbol: M_l (The M is chosen to make us think of a mirror reflection.)
- Spoken: The reflection with axis l.
- Usage: l is a line, called the *axis of reflection*.
- Meaning: The function $M_l:\mathbb{R}^2 \to \mathbb{R}^2$ defined in the following way. If P is a point to be used as input to the function, let Q be the foot of the perpendicular dropped from P to l. Denote by $\vec{v}_P$ the vector $\overrightarrow{PQ}$. Then the output of the function is the point $M_l(P)=P+2\vec{v}_P$

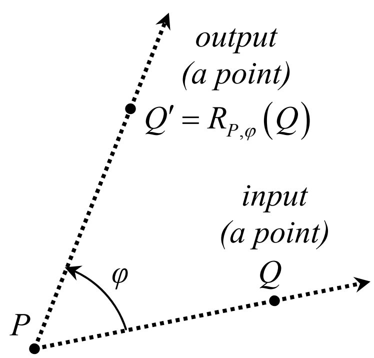
- Picture:



Note that in the definition of the translation of the plane by a vector, the translation vector $\vec{v}$ does not depend on the point being translated. But in the definition of reflection of the plane across a line, the vector $\vec{v}_P$ does depend on the point *P* being reflected.

Definition 16 rotation of the plane

- Symbol: $R_{P,\varphi}$
- Spoken: the (counterclockwise) rotation of the plane about *P* through φ.
- Usage: *P* is a point and φ is an angle.
- Meaning: The function $R_{P,\varphi}:\mathbb{R}^2 \to \mathbb{R}^2$ defined as follows. When point $Q \in \mathbb{R}^2$ is used as input, the resulting output is the point $Q' = R_{P,\varphi}(Q) \in \mathbb{R}^2$ that has the following two properties:
  - The rays $\overrightarrow{PQ}$ and $\overrightarrow{PQ'}$ form an angle congruent to φ, with $\overrightarrow{PQ'}$ oriented counter clockwise from $\overrightarrow{PQ}$.
  - $\overline{PQ} \cong \overline{PQ'}$
- Picture:



3.2. Relationships between Translations, Reflections, and Rotations

3.2.1. Alternate description of reflections across lines not through the origin

The definition of the reflection of the plane across a line did not assume that the line went through the origin. It is interesting to note that any reflection of the plane across a line *m* that does not go through the origin can be given an alternate description involving a reflection across a line *n* that does go through the origin, but the alternate description will also involve translations. To see how, pick some point *P* on the line *m*. (*Any* point.) Then the symbol $\overrightarrow{P}$ stands for the vector that can be represented by an arrow drawn from the origin to point *P*. (See Definition 3) The translation $T_{\overrightarrow{P}}$ will move the origin to point *P*, and the translation $T_{-\vec{P}}$ will move the point *P* to the origin. Define a new line *n* by $n = T_{-\vec{P}}(m)$. Then line *n* will be parallel to *m*, but line *n* will go through the origin. Now observe that if we want to reflect some point *Q* across line *m*, we can accomplish the reflection by first moving the point *Q* with the translation $T_{-\vec{P}}$, then flipping the translated point $T_{-\vec{P}}(Q)$ point across line *n* using the reflection M_n, then moving the resulting point $M_n\left(T_{-\vec{P}}(Q)\right)$ with the translation $T_{\vec{P}}$. That is,

$$M_m(Q) = T_{\vec{P}}\left(M_n\left(T_{-\vec{P}}(Q)\right)\right) = T_{\vec{P}} \circ M_n \circ T_{-\vec{P}}(Q).$$

In other words, $M_m = T_{\vec{P}} \circ M_n \circ T_{-\vec{P}}$.

3.2.2. Alternate description of rotations around a point P that is not the origin

There is an analogous alternate description for rotations. Note that definition of the rotation of the plane around a point *P* did not assume that point *P* was the origin. But any rotation of the plane around a point *P* that is not the origin can be given an alternate description involving a rotation of the plane around the origin, and the alternate description will also involve translations. The description works as follows. The symbol $\vec{P}$ stands for the vector that can be represented by an arrow drawn from the origin to point *P*. The translation $T_{\vec{P}}$ will move the origin to point *P*, while $T_{-\vec{P}}$ will move the point *P* to the origin. Observe that if we want to rotate some point *Q* around point *P*, we can accomplish the same thing by instead doing a three-step process. First, we move the point *Q* with the translation $T_{-\vec{P}}$. The resulting translated point is $Q' = T_{-\vec{P}}(Q)$. We then rotate this point Q' around the origin using the rotation $R_{(0,0),\varphi}$. The resulting rotated point is $Q'' = R_{(0,0),\varphi}\left(T_{-\vec{P}}(Q)\right)$. We then translate the point Q'' with the translation $T_{\vec{P}}$. The resulting point is $Q''' = T_{\vec{P}}\left(R_{(0,0),\varphi}\left(T_{-\vec{P}}(Q)\right)\right) = T_{\vec{P}} \circ R_{(0,0),\varphi} \circ T_{-\vec{P}}(Q)$. This final point Q''' is the same point that we would have gotten if we had simply rotated our original point *Q* around point *P* through an angle *φ*. In other words, $R_{P,\varphi}(Q) = T_{\vec{P}} \circ R_{(0,0),\varphi} \circ T_{-\vec{P}}(Q)$. Since this works regardless of the point *Q*, we can rewrite the expression keeping just the functions, without reference to *Q*. That is, as functions, $R_{P,\varphi} = T_{\vec{P}} \circ R_{(0,0),\varphi} \circ T_{-\vec{P}}$.

3.2.3. A connection between translations and reflections

The following theorem and its corollary establish a connection between translations and reflections.

Theorem 2 Let *l* and *m* be parallel lines. Then the mapping $M_m \circ M_l$ is a translation.
Proof:

To show that $M_m \circ M_l$ does qualify as a translation, we will have to demonstrate that there is some vector $\vec{v}$ such that $M_m \circ M_l$ does translate all the points of the plane by that same vector. That is, for every $P \in \mathbb{R}^2$, $M_m \circ M_l(P) = P + \vec{v}$. The vector $\vec{v}$ must not depend on the point *P*. Then we can say that the mapping $M_m \circ M_l$ is actually just the translation mapping, $T_{\vec{v}}$.

Suppose that *P* is given, that $M_l(P) = P'$ where $P' = P + 2\vec{w}_P$, and $M_m(P) = P'' = P' + 2\vec{z}_{P'}$. Then

$$\begin{aligned} M_m \circ M_l(P) &= M_m(M_l(P)) \\ &= M_m(P') \\ &= P' + 2\vec{z}_{P'} \\ &= (P + 2\vec{w}_P) + 2\vec{z}_{P'} \\ &= P + 2(\vec{w}_P + \vec{z}_{P'}) \end{aligned}$$

We claim that the vector $2(\vec{w}_P + \vec{z}_{P'})$ does not depend on the choice of *P* and hence is the sought- after vector $\vec{v}$. We prove this by showing the vector $\vec{w}_P + \vec{z}_{P'}$ always will have the following two properties, regardless of the choice of the point *P*:

1) The vector $\vec{w}_P + \vec{z}_{P'}$ is perpendicular to both lines *l* and *m*.
2) If the tail of the vector is placed on line *l*, then the head of the vector will lie on line *m*.

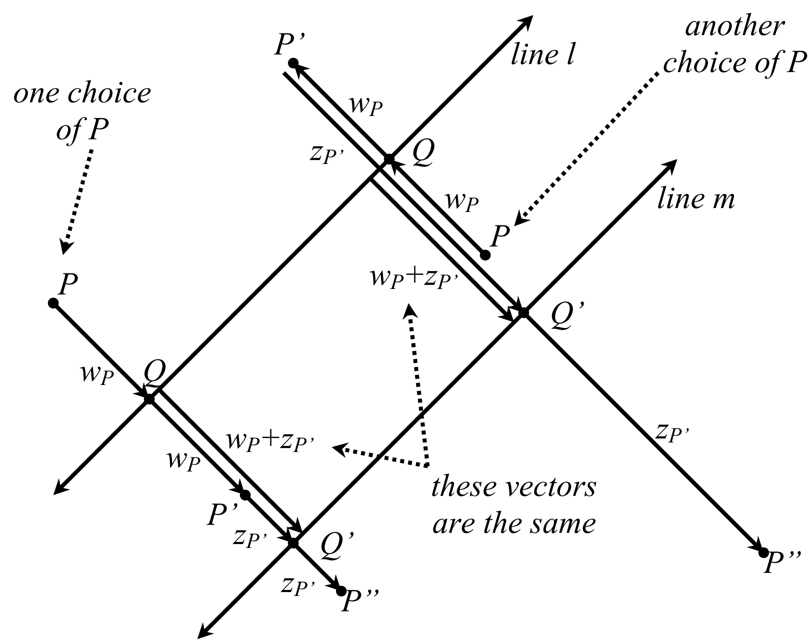
Any two vectors that have both of these properties can be floated to lie on top of each other, which indicates that they actually are the same vector.

To see that the vector $\vec{w}_P + \vec{z}_{P'}$ has property (1), remember that vector $\vec{w}_P$ was constructed to be perpendicular to line *l*, and the vector $\vec{z}_{P'}$ was constructed to be perpendicular to line *m*. Since the lines *l* and *m* are parallel, the two vectors $\vec{w}_P$ and $\vec{z}_{P'}$ point in the same direction, and their vector sum $\vec{w}_P + \vec{z}_{P'}$ will be perpendicular to both lines *l* and *m*.

To see that the vector $\vec{w}_P + \vec{z}_{P'}$ has property (2), let point Q be the point that is the foot of the perpendicular dropped from *P* to *l*, and let Q' be the point that is the foot of the perpendicular from P' to *m*. Because $P' = M_l(P) = P + 2\vec{w}_P$, we can imagine two copies of the vector $\vec{w}_P$: one copy with its tail on *P* and its head on *Q*, and another copy with its tail on *Q* and its head on P' . Similarly, because $P'' = M_m(P') = P' + 2\vec{z}_{P'}$, we can imagine two copies of the vector $\vec{z}_{P'}$: one copy with its tail on P' and its head on Q' , and another copy with its tail on Q' and its head on P'' . Now consider the configuration of the second copy of the vector $\vec{w}_P$ and the first copy of the vector $\vec{z}_{P'}$. They are placed head-to-tail, so an arrow drawn with tail at the tail of $\vec{w}_P$ and head at the head of $\vec{z}_{P'}$ represents the vector $\vec{w}_P + \vec{z}_{P'}$. The tail of this arrow is on *Q* (which is on line *l*), and the head is on Q' (which is on line *m*). Therefore, the vector $\vec{w}_P + \vec{z}_{P'}$ has property (2).

In conclusion, we see that $M_m \circ M_l(P) = P + 2(\vec{w}_P + \vec{z}_{P'})$, where the vector $2(\vec{w}_P + \vec{z}_{P'})$ does not depend on the choice of *P* and hence is the sought-after vector $\vec{v}$. That is, we can say that $M_m \circ M_l(P) = T_{\vec{v}}(P)$.

The figure below illustrates the construction described in the proof. Vectors that lie on top of each other have been offset slightly for clarity. Notice that regardless of the choice of input point P, the resulting vector $\vec{v}_P + \vec{z}_{P'}$ will extend from point *Q* (on line *l*) to point Q' (on line *m*), and will be perpendicular to both lines.



End of proof.

The constructions above can essentially be run backwards to prove that the converse of the statement just proved is also true. We present it here as a theorem.

Theorem 3 If $T_{\vec{v}}$ is a translation, then there are parallel lines *l* and *m* such that $T_{\vec{v}} = M_m \circ M_l$.
Proof:

Let *S* be any point in $\mathbb{R}^2$ and let *l* be the line containing *S* perpendicular to $\vec{v}$. Let *m* be the line $m = l + \frac{1}{2}\vec{v} = \left\{ V \mid V = P + \frac{1}{2}\vec{v} \text{ for some } P \in l \right\}$. Note that the proof of the preceding theorem yields $M_m \circ M_l = T_{2\left(\frac{1}{2}\vec{v}\right)} = T_{\vec{v}}$.

End of proof

3.2.4. A connection between rotations and reflections

The following theorem describes a relationship between rotations and reflections. We won't prove it yet, primarily because I don't have a good drawing for it. But in the exercises, you will be asked to do a few graphical examples that should convince you of the theorem's truth.

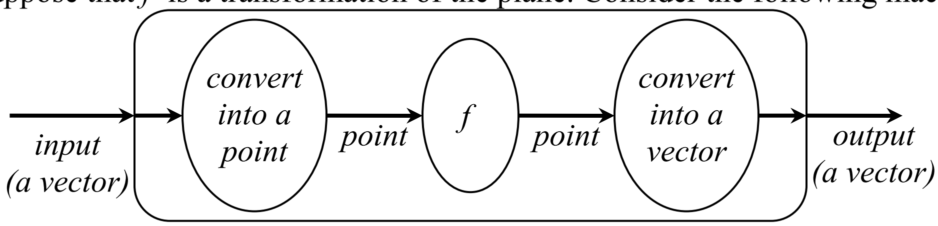
Theorem 4 If *l* and *m* are lines that intersect at point *P* forming an angle φ (measured counterclockwise from line *l* to line *m*), then $M_m \circ M_l = R_{P,2\varphi}$.

3.3. The Relationship between Rotations, Reflections, and Matrices

We will see that rotations about the origin and reflections across lines through the origin have a very nice representation in terms of matrices. Before exploring that, we should be clear about the relationship between transformation of the plane $\mathbb{R}^2$ and transformations of the vector space $\left[\mathbb{R}^2, +, \cdot\right]$.

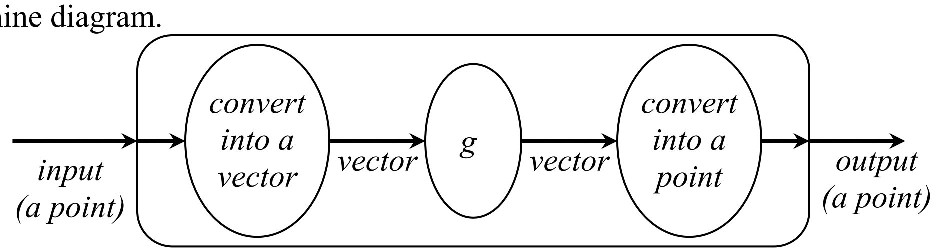
First, recall that a transformation of the plane $\mathbb{R}^2$ takes points as input and produces points as output. A transformation of the vector space $\left[\mathbb{R}^2,+,\cdot\right]$ takes vectors as input and produces vectors as output. But a transformation of either type can be converted to a transformation of the other type.

For example, suppose that f is a transformation of the plane. Consider the following machine diagram.



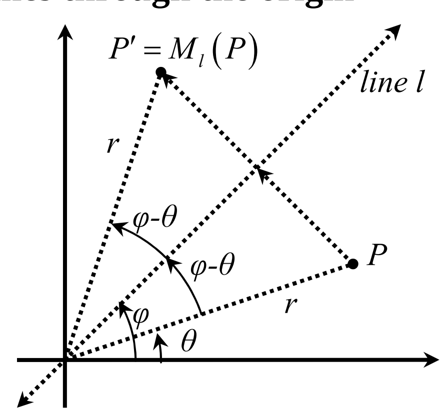
The machine shown takes vectors as input and produces vectors as output. Observe that the conversion functions shown inside of the machine are just the functions $\overrightarrow{(\ )}$ and $\bullet(\ )$ described in Definition 3 and Definition 4. The machine is a transformation of the vector space $\left[\mathbb{R}^2,+,\cdot\right]$. So, given any transformation of the plane $\mathbb{R}^2$, there is an associated transformation of the vector space $\left[\mathbb{R}^2,+,\cdot\right]$.

This works both ways: suppose that g is a transformation of the vector space. Then consider the following machine diagram.



The machine shown takes points as input and produces points as output. As above, the conversion functions shown inside of the machine are just the functions $\overrightarrow{(\ )}$ and $\bullet(\ )$. The machine is a transformation of the plane. So, given any transformation of the vector space $\left[\mathbb{R}^2,+,\cdot\right]$, there is an associated transformation of the plane $\mathbb{R}^2$.

3.3.1. Matrix representation of reflections across lines through the origin



Consider the reflection function $M_l : \mathbb{R}^2 \to \mathbb{R}^2$, where l is a line that makes an angle φ with the positive x-axis. Suppose that an input point P has polar coordinates (r,θ) and rectangular coordinates $(r\cos(\theta), r\sin(\theta))$. Let $P' = M_l(P)$ be the resulting output point. From the picture at right, one can see that the point P' will have an angle of $\theta + (\varphi-\theta) + (\varphi-\theta) = 2\varphi - \theta$. Therefore, point P' will have polar coordinates $(r, 2\varphi-\theta)$ and rectangular coordinates $(r\cos(2\varphi-\theta), r\sin(2\varphi-\theta))$.

Corresponding to the function M_l, which takes points and input and produces points as output, there is an associated function which would use vectors as inputs and outputs. For that function, if the vector

$\begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix}$ is used as input, then the vector $\begin{bmatrix} r\cos(2\varphi-\theta) \\ r\sin(2\varphi-\theta) \end{bmatrix}$ will be the resulting output. The addition formulas for the sine and cosine functions can be exploited to rewrite this output vector as a matrix times the input vector.

$$
\begin{aligned}
\textit{output vector} &= \begin{bmatrix} r\cos(2\varphi-\theta) \\ r\sin(2\varphi-\theta) \end{bmatrix} \\
&= \begin{bmatrix} r\cos(2\varphi)\cos(\theta)+r\sin(2\varphi)\sin(\theta) \\ r\sin(2\varphi)\cos(\theta)-r\cos(2\varphi)\sin(\theta) \end{bmatrix} && \text{by angle addition formulas} \\
&= \begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix} \cdot \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix} && \text{matrix multiplication} \\
&= \begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix} \cdot \textit{input vector}
\end{aligned}
$$

We see that to obtain the output vector, we should left-multiply the input vector by the matrix $\begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos 2(\varphi) \end{bmatrix}$. In other words if a line *l* makes an angle *φ* with the positive *x*-axis, then the following two transformations are associated in the sense described earlier:

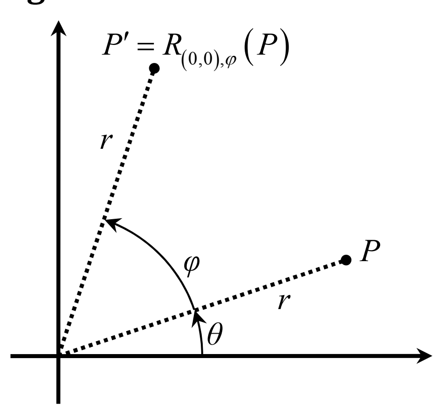
- The function $M_l : \mathbb{R}^2 \to \mathbb{R}^2$ defined as reflection across the line *l*.
- The mapping of $\left[\mathbb{R}^2,+,\cdot\right]$ defined by left multiplication by the matrix $\begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos 2(\varphi) \end{bmatrix}$.

We say that the function M_l is represented by the matrix.

Remark: Because $(\cos(\varphi))^2+(\sin(\varphi))^2=1$, the reflection matrix $\begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix}$ is of the form $\begin{bmatrix} a & b \\ b & -a \end{bmatrix}$, where $a^2+b^2=1$. Conversely, any matrix of the form $\begin{bmatrix} a & b \\ b & -a \end{bmatrix}$, where $a^2+b^2=1$ can be thought of as a reflection matrix. This will be explored in the exercises.

3.3.2. Matrix represetation of rotations about the origin

Consider the reflection function $R_{(0,0),\varphi} : \mathbb{R}^2 \to \mathbb{R}^2$. Suppose that an input point *P* has polar coordinates (r,θ) and rectangular coordinates $(r\cos(\theta), r\sin(\theta))$. Let $P' = R_{(0,0),\varphi}(P)$ be the resulting output point. From the picture at right, one can see that the point P' will have an angle of $\theta+\varphi$. Therefore, point P' will have polar coordinates $(r,\theta+\varphi)$ and rectangular coordinates $(r\cos(\theta+\varphi), r\sin(\theta+\varphi))$.



Corresponding to the function $R_{(0,0),\varphi}$, which takes points and input and produces points as output, there is an associated function which would use vectors as inputs and outputs. For that function, if the vector $\begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix}$ is used as input, then the vector $\begin{bmatrix} r\cos(\theta+\varphi) \\ r\sin(\theta+\varphi) \end{bmatrix}$ will be the resulting output. As in the previous example, the addition formulas for the sine and cosine functions can be exploited to rewrite this output vector as a matrix times the input vector.

$$\begin{aligned} \textit{output vector} &= \begin{bmatrix} r\cos(\theta+\varphi) \\ r\sin(\theta+\varphi) \end{bmatrix} \\ &= \begin{bmatrix} r\cos(\theta)\cos(\varphi) - r\sin(\theta)\sin(\varphi) \\ r\cos(\theta)\sin(\varphi) + r\sin(\theta)\cos(\varphi) \end{bmatrix} && \text{by angle addition formulas} \\ &= \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \cdot \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix} && \text{matrix multiplication} \\ &= \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \cdot \textit{input vector} \end{aligned}$$

We see that to obtain the output vector, we should left-multiply the input vector by the matrix $\begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$. In other words, the following two transformations are associated in the sense described earlier:

- The function $R_{(0,0),\varphi} : \mathbb{R}^2 \to \mathbb{R}^2$ defined as counterclockwise rotation about the origin through the angle φ.
- The mapping of $\left[\mathbb{R}^2, +, \cdot\right]$ defined by left multiplication by the matrix $\begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$.

Again, we will say that the function $R_{(0,0),\varphi}$ is *represented by* the matrix.

Remark: Because $(\cos(\varphi))^2 + (\sin(\varphi))^2 = 1$, the rotation matrix $\begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$ is of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, where $a^2 + b^2 = 1$. This will be explored in the exercises.

3.4. Exercises

[1] Draw a picture showing the lines *l* and *m* that make angles 30° and 60° with the positive *x*-axis, respectively. Draw points *P*, *Q*, *R*, *S*, and *T* in the first quadrant with these properties.

Point *P* is below line *l*.
Point *Q* is on line *l*.
Point *R* is between lines *l* and *m*.
Point *S* is on line *m*.
Point *T* is above line *m*.

Show the outputs that result when each point is used as input to the map $M_m \circ M_l$.

[2] Prove the following:

Theorem 5 Translations are Euclidean isometries.

Hint: Recycle the work that you did for Section 2.3 Exercise [2](d).

[3] Prove the following:

Theorem 6 Reflections across lines through the origin are Euclidean isometries.

[4] Prove the following:

Theorem 7 Rotations around the origin are Euclidean isometries.

[5] Prove the following:

Theorem 8 Reflections across lines are Euclidean isometries.

Hint: Use the alternate description of reflections across lines that do not go through the origin, in conjunction with Theorem 1 and Theorem 5.

[6] Prove the following:

Theorem 9 Rotations are Euclidean isometries.

Hint: Use the alternate description of rotations around points that are not the origin, in conjunction with Theorem 1 and Theorem 5.

[7] Decide where each of the following maps belongs on the Venn diagram shown in Section 2.3. That is, determine which properties each map has and then determine which one of the eight regions in the diagram the map belongs in. Give the number of that region. Show your work and explain your answers.

a) The rotation $R_{(0,0),\varphi}$ counterclockwise, through an angle φ, about the origin
b) The rotation $R_{P,\varphi}$ counterclockwise, through an angle φ, about a point *P* that is not the origin.
c) Reflection M_l across a line *l* that goes through the origin
d) Reflection M_m across a line *m* that does not go through the origin.
e) Translation by a non-zero vector $\vec{v}$

[8] (a) Find the matrix representing a rotation $R_{(0,0),30}$ of 30 degrees.

(b) Let *l* be a the line that makes an angle of 30 degrees with the positive *x*-axis. Find the matrix representing the reflection M_l.

(c) Use the matrices from parts (a) and (b) to find the images of the points $P=(2,5)$, $Q=(-1,3)$, and $S=(-2,3)$ under the transformations $R_{(0,0),30}$ and M_l.

[9] The goal of this problem is to show that if *l* and *m* are two lines through the origin, then $M_m \circ M_l$ is a rotation about the origin.

(a) Let the reflection matrix $\begin{bmatrix}\cos(2\varphi) & \sin(2\varphi)\\ \sin(2\varphi) & -\cos(2\varphi)\end{bmatrix}$ represent M_m and the reflection matrix $\begin{bmatrix}\cos(2\theta) & \sin(2\theta)\\ \sin(2\theta) & -\cos(2\theta)\end{bmatrix}$ represent M_l. Multiply these matrices.

(b) Then use the following angle subtraction identities for sine and cosine to simplify the result from (a).

$$\cos(A-B)=\cos(A)\cos(B)+\sin(A)\sin(B)$$
$$\sin(A-B)=\sin(A)\cos(B)-\cos(A)\sin(B)$$

(c) Show that the result from (b) is actually a rotation matrix.

4. Mapping Diagrams

In math, we spend a lot of time thinking about functions. In many computations, there is more than one function involved, with some functions feeding into others, and some providing alternate ways of producing outputs. It can be difficult just to keep all the various function straight in one's mind: what are their inputs and outputs? Which ones are connected, and how? It is natural to look for effective ways of illustrating such function concepts. One common type of illustration uses arrows to represent functions. Because functions are also called maps, we could refer to these illustrations as *mapping diagrams*. In this chapter, we will learn to read, and create, mapping diagrams. The goal will be to use them to illustrate discussions involving distance functions and isometries. In the next chapter, we will use them to depict what we will call *coordinate functions* for lines.

4.1. Mapping Diagram for a Function

You have already seen the simplest example of a mapping diagram: one that depicts a single function. Recall that the symbol $f : A \to B$, which means that f is a function whose domain is the set *A* and whose codomain is the set *B*. The symbol could be considered the most primitive of mapping diagrams. But because mapping diagrams often need to involve more than one function, it is useful to move the symbol f to a spot right above the arrow:

$$A \xrightarrow{f} B$$

In spite of the fact that such a diagram only relates three pieces of information—the name of the function, its domain, and its codomain—the diagram can be a very powerful illustrating tool. Still, there are ways of enhancing such a diagram so that it conveys more information. One way is to add to the diagram a description of what the map f actually does to produce the output for a given input. A common way of doing this is to put a second arrow on the diagram, parallel to the first. Symbols representing a generic input—an element of the domain—are put at the tail of the arrow. The corresponding output is put at the head. This second arrow is often given a little crossbar at the tail end. Sometimes the second arrow is dotted. For example, consider the map $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 + 3x + 2$. A mapping diagram for this map f could be drawn as shown in the figure below.

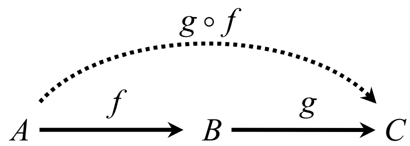
$$\begin{array}{ccc} x & \longmapsto & x^2 + 3x + 2 \\ \mathbb{R} & \xrightarrow{f} & \mathbb{R} \end{array}$$

In some cases, one may want to follow the trajectory of a particular input as it makes its way through a mapping diagram. This can be done by using the same sort of drawing just introduced for generic inputs. For the function just introduced, such a diagram might be as follows.

$$\begin{array}{ccc} 5 & \longmapsto & 42 \\ \mathbb{R} & \xrightarrow{f} & \mathbb{R} \end{array}$$

4.2. Mapping Diagram for Compositions

The composition of functions can be illustrated nicely in a diagram. If f is a function with domain *A* and range *B*, and *g* is a function with domain *B* and range *C*, then the composition $g \circ f$ is a function with domain *A* and range *C*. The mapping symbols for these three functions are $f : A \to B$, $g : B \to C$, and $g \circ f : A \to C$. The mapping diagram below combines these three mapping symbols.



4.3. Products of Maps

When discussing functions, we often consider the two outputs that result from two inputs. It is useful to be able to describe this process as a function, if for no other reason than to be able to illustrate it in a diagram. For that reason, we define the notion of the product mapping. (The name is totally made up: don't use it in public or people will look at you funny.)

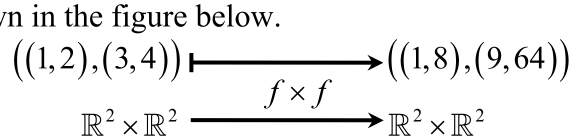
Definition 17 the product mapping

- Symbol: $f \times f$
- Usage: f is a function $f : A \to B$
- Meaning: The function $f \times f : A \times A \to B \times B$ defined by $f \times f\left(\left(a_1, a_2\right)\right) = \left(f\left(a_1\right), f\left(a_2\right)\right)$

Remark: Notice that the "×" symbol is used two different ways in the above definition. When this symbol is encountered, its meaning must be inferred from the context. That is, if the "×" appears between two symbols that stand for sets, then it denotes the Cartesian product of the two sets. On the other hand, if the "×" appears between two symbols that stand for functions, then it denotes the product mapping defined above.

Observe that the domain of $f \times f$ is the set $A \times A$. That is, $f \times f$ takes as input a pair of elements of the set *A*. And note that $f \times f$ produces as output a pair of elements of the set *B*.

For example, consider the map $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = \left(x^2, y^3\right)$. Let $P = (1,2)$ and $Q = (3,4)$. These are both elements of the domain of *f*. When these points are used as inputs, the resulting outputs are $f\left((1,2)\right) = (1,8)$ and $f\left((3,4)\right) = (9,64)$. This information can be conveyed in one expression by using the product map symbol: $f \times f\left(\left((1,2),(3,4)\right)\right) = \left((1,8),(9,64)\right)$. In this expression, the map $f \times f$ is a function $f \times f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2 \times \mathbb{R}^2$ A mapping diagram for this map f could be drawn as shown in the figure below.



As another example, let *l* be some line in the plane, and let $M_l : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection across *l*. When points *P* and *Q* are used as inputs, the resulting outputs are $M_l(P)$ and $M_l(Q)$. This information can be conveyed in one expression by using the product map: $M_l \times M_l\left((P,Q)\right) = \left(M_l(P), M_l(Q)\right)$.

4.4. Mapping Diagrams for Isometries

In this section, we will see that the behavior of an isometry can be nicely illustrated using a mapping diagram. Recall that the term isometry was introduced in Definition 12, in Section 2.2. Consider the equation $d(P,Q) = d(f(P), f(Q))$ that appears in the definition. One can think of this equation as a comparison between two different processes that produce a real number from a pair of points.

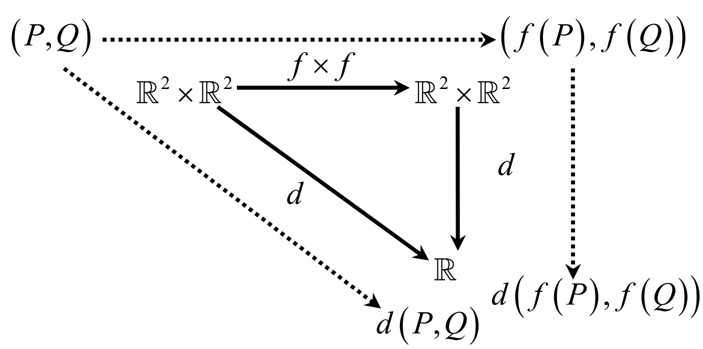
Process 1: Given a pair of points *P* and *Q*, compute the distance between them. This could be illustrated in a diagram as shown.

$$\begin{array}{ccc} (P,Q) & \longmapsto & d(P,Q) \\ \mathbb{R}^2 \times \mathbb{R}^2 & \xrightarrow{\ d\ } & \mathbb{R} \end{array}$$

Process 2: Given a pair of points *P* and *Q*, feed both points into the function f to obtain outputs $f(P)$ and $f(Q)$, and then compute the distance between the two output points $f(P)$ and $f(Q)$. Using the concept of a product mapping, and our method of conveying compositions of maps in diagrams, this processes can be illustrated as shown.

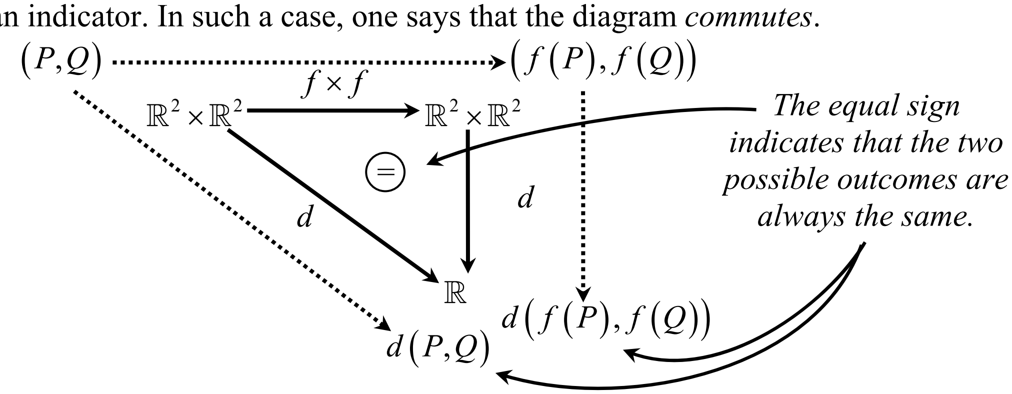
$$\begin{array}{ccccc} (P,Q) & \longmapsto & (f(P), f(Q)) & \longmapsto & d(f(P), f(Q)) \\ \mathbb{R}^2 \times \mathbb{R}^2 & \xrightarrow{\ f \times f\ } & \mathbb{R}^2 \times \mathbb{R}^2 & \xrightarrow{\ d\ } & \mathbb{R} \end{array}$$

The two diagrams showing different ways of getting a real number from a pair of points can be combined into a single diagram, as shown. For clarity, the first diagram has been tilted and the second diagram has been bent.



The diagram shows clearly that there are two different routes that lead from the upper left set, $\mathbb{R}^2 \times \mathbb{R}^2$, to the lower right set, $\mathbb{R}$, resulting in possible outcomes $d(P,Q)$ and $d(f(P), f(Q))$. To say that the function f is an isometry means that these two outcomes are always the same. Such a function is said to *preserve distance*.

Additional terminology and notation: On a mapping diagram, if two different routes from a starting place to a destination always yield the same result, an equal sign is often placed in a circle inside the diagram as an indicator. In such a case, one says that the diagram *commutes*.



4.5. Exercises

[1] Make a mapping diagram to illustrate the fact that the rotation $R_{(0,0),45°}$ (introduced in Definition 16) is a Euclidean isometry. Your diagram should include an arrow that represents the product mapping $R_{(0,0),45°} \times R_{(0,0),45°}$ and two arrows that represent the Euclidean distance function, d_2. It should also include the sets that are the domains and codomains of these functions. Add to your diagram some dotted arrows showing what happens to a particular sample pair of points in $\mathbb{R}^2 \times \mathbb{R}^2$. Use the pair ((1,0),(2,0)) as the sample pair. That is, put ((1,0),(2,0)) above and to the left of the diagram, and show what happens when this pair is used as input.

[2] The Section 4.4 discussion of what it means for a map to be an isometry involved a comparison between the outputs of two mathematical processes. The two processes were then depicted together in a single mapping diagram. In the Section 2.2 discussion following Definition 10, you read that a linear map is one that preserves scalar multiplication and also preserves vector addition. The explanation of what it means to preserve scalar multiplication used a comparison between the outputs of two mathematical processes. So did the explanation of what it means to preserve vector addition. Mapping diagrams can be used to illustrate those explanations.

(a) Make a mapping diagram to illustrate what it means for a map f to preserve scalar multiplication.

(b) Make a mapping diagram to illustrate what it means for a map f to preserve vector addition.

5. Distance Functions

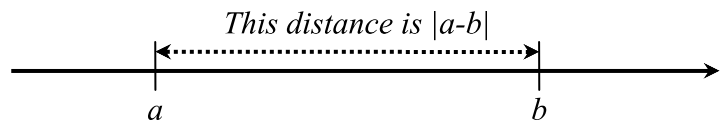
As mentioned in the brief introduction (Chapter 1), one of the goals of this unit is to come to an understanding of a postulate that is often included in the lists of postulates for Euclidean geometry and that pertains to the ability to measure distance. Recall that Definition 11 in Section 2.2 introduced the concept of *distance functions*. The distance function usually plays a very important role in a variety of postulates and definitions related to the geometry. Furthermore, one often answers questions about the properties of a metric geometry by investigating the distance function. In this chapter, we will learn about two particular ways of investigating distance functions.

5.1. Examples of Distance Functions

In this section, we will present distance functions for three different sets: the real numbers, the Cartesian plane, and the Poincare disk model of hyperbolic geometry. We start by introducing the distance function for the real numbers.

5.1.1. The real number distance function

You are used to the idea of visualizing real numbers on a number line, and you are used to the idea that the distance between two spots on the number line is just the absolute value of the difference of the two real numbers.



Realize that that this process actually describes a distance function for the set of real numbers. The function is usually not given a symbol, but we will invent one because we want to distinguish this distance function from the many others that we will be discussing.

Definition 18 the real number distance function

- Meaning: the function $d_{\mathbb{R}}:\mathbb{R}\times\mathbb{R}\to[0,\infty)$ defined by $d_{\mathbb{R}}(a,b)=|a-b|$
- Mapping diagram:

$$\begin{array}{ccc} (a,b) & \longmapsto & |a-b| \\ & d_{\mathbb{R}} & \\ \mathbb{R}\times\mathbb{R} & \longrightarrow & [0,\infty) \end{array}$$

5.1.2. Four distance functions for the Cartesian plane

There is often more than one way to define a distance function for a given set. We will consider four different distance functions for the Cartesian plane, which is just the set $S=\mathbb{R}^2$, the traditional setting for analytic geometry. Their definitions are listed below. In the definitions, let $A=(x,y)$ and $B=(w,z)$.

Definition 19 the discrete distance function

Meaning: the function $disc:\mathbb{R}^2\times\mathbb{R}^2\to[0,\infty)$, defined by $disc(A,B)=\begin{cases}1 \text{ } if \text{ } A\neq B\\ 0 \text{ } if \text{ } A=B.\end{cases}$

Definition 20 the taxi-cab distance function

Meaning: the function $d_1:\mathbb{R}^2\times\mathbb{R}^2\to[0,\infty)$, defined by $d_1(A,B)=|w-x|+|z-y|$.

Definition 21 the Euclidean distance function

Meaning: the function $d_2:\mathbb{R}^2\times\mathbb{R}^2\to[0,\infty)$, defined by $d_2(A,B)=\sqrt{(w-x)^2+(z-y)^2}$.

Definition 22 the worst-case distance function

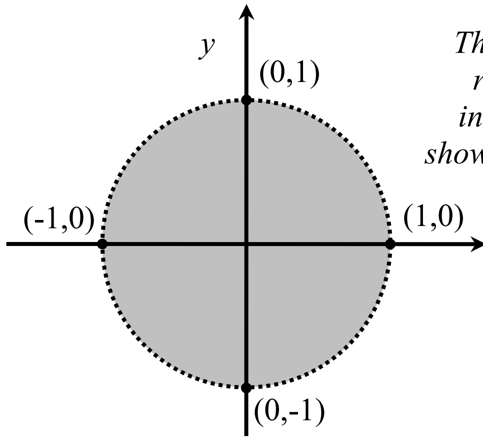
Meaning: the function $d_\infty:\mathbb{R}^2\times\mathbb{R}^2\to[0,\infty)$ defined by $d_\infty(A,B)=\max\{|w-x|,|z-y|\}$.

Notice that each of the above functions is *well-defined*. That is, each of the rules assigns only one non-negative real number to each pair of points. Furthermore, observe that each of the functions does have the three properties required of any distance function.

5.1.3. Hyperbolic distance

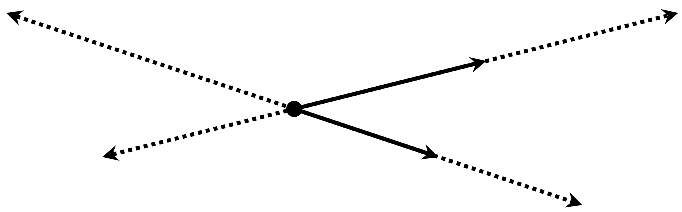
In this section, we will briefly review the objects of hyperbolic geometry, and then introduce a distance function for the geometry. Remember that hyperbolic geometry is described by a postulate list consisting of all of the postulates for absolute geometry, plus the hyperbolic parallel postulate. (This last postulate is just the negation of the statement of the Euclidean parallel postulate.) Therefore, hyperbolic geometry will have points and lines. Because the absolute geometry axioms include the distance function postulate and the ruler postulate, hyperbolic geometry will have a distance function, and lines will have rulers.

Hyperbolic points, called *Hpoints*, are ordered pairs (x,y) in $\mathbb{R}^2$ such that $x^2+y^2<1$. That is, *Hpoints*, are points in $\mathbb{R}^2$ that lie in the interior of the unit circle. The hyperbolic plane, denoted *Hplane*, is the set of all *Hpoints*. In set notation, $Hplane=\{(x,y):x^2+y^2<1\}$.

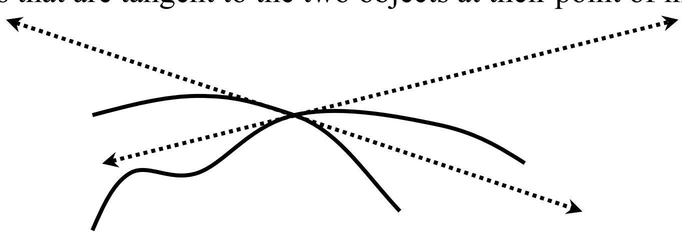


The Hplane is the shaded region. The boundary, including the four points shown, is NOT part of the set.

We need to discuss something called "orthogonal circles". To do that, we must first introduce the idea of the angle of intersection of curvy objects. You are used to the idea of the angles created by two intersecting lines: One simply considers the angle created by a pair of rays that emanate from the intersection point and that lie on the two lines.

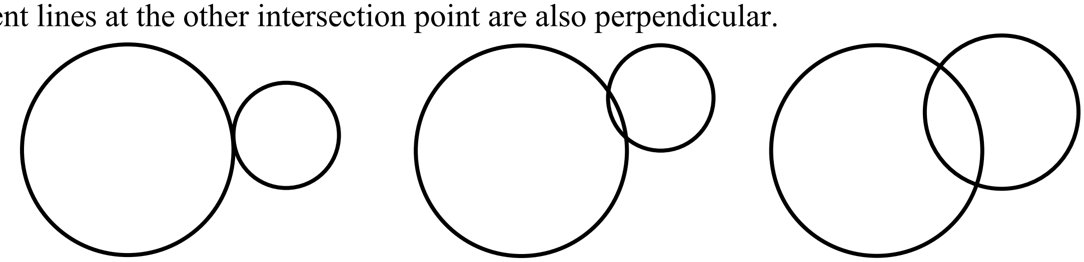


This same idea can be generalized to the intersection of curvy objects. Given two intersecting objects, there are two straight lines that are tangent to the two objects at their point of intersection.



We say that the two curvy objects are perpendicular (at the place where they intercept) if the tangent lines are perpendicular.

Now consider two intersecting circles. The circles may intersect in just one point. But notice that if the circles intersect at more than one point, then the angles created at the two intersection points (created in the sense described above) will be congruent. (We don't prove this, but you might have proven it in Math 330A.) In particular, if the two tangent lines at one intersection point are perpendicular, then the two tangent lines at the other intersection point are also perpendicular.



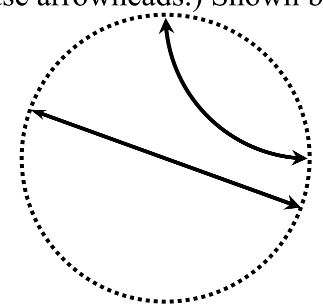
one intersection point | *two intersection points with non-perpendicular tangent lines* | *two intersection points with perpendicular tangent lines*

We will say that a circle is *orthogonal* to the unit circle if the tangent lines at the point of intersection are perpendicular. Such circles will play a key role in our definition of hyperbolic lines.

Hyperbolic lines, denoted *Hlines*, are sets of *Hpoints*. They come in two flavors:

- A "straight-looking" *Hline* is the set of *Hpoints* that lie on a diameter of the unit circle.
- A "curved-looking" *Hline* is the set of *Hpoints* that lie on a circle that is orthogonal to the unit circle.

Remember that because an *Hline* must be made up of *Hpoints*, and all *Hpoints* lie in the interior of the unit circle, an *Hline* (of either flavor) therefore does not include the two "endpoints". We will refer to these as the "missing endpoints". And remember that the missing endpoints are not *Hpoints*, because they are not elements of the set *Hplane*, but they are elements of the set $\mathbb{R}^2$. To indicate that the endpoints are missing, we will put open circles or arrowheads on the ends of our *Hlines*. (My typesetting program won't do open circles, so I'll use arrowheads.) Shown below are some examples of *Hlines*.



We will not prove here that *Hpoints* and *Hlines* as defined above do actually satisfy the postulates for hyperbolic geometry, but we will use those facts, including the fact that given two *Hpoints* *A* and *B*, there is exactly one *Hline* containing both points. This *Hline* will be denoted by $\overleftrightarrow{AB}$.

We now consider a distance function for the hyperbolic plane.

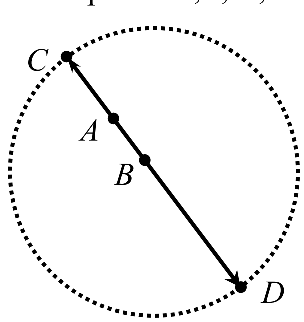
Definition 23 hyperbolic distance

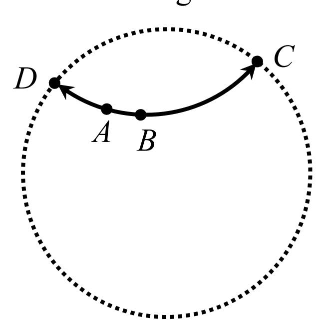
Meaning: the function $d_H : Hplane \times Hplane \to [0,\infty)$ defined by the following formula

$$d_H(A,B) = \left| \ln\left(\frac{AC/AD}{BC/BD} \right) \right|,$$

where *C* and *D* are the "missing endpoints" of the line $\overleftrightarrow{AB}$, and the symbols *AC*, *AD*, *BC*, and *BD* stand for $d_2(A,C)$, $d_2(A,D)$, $d_2(B,C)$, and $d_2(B,D)$.

Two typical configurations of points *A*,*B*, *C*, and *D* are shown in the figures below.





You might be a little disturbed by the fact that in one of these figures, the ordering of points on the line is *C-A-B-D*, while in the other, the ordering is *D-A-B-C*. Shouldn't it matter what names we give to the missing endpoints, at least when using the function d_H? The answer is no, it doesn't matter. To see why, consider the effect interchanging the names of the missing endpoints *C* & *D*. That would mean that we should interchange the symbols *C* & *D* in the formula for the distance between *A* & *B*. But rules of logarithms can be used to show that if *u*, *v*, *x*, and *y* are positive numbers, then

$$\left|\ln\left(\frac{u/v}{x/y}\right)\right| = \left|\ln\left(\frac{v/u}{y/x}\right)\right|.$$

(You will prove this in the exercises.) Therefore, interchanging the symbols *C* & *D* in the formula would have no effect on the outcome.

For example of the use of the distance function d_H, let $A=\left(0,-\frac{1}{4}\right)$ and let $B=\left(,\frac{3}{4}\right)$. Both of these points qualify as *Hpoints*, and they lie on the *Hline* consisting of the portion of the *y*-axis that lies inside the unit circle. The missing endpoints of this line are $C=(0,1)$ and $D=(0,-1)$. To compute the hyperbolic distance between *A* and *B*, we first compute the Euclidean distances needed in the formula for d_H. The results are $d_2(A,C)=\frac{5}{4}$, $d_2(A,D)=\frac{3}{4}$, $d_2(B,C)=\frac{1}{4}$, and $d_2(B,D)=\frac{7}{4}$. Plugging these into the formula for d_H, we obtain

$$d_H(A,B)=\left|\ln\left(\frac{AC/AD}{BC/BD}\right)\right| = \left|\ln\left(\frac{\frac{5}{4}/\frac{3}{4}}{\frac{1}{4}/\frac{7}{4}}\right)\right| = \left|\ln\left(\frac{5\cdot 7}{1\cdot 3}\right)\right| = \ln\left(\frac{35}{3}\right) \approx 2.46$$

Notice that the hyperbolic distance between *Hpoints A* & *B* is not the same as the Euclidean distance between them, which is just $d_2(A,B)=1$.

In the exercises, you will compute some examples of distances between *Hpoints*. And, you will turn the process around, and try to find a point that is a specified distance from a given point. You will also confirm that the function d_H really meet two of the requirements of a distance function.

5.2. Investigating Distance Functions by Considering Their Unit Circles

In this section, and in the exercises, we will begin studying the four distance functions for the Cartesian plane that were introduced in Section 5.1.2. Our study will start by focusing on the unit circles that correspond to each distance function. Here is a definition of the *circle* and of the *unit circle*.

Definition 24 circle

- Words: *the circle of radius r , centered at A.*
- Usage: *A* is a point, *r* is a non-negative real number, and there is some distance function in use, which we denote by the symbol *d*.
- Meaning: the set $\{B : d(A,B)=r\}$
- Additional terminology: The *unit circle* is the circle of radius 1 centered at the origin. That is, the set $\{B : d((0,0),B)=1\}$.

Earlier it was mentioned that in a metric geometry, a variety of definitions will use the distance function. Observe that the distance function, whatever it may be, is part of the definition of the circle. Notice also that there is nothing in the definition that says that a circle is "one of those round-looking things". In fact, circles may look quite strange, depending on the distance function being used.

For example, let us determine the unit circle for the Taxi-cab distance function, introduced in Definition 20. Recall that it is the function $d_1 : \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\infty)$, defined by $d_1(A,B) = |w-x| + |z-y|$. Using this definition of distance, and letting $B = (x,y)$, the unit circle is the set

$$\{(x,y) : d_1((0,0),(x,y)) = 1\} = \{(x,y) : |x-0| + |y-0| = 1\} = \{(x,y) : |x| + |y| = 1\}$$

Right away, we can see that the four points $(1,0)$, $(0,1)$, $(-1,0)$, and $(0,-1)$ are on the Taxi-cab unit circle. To find out about other points on the circle, it will help to write the equation $|x| + |y| = 1$ without the absolute value signs. To do this, however, we need to know something about the signs of the numbers x and y. It is easiest to consider the cases corresponding to the four quadrants.

Case 1: (The 1st quadrant) $x > 0$ and $y > 0$

In this case, $|x| = x$ and $|y| = y$, so the equation defining the unit circle becomes simply $x + y = 1$. Solving this equation for y, we find $y = -x + 1$. In other words, the portion of the taxi-cab unit circle that lies in the first quadrant is just a piece of the line $y = -x + 1$.

Case 2: (The 2nd quadrant) $x < 0$ and $y > 0$

In this case, $|x| = -x$ and $|y| = y$, so the equation defining the unit circle becomes $-x + y = 1$. Solving this equation for y, we find $y = x + 1$. So the portion of the taxi-cab unit circle that lies in the second quadrant is a piece of the line $y = x + 1$.

Case 3: (The 3rd quadrant) $x < 0$ and $y < 0$

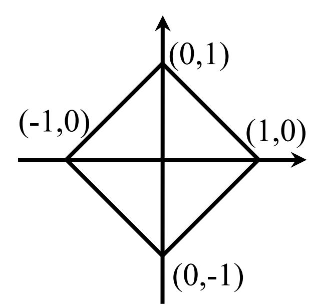
In this case, $|x| = -x$ and $|y| = -y$, so the equation defining the unit circle becomes $-x - y = 1$. Solving this equation for y, we find $y = -x - 1$. So the portion of the taxi-cab unit circle that lies in the third quadrant is a piece of the line $y = -x - 1$.

Case 4: (The 4th quadrant) $x > 0$ and $y < 0$

In this case, $|x| = x$ and $|y| = -y$, so the equation defining the unit circle becomes $x - y = 1$. Solving this equation for y, we find $y = x - 1$. So the portion of the taxi-cab unit circle that lies in the fourth quadrant is a piece of the line $y = x - 1$.

Combining all the information that we have, we see that the taxi-cab unit circle will have the diamond shape shown in the drawing at right.

In the exercises, you will be asked to draw unit circles for the discrete distance function, *disc*, and for the worst-case distance function, d_∞.



unit circle for the taxi-cab distance function

5.3. Investigating Distance Functions by Considering Their Isometries

In this section, and in the exercises, we will continue our study of the four distance functions for the Cartesian plane that were introduced in Section 5.1.2. This time, our study will focus on the isometries for each distance function. Remember that when isometries were introduced Definition 12 in Section 2.2, the definition assumed that some distance function was in use. A map f was deemed to be an isometry for the distance function *d* if the map preserved distance as defined by that distance function. But we have seen that there can be more than one definition of distance for a given set. A map f that preserves one kind of distance will not necessarily preserve another. Determining which maps qualify as isometries for a given distance function gives us a better appreciation of the distance function, and gives us a way to compare different distance functions.

It is not difficult to show that the translation mapping $T_{\vec{v}}$, introduced in Definition 14 in section 3.1, is an isometry for each of the four distance functions that we have used for the Cartesian plane. The fact that translations are Euclidean isometries was the subject of Theorem 5, found in the exercises of Section 3.4. A proof similar to the proof of Theorem 5 can be used to prove the following claim.

Claim: Translations are isometries for the taxi-cab distance function, d_1.
Proof

Let $\vec{v} = \langle v_1, v_2 \rangle$ be any vector. To show $T_{\vec{v}}$ is an isometry for the distance function d_1, we must show that for any points $P = (x, y)$ and $Q = (w, z)$, $d_1(T_{\vec{v}}(P), T_{\vec{v}}(Q)) = d_1(P, Q)$. Our approach is to build the expression on the left side of the equation, and then simplify it.

$d_1(T_{\vec{v}}(P), T_{\vec{v}}(Q))$	$=$	$d_1((x+v_1, y+v_2),(w+v_1, z+v_2))$	definition of $T_{\vec{v}}$, *P*, and *Q*
	$=$	$\lvert (w+v_1)-(x+v_1) \rvert + \lvert (z+v_2)-(y+v_2) \rvert$	definition of d_1
	$=$	$\lvert w-x \rvert + \lvert z-y \rvert$	algebra
	$=$	$d_1(P, Q)$	definition of d_1

End of proof

In the exercises, you will be asked to prove that Translations are isometries for the worst-case distance function d_∞. The proof structure above can be adapted for that proof.

The same proof structure can be used to prove the following claim.

Claim: Translations are isometries for the discrete distance function, *disc*.
Proof #1 (using structure similar to that of the previous proof)
Let $\vec{v} = \langle v_1, v_2 \rangle$ be any vector. To show $T_{\vec{v}}$ is an isometry for the distance function *disc*, we must show that for any points $P = (x, y)$ and $Q = (w, z)$, $d_1(T_{\vec{v}}(P), T_{\vec{v}}(Q)) = d_1(P, Q)$. Our approach is to build the expression on the left side of the equation, and then simplify it.

$disc(T_{\vec{v}}(P), T_{\vec{v}}(Q))$	$=$	$disc((x+v_1, y+v_2),(w+v_1, z+v_2))$	definition of $T_{\vec{v}}$, *P*, and *Q*
	$=$	$\begin{cases} 0 & if \quad (x+v_1, y+v_2) = (w+v_1, z+v_2) \\ 1 & otherwise \end{cases}$	definition of *disc*

$$= \begin{cases} 0 & if \quad x+v_1 = w+v_1 \ and \ y+v_2 = z+v_2 \\ 1 & otherwise \end{cases} \qquad \text{definition of equality of points}$$

$$= \begin{cases} 0 & if \quad x = w \ and \ y = z \\ 1 & otherwise \end{cases} \qquad \text{algebra}$$

$$= \begin{cases} 0 & if \quad P = Q \\ 1 & otherwise \end{cases} \qquad \text{definition of equality of points}$$

$$= disc(P,Q) \qquad \text{definition of } disc$$

End of proof #1

But in fact, there is a much simpler proof available for the same claim, made possible by the simplicity of the definition of the discrete distance function. This proof relies on the fact that the translation $T_{\vec{v}}$ is a one-to-one function. If we use this fact, we can replace the four middle steps of the proof with a single step. Here is the proof.

Proof #2 (using simpler structure)

Let $\vec{v} = \langle v_1, v_2 \rangle$ be any vector. To show $T_{\vec{v}}$ is an isometry for the distance function *disc*, we must show that for any points $P = (x, y)$ and $Q = (w, z)$, $d_1(T_{\vec{v}}(P), T_{\vec{v}}(Q)) = d_1(P,Q)$. Our approach is to build the expression on the left side of the equation, and then simplify it.

$$disc(T_{\vec{v}}(P), T_{\vec{v}}(Q)) = \begin{cases} 0 & if \quad T_{\vec{v}}(P) = T_{\vec{v}}(Q) \\ 1 & otherwise \end{cases} \qquad \text{definition of } disc$$

$$= \begin{cases} 0 & if \quad P = Q \\ 1 & otherwise \end{cases} \qquad \text{because } T_{\vec{v}} \text{ is one-to-one}$$

$$= disc(P,Q) \qquad \text{definition of } disc$$

End of proof

Having just been told that the translation map, $T_{\vec{v}}$, is an isometry for all four of the distance functions *disc*, d_1, d_2, and d_∞, it would be natural to wonder if the four distance functions share other isometries, as well. It turns out that some of the functions that we have seen that are Euclidean isometries are *not* isometries for some of the other distance functions. The following theorem illustrates.

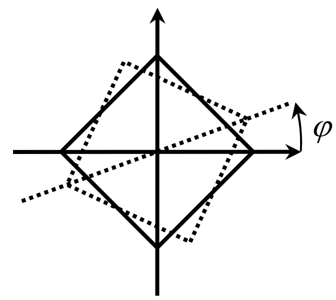
Theorem 10 (Only certain rotations are isometries for the Taxi-Cab distance function.)
The rotation $R_{(0,0),\varphi}$ about the origin is an isometry for the taxi-cab distance function, d_1, only when the angle φ is of the form $\varphi = k \cdot 90^\circ$ for some integer *k*. That is,

$$\varphi \in \{\cdots, -450, -360, -270, -180, -90, 0, 90, 180, 270, 360, 450, \cdots\}$$

Proof:

It would be nice to solve this problem algebraically. An obvious way to start would be to let $A = (u, v)$ and $B = (x, y)$, then write out the expressions for $R_{(0,0),\varphi}(A)$, $R_{(0,0),\varphi}(B)$, $d_1(A,B)$, and $d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$, and then set $d_1(A,B) = d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$. But this quickly turns quite frustrating, and a simple solution does not seem to emerge. Let's instead try to solve the problem in three steps. In the first step, choose a particular pair of points *A* and *B* for which it will be easy to see whether or not $d_1(A,B) = d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$. Using this pair of points, we can narrow our search for angles φ to a smaller list of candidates. In step 2, we will find the matrices that correspond to the angles on our list of candidates. In the third step, we go back to letting *A* and *B* be any two points, and check to see whether or not the angles φ on our list still satisfy $d_1(A,B) = d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$.

Step 1: Use a particular choice of points A and B, and find angles that work.
Let $A = (0,0)$, and let *B* be any random point such that $d_1\left((0,0), B\right) = 1$. In other words, let *B* be any point on the taxi-cab unit circle. Then, look for angles φ such that $d_1\left(R_{(0,0),\varphi}(0,0), R_{(0,0),\varphi}(B)\right) = 1$. But any rotation about the origin fixes the origin, so $R_{(0,0),\varphi}(0,0) = (0,0)$. So we are looking for angles φ such that $d_1\left((0,0), R_{(0,0),\varphi}(B)\right) = 1$. But the only points $R_{(0,0),\varphi}(B)$ that satisfy this equation will be those that lie on the unit circle.



In other words, we want to find a rotation $R_{(0,0),\varphi}$ with the property that if a point on the unit circle is used as input, then the resulting output will also be a point on the unit circle. This can be visualized: we imagine rotating the taxi-cab unit circle, and we look for angles of rotation that will make the resulting shape lie on top of the original shape. We say that such a rotation *preserves the unit circle*.

We see that the only angles for which this will happen are angles φ of the form $\varphi = k \cdot 90^\circ$ for some integer *k*. That is, $\varphi \in \{\cdots, -450, -360, -270, -180, -90, 0, 90, 180, 270, 360, 450, \cdots\}$. What we have shown is if $A = (0,0)$ and *B* is any point on the taxi-cab unit circle, and if φ is any angle in the list above, then $d_1(A,B) = d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$. Since any rotation $R_{(0,0),\varphi}$ that hopes to qualify as an isometry will have to preserve the unit circle, we realize that the list of angles above is a narrowed-down list of candidate angles that might produce rotations that are isometries.

Step 2: Find the matrices that represent the angles on our list of candidates.
In Section 3.3.2, we saw that the rotation $R_{(0,0),\varphi}$ corresponds to the matrix $\begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$.

Because of the periodicity of the trigonometric functions, we know that the matrices corresponding to the angles on our list of candidates will follow a repeating pattern built from the matrices corresponding to the four angles $\varphi = 0, 90, 180, 270$. We compute the matrices corresponding to these four angles.

- The matrix corresponding to the angle $\varphi = 0$ is $\begin{bmatrix} \cos(0) & -\sin(0) \\ \sin(0) & \cos(0) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
- The matrix corresponding to the angle $\varphi = 90$ is $\begin{bmatrix} \cos(90) & -\sin(90) \\ \sin(90) & \cos(90) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.
- The matrix corresponding to the angle $\varphi = 180$ is $\begin{bmatrix} \cos(180) & -\sin(180) \\ \sin(180) & \cos(180) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

- The matrix corresponding to the angle $\varphi = 270$ is $\begin{bmatrix} \cos(270) & -\sin(270) \\ \sin(270) & \cos(270) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

The angles on our list of candidates have matrices of one of the following four types shown above.

Step 3: Let A and B be any two points and see if the angles on our list work.

For the rotation $R_{(0,0),\varphi}$ be an isometry, it will have to satisfy $d_1(A,B) = d_1\left(R_{(0,0),\varphi}(A), R_{(0,0),\varphi}(B)\right)$ for any two points *A* and *B*, not just the special points that we considered in *Step 1*. We will consider each of the four matrix types above.

Matrix type 1: The rotation matrix is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$*, and* $A = \begin{bmatrix} u \\ v \end{bmatrix}$ *and* $B = \begin{bmatrix} x \\ y \end{bmatrix}$ *are any two points.*

In this case, the rotated points are the same as the original points, so the distance is preserved.

Matrix type 2: The rotation matrix is $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$*, and* $A = \begin{bmatrix} u \\ v \end{bmatrix}$ *and* $B = \begin{bmatrix} x \\ y \end{bmatrix}$ *are any two points.*

In this case, the rotated points are $A' = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -v \\ u \end{bmatrix}$ and $B' = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$. Using these, we have

$$d_1(A',B') = d_1\left((-v,u),(-y,x)\right) = |-v-(-y)| + |u-x| = |u-x| + |v-y| = d_1(A,B)$$

so the distance is preserved.

Matrix type 3: The rotation matrix is $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$*, and* $A = \begin{bmatrix} u \\ v \end{bmatrix}$ *and* $B = \begin{bmatrix} x \\ y \end{bmatrix}$ *are any two points.*

In this case, the rotated points are $A' = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -u \\ -v \end{bmatrix}$ and $B' = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ -y \end{bmatrix}$.

Using these, we have

$$d_1(A',B') = d_1\left((-u,-v),(-x,-y)\right) = |-u-(-x)| + |-v-(-y)| = |u-x| + |v-y| = d_1(A,B)$$

so the distance is preserved.

Matrix type 4: The rotation matrix is $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$*, and* $A = \begin{bmatrix} u \\ v \end{bmatrix}$ *and* $B = \begin{bmatrix} x \\ y \end{bmatrix}$ *are any two points.*

In this case, the rotated points are $A' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} v \\ -u \end{bmatrix}$ and $B' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}$. Using these, we have

$$d_1(A',B') = d_1\left((v,-u),(y,-x)\right) = |v-y| + |-u-(-x)| = |u-x| + |v-y| = d_1(A,B)$$

so the distance is preserved.

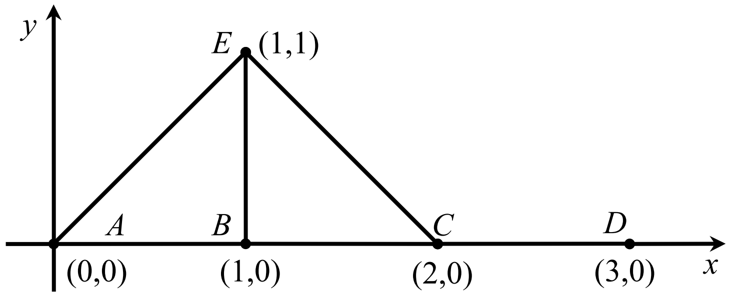
We see that all four cases, the rotations are isometries. That means that for any angle φ is of the form $\varphi = k \cdot 90^\circ$ for some integer *k*, the rotation $R_{(0,0),\varphi}$ is an isometry.

End of proof

In the exercises, you will be asked to find which lines through the origin produce reflections that are isometries for the taxi-cab distance function. The structure of the proof above can be adapted to solve that problem.

5.4. Exercises

[1] In Section 5.1.2, you were introduced to four distance functions for the Cartesian plane: *disc*, d_1, d_2, and d_∞. Using the diagram shown at right, for each of the four distance functions find the following five quantities: $d(A,B)$, $d(A,C)$, $d(A,D)$, $d(A,E)$, $d(B,E)$



[2] In Section 5.1.3, you were told that $\left|\ln\left(\frac{u/v}{x/y}\right)\right| = \left|\ln\left(\frac{v/u}{y/x}\right)\right|$. Prove this.

[3] Let *l* be the *Hline* consisting of the portion of the *x*-axis that lies inside the unit circle. The "missing endpoints" of this line are $R=(-1,0)$ and $S=(1,0)$. (They are not *Hpoints*.) Let *A*, *B*, *C*, and *D* be the *Hpoints* $A=(0,0)$, $B=\left(\frac{1}{4},0\right)$, $C=\left(\frac{1}{2},0\right)$, and $D=\left(\frac{3}{4},0\right)$. Find $d_H(A,B)$ and $d_H(C,D)$. (Be sure to use *R* & *S* as the missing endpoints!)

[4] Compute $d_H((b,0),(0,0))$ where *b* is some unspecified real number such that $0<b<1$.

[5] Given some unspecified positive number *r*, find a number *b* such that $0<b<1$ and $d_H((b,0),(0,0))=r$. (Hint: Think of the result of the previous problem as a function that computes distance as a function of *b*. Call the resulting distance *r*. This gives you an equation involving *b* and *r*. Solve that equation for *b* in terms of *r*. The fact that you know $0<b<1$ enables you deal with the absolute value sign.)

[6] In Definition 23, the function d_H was presented as a distance function for hyperbolic geometry, but it was never verified that the function really does meet the requirements of a distance function. In this exercise, you will verify that d_H meets some of the requirements.
(a) (the *positive definite* property) Prove that $d_H(P,Q)=0$ if and only if *P* and *Q* are the same point.
(b) (the *symmetric* property) Prove that For any *H-points P* and *Q*, $d(P,Q)=d(Q,P)$.

[7] You are familiar with the unit circle for the Euclidean distance function, d_2, and in Section 5.2, you saw the unit circle for the Taxi-cab distance function, d_1.

(a) Sketch the unit circle for the discrete distance function, *disc*.
(b) Sketch the unit circle for the worst-case distance function, d_∞.

[8] Prove that the translation, $T_{\vec{v}}$, is an isometry for the worst-case distance function, d_∞. (Hint: Use a proof structure similar to the ones used for proofs of similar claims in Section 5.3.)

[9] Consider a line *l* that goes through the origin, making an angle φ with the positive *x*-axis. For which angles φ will the reflection M_l be an isometry for the taxi-cab distance function, d_1? (Hint: use a proof structure similar to the one used for the proof of Theorem 10 in Section 5.3.)

6. Coordinate Functions

6.1. The definition of a coordinate function

In this chapter, we continue on our program of learning background mathematical concepts that will be helpful for understanding the so-called *distance postulate* and *ruler postulate*. The concept that we will discuss is that of a *coordinate function* for a line. We start with a definition.

Definition 25 Coordinate function for a line

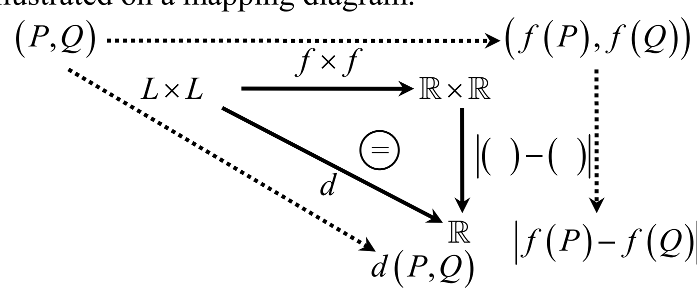
- words: f is a coordinate function for line *L*.
- usage: When the words above are written, it is implied that some distance function *d* is in use.
- meaning: f is a bijection, $f : L \to \mathbb{R}$, such that for each $A, B \in L$, $d(A,B) = |f(A) - f(B)|$.
- additional terminology: The output number $f(A)$ is called *the coordinate of A*.

Consider the equation $d(A,B) = |f(A) - f(B)|$ that appears in the definition of a coordinate function. One can think of this equation as a comparison between two different processes for getting a real number from a pair of points on line *L*.

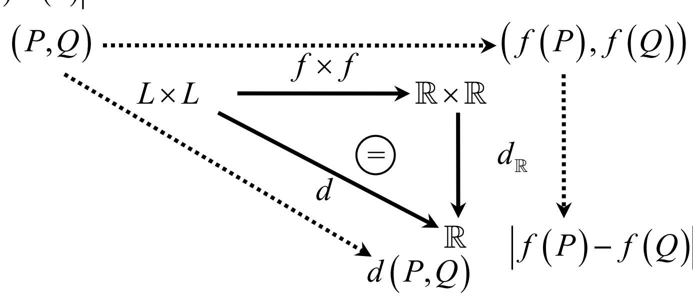
Process 1: Given a pair of points *P* and *Q*, compute the distance between them.

Process 2: Given a pair of points *P* and *Q*, feed both points into the function f to obtain outputs $f(P)$ and $f(Q)$. Note that these outputs are real numbers, not points. Compute the absolute value of the difference between the real numbers $f(P)$ and $f(Q)$.

To say that f is a coordinate function means that the outputs of these two processes are always the same. This idea can be illustrated on a mapping diagram.



Consider the right vertical leg on the diagram above. It depicts the operation of taking the absolute value of the difference between two real numbers. But in Definition 18 (in Section 5.1.1), we saw that operation defined as the real number distance function, $d_{\mathbb{R}}$. Let's use this symbol in the diagram above, instead of the symbol $|(\ )-(\ )|$.



This diagram has much in common with the diagram for an isometry found in Section 4.4. In fact, some of the same terminology that we used when describing isometries (see Definition 12 in Section 2.2) can be used when describing coordinate functions. Remember that for an isometry, the distance between the two *input points* will always be the same as the distance between the two *output points*. Because of that behavior just described, we say that *isometries preserve distance*. Now notice that for a coordinate function, the distance between two *input points* will always be the same as the distance between the two *real number outputs*. For that reason, we could also say that *coordinate functions preserve distance*.

6.2. Examples of Coordinate Functions

Example #1: a coordinate function for a line in Euclidean geometry

Claim: If L is the line $\{(x,y): y = 4x+3\}$ in $\mathbb{R}^2$ with the Euclidean distance function, then the function $f: L \to \mathbb{R}$ defined by $f(x,y) = x\sqrt{17}$ is a coordinate function for *L*.

Proof (We must show that f does meet all the requirements of a coordinate function.)

Step 1: Confirm that f is one-to-one.

Observe that if the domain of f were all of $\mathbb{R}^2$, then f would not be a one-to-one function. For example, let $P=(3,0)$ and $Q=(3,1)$. Then $P \neq Q$ and yet $f(P) = f(3,0) = 3\sqrt{17}$ and $f(Q) = f(3,1) = 3\sqrt{17}$. But neither point *P* nor *Q* is on line *L*, so the fact that they both would produce the same output is irrelevant because they will never be used as inputs.

So let's start over and consider two generic points *P* and *Q* that *are* on line *L*. Points *P* and *Q* must be of the form $P=(x,4x+3)$ and $Q=(w,4w+3)$. Suppose that $f(P)=f(Q)$. Then $x\sqrt{17} = w\sqrt{17}$. Therefore, $x = w$. But that means that the points *P* and *Q* are in fact the same point. This proves that f is one-to-one.

Step 2: Confirm that f is *ONTO*.

Let $b \in \mathbb{R}$ be some desired output. Then let $x = \frac{b}{\sqrt{17}}$ and $y = \frac{4b}{\sqrt{17}}+3$. Then the point $P=(x,y)=\left(\frac{b}{\sqrt{17}},\frac{4b}{\sqrt{17}}+3\right)$ lies on line *L* and $f(P)=f\left(\frac{b}{\sqrt{17}},\frac{4b}{\sqrt{17}}+3\right)=\frac{b}{\sqrt{17}}\cdot\sqrt{17}=b$.

This confirms that f is *ONTO*.

Step 3: Confirm that f preserves distance.

Let $P=(x,4x+3)$ and $Q=(w,4w+3)$ be any two points on line *L*. We compare the results of the two processes shown in the mapping diagram.

$$d_2(P,Q)=d_2((x,4x+3),(w,4w+3))=\sqrt{(x-w)^2+((4x+3)-(4w+3))^2}$$
$$=\sqrt{(x-w)^2+(4(x-w))^2}=\sqrt{(x-w)^2+4^2(x-w)^2}=\sqrt{(x-w)^2+16(x-w)^2}$$
$$=\sqrt{17(x-w)^2}=\sqrt{17}\cdot\sqrt{(x-w)^2}=\sqrt{17}|x-w|$$
$$|f(P)-f(Q)|=|f(x,4x+3)-f(w,4w+3)|=|x\sqrt{17}-w\sqrt{17}|=|\sqrt{17}(x-w)|=\sqrt{17}|x-w|$$

The fact that these two results are the same confirms that f preserves distance.

End of Proof

Example #2: a coordinate function for a line with the Taxi-cab distance function d_1

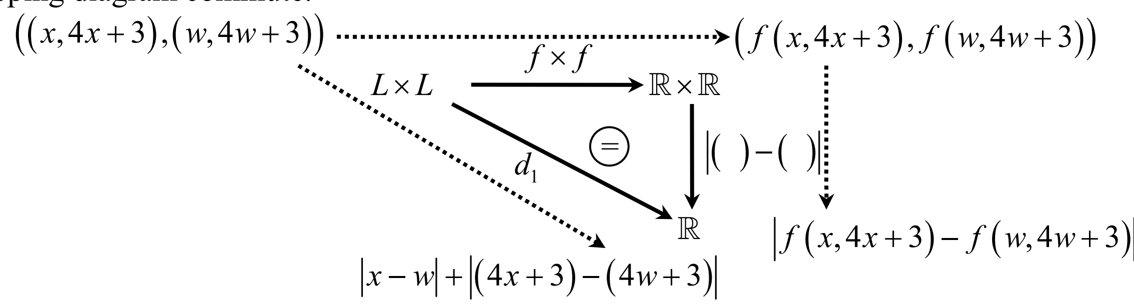
Let L be the line $\{(x,y): y=4x+3\}$ in $\mathbb{R}^2$ with the Taxi-cab distance function d_1. Find a coordinate function f for *L*.

Solution:

Let $P=(x,4x+3)$ and $Q=(w,4w+3)$ be any two points on line *L*. Our goal is to find a function $f:L\to\mathbb{R}$ that is a bijection and that makes the following equation true:

$$\left|f(x,4x+3)-f(w,4w+3)\right|=|x-w|+\left|(4x+3)-(4w+3)\right|$$

This can be visualized: We want to find a function $f:L\to\mathbb{R}$ that will make the following mapping diagram commute.



We can rewrite the expression $|x-w|+\left|(4x+3)-(4w+3)\right|$ in the following way:

$$\begin{aligned}|x-w|+\left|(4x+3)-(4w+3)\right| &= |x-w|+\left|4(x-w)\right| \\ &= |x-w|+4\left|(x-w)\right| \\ &= 5\left|(x-w)\right| \\ &= |5x-5w|\end{aligned}$$

So the initial goal of finding a function f such that

$$\left|f(x,4x+3)-f(w,4w+3)\right|=|x-w|+\left|(4x+3)-(4w+3)\right|$$

can be changed to the easier goal of finding a function f such that

$$\left|f(x,4x+3)-f(w,4w+3)\right|=|5x-5w|$$

We see that a function $f:L\to\mathbb{R}$ defined by $f(x,y)=5x$ would work. It should be confirmed that the function f is a bijection. The proof is not hard: it is almost identical to the proof that the function f from Example #1 is a bijection.

Example #3: a coordinate function for a line with the worst-case distance function d_∞

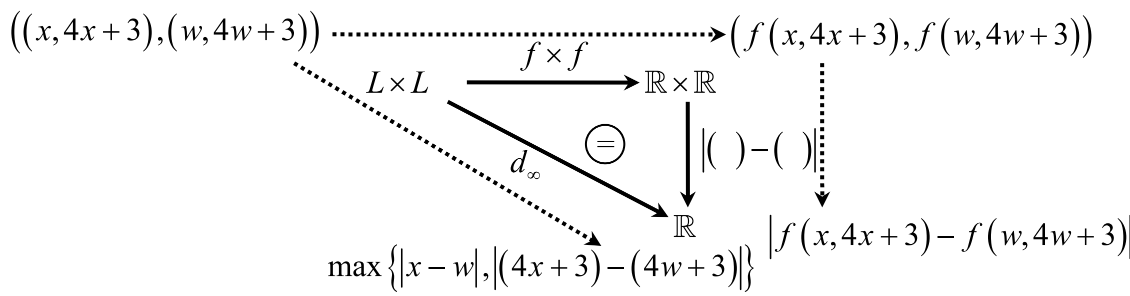
Let L be the line $\{(x,y):y=4x+3\}$ in $\mathbb{R}^2$ with the worst-case distance function d_∞. Find a coordinate function f for *L*.

Solution:

Let $P=(x,4x+3)$ and $Q=(w,4w+3)$ be any two points on line *L*. Our goal is to find a function $f:L\to\mathbb{R}$ that is a bijection and that makes the following equation true:

$$\left|f(x,4x+3)-f(w,4w+3)\right|=\max\left\{|x-w|,\left|(4x+3)-(4w+3)\right|\right\}$$

This can be visualized: We want to find a function $f:L\to\mathbb{R}$ that will make the following mapping diagram commute.



We can rewrite the expression $\max\{|x-w|,|(4x+3)-(4w+3)|\}$ in the following way:

$$\begin{aligned}\max\{|x-w|,|(4x+3)-(4w+3)|\} &= \max\{|x-w|,|4(x-w)|\}\\ &= \max\{|x-w|,4|x-w|\}\\ &= 4|(x-w)|\\ &= |4x-4w|\end{aligned}$$

So the initial goal of finding a function f such that

$$|f(x,4x+3)-f(w,4w+3)| = \max\{|x-w|,|(4x+3)-(4w+3)|\}$$

can be changed to the easier goal of finding a function f such that

$$|f(x,4x+3)-f(w,4w+3)| = |4x-4w|$$

We see that a function $f: L \to \mathbb{R}$ defined by $f(x,y)=4x$ would work. Again, it should be confirmed that the function f is a bijection. But again, the proof is almost identical to the proof that the function f from Example #1 is a bijection, so we skip the proof.

Example #4: There can be more than one coordinate function for a given line.

In Example #3, inspired by the requirement that the function $f: l \to \mathbb{R}$ had to satisfy

$$|f(x,4x+3)-f(w,4w+3)| = |4x-4w|,$$

we defined a function $f: L \to \mathbb{R}$ by $f(x,y)=4x$. But that is not the only way that we could have defined f. Suppose that we had defined f by $f(x,y)=-4x+11$, instead. Right away, we observe that this new function f is certainly a bijection. And computing $|f(x,4x+3)-f(w,4w+3)|$, we find

$$|f(x,4x+3)-f(w,4w+3)| = |(-4x+11)-(-4w+11)| = |-(4x-4w)| = |4x-4w|.$$

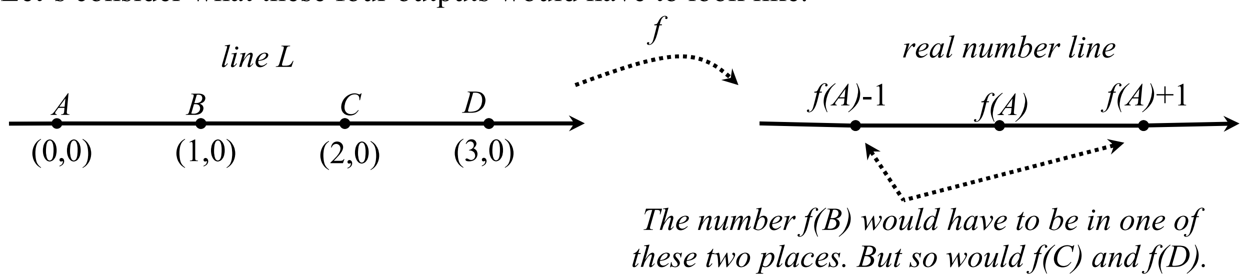
In other words, this new function f satisfies both of the requirements of a coordinate function for line *L*. So f is a coordinate function for line *L* with the worst-case distance function

Example #5: With some distance functions, there cannot be a coordinate function.

Look back at the picture used for exercise [1] in Section 5.4. Let line *L* be the *x*-axis, and points *A*, *B*, *C*, and *D* be the points as shown. In that exercise, using the discrete distance function, you found that $disc(A,B)=1$, $disc(A,C)=1$, and $disc(A,D)=1$. Suppose that a coordinate function exists for line *L*. Call the coordinate function *f*. When we feed the four points *A*, *B*, *C*, and *D* into the coordinate function *f*, the resulting outputs must be four distinct real numbers (because *f* must be 1-to-1). But because *f* is a coordinate function, we also know that for any two points P, Q, the equation $\left|f(P)-f(Q)\right| = disc(P,Q)$ must be true. Therefore, we can say that the outputs $f(A)$, $f(B)$, $f(C)$, and $f(D)$ must satisfy these three equations

$$\left|f(A)-f(B)\right| = disc(A,B) = 1$$
$$\left|f(A)-f(C)\right| = disc(A,C) = 1$$
$$\left|f(A)-f(D)\right| = disc(A,D) = 1.$$

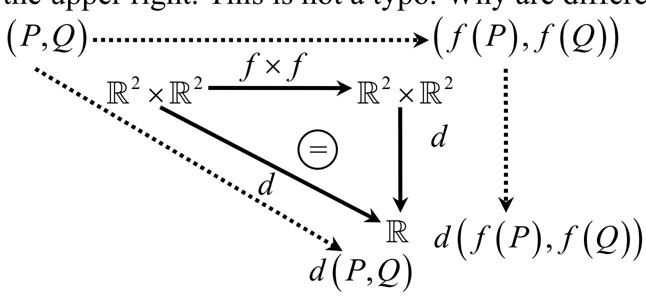
Let's consider what these four outputs would have to look like.

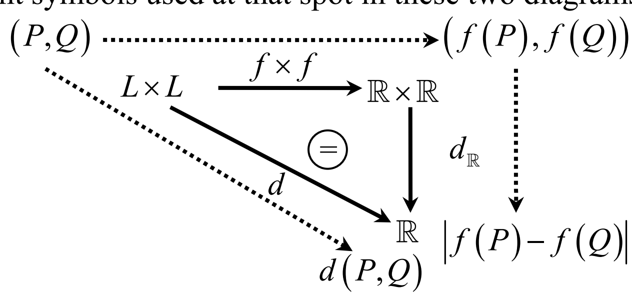


We see that the numbers $f(A)$, $f(B)$, $f(C)$, and $f(D)$ cannot all be different. In other words, the function f could not be 1-to-1. Therefore, we were wrong to assume that there could be a coordinate function for line L with the discrete distance function: there cannot be one.

6.3. Exercises

[1] In Section 4.4, you saw a mapping diagram that illustrates what it means for a function to be an isometry of $\mathbb{R}^2$. In Section 6.1, you saw a mapping diagram that illustrates what it means for a function to be a coordinate function for a line in $\mathbb{R}^2$. Those diagrams are shown below. In one of these diagrams, the symbol $\mathbb{R}^2 \times \mathbb{R}^2$ appears in the upper right, while in the other diagram, the symbol $\mathbb{R} \times \mathbb{R}$ appears in the upper right. This is not a typo. Why are different symbols used at that spot in these two diagrams?





[2] Let L be the line $\left\{(x,y) : y = -\frac{1}{2}x + 5\right\}$ in $\mathbb{R}^2$.

(a) Suppose the distance function is d_2. Find a coordinate function for L.

(b) Suppose the distance function is d_1. Find a coordinate function for L.

(c) Suppose the distance function is d_∞. Find a coordinate function for L.

[3] Let L be a line in hyperbolic geometry (an *"H-line"*). The goal is to produce a coordinate function f for L. That is, we want a bijective function $f : L \to \mathbb{R}$ such that for all *H-points* P and Q on L,
$d_H(P,Q) = \left|f(P)-f(Q)\right|$.

(a) Let R and S denote the "missing endpoints" of the line L. Using the definition of hyperbolic distance d_H ,write down the expression for the distance $d_H(P,Q)$ between two *H-points* P and Q on L.

(b) Observe that the expression for $d_H(P,Q)$ involves the logarithm of a quotient. Use a rule of logarithms to rewrite this as the difference of two logarithms.

(c) The expression for the distance should now look like $d_H(P,Q) = |\mathit{something} - \mathit{something\ else}|$. Compare the expression on the right side, $|\mathit{something} - \mathit{something\ else}|$, to the expression that we want on the right side, $|f(P) - f(Q)|$. Determine a function f that will work.

(d) Is the function $f : L \to \mathbb{R}$ a 1-to-1 function? Explain why you think so.

(e) Is the function $f : L \to \mathbb{R}$ an ONTO function? Explain why you think so.

[4] Prove the following:

Theorem 11 If $f : L \to \mathbb{R}$ is a coordinate function and $g : L \to \mathbb{R}$ is the function defined by $g(x) = f(x) + C$, where $C \in \mathbb{R}$ is a constant, then g is also a coordinate function.

[5] Prove the following:

Theorem 12 If $f : L \to \mathbb{R}$ is a coordinate function and $g : L \to \mathbb{R}$ is the function defined by $g(x) = -f(x)$, then g is also a coordinate function.

[6] Prove the following:

Theorem 13 If a line L has a coordinate function $f : L \to \mathbb{R}$, and P and Q are distinct points on L, then there exists a coordinate function $g : L \to \mathbb{R}$ for L such that $g(P) = 0$ and $g(Q)$ is positive.

[7] Consider again the line $L = \left\{(x,y) : y = -\frac{1}{2}x + 5\right\}$ in $\mathbb{R}^2$ from exercise [2] above. Observe that $P = (6,2)$ and $Q = (4,3)$ are points on line L.

(a) Suppose the distance function is d_2. Find a coordinate function g for L such that $g(P) = 0$ and $g(Q)$ is positive.

(b) Suppose the distance function is d_1. Find a coordinate function g for L such that $g(P) = 0$ and $g(Q)$ is positive.

(c) Suppose the distance function is d_∞. Find a coordinate function g for L such that $g(P) = 0$ and $g(Q)$ is positive.

[8] In this exercise, we revisit the hyperbolic line and points used in exercise [3] of Section 5.4. Let L be the *H-line* consisting of the portion of the x-axis that lies inside the unit circle. The "missing endpoints" of this line are $R = (-1,0)$ and $S = (1,0)$. (They are not *Hpoints*.)

(a) Find a coordinate function f for the line L.

(b) Consider the *Hpoints* $A = (0,0)$, $B = \left(\frac{1}{4},0\right)$, $C = \left(\frac{1}{2},0\right)$, and $D = \left(\frac{3}{4},0\right)$. Using your coordinate function, compute $f(A)$, $f(B)$, $f(C)$ and $f(D)$.

(c) For the line L, find a new coordinate function g such that $g(C) = 0$ and $g(B)$ is positive.

7. Distance Functions & Coordinate Functions in Geometry Postulates

Having learned much terminology pertaining to transformations of the plane, distance functions, and coordinate functions, we are now well-equipped to study the "distance" and "ruler" postulates that are frequently included in lists of Euclidean geometry postulates.

7.1. Discussion of the SMSG postulates for Euclidean geometry

In the 1960's, the *School Mathematics Study Group* (*SMSG*) developed a set of postulates to be used for teaching Euclidean geometry in high schools. Included in their list of postulates are the following three.

SMSG Postulate 2 *The Distance Postulate* To every pair of distinct points there corresponds a unique positive number. This number is called the *distance* between the two points.

SMSG Postulate 3 *The Ruler Postulate* The points of a line can be placed in a correspondence with the real numbers such that

1. To every point of the line there corresponds exactly one real number.
2. To every real number there corresponds exactly one point of the line
3. The distance between two distinct points is the absolute value of the difference of the corresponding real numbers.

SMSG Postulate 4 *The Ruler Placement Postulate* Given two points P and Q of a line, the coordinate system can be chosen in such a way that the coordinate of P is zero and the coordinate of Q is positive.

We see that in essence, Postulate 2 is saying that there exists a distance function, which we could call *d*. And Postulate 3 is saying that for any line L, there exists a coordinate function, which we could call f. There are many other things in these three postulates that we can observe. We will discuss a few.

Observe that the word "function" does not appear anywhere in the statements of the three postulates. This is because the concept of a "function" does not play a large role in high school geometry. In fact, one would probably not find anywhere in a high school math book the term "function" defined as we are used to defining it (as a sort-of "mathematical machine", with input and output). An exception might be the calculus books used in high schools, especially since college-level books are often used for high school advanced placement calculus courses. But even if the abstract definition of "function" does appear in a high-school book, it is probably overlooked. It is usually at the college level, hopefully in calculus courses but often not until later courses, that the concept is introduced and used.

Observe that Postulate 2 and Postulate 3.3 only refer to *distinct* points. There seems to be no provision for measuring $d(P,P)$. And because of that, there is no mention of any requirement that $d(P,P)=0$.

Observe that Postulate 2 does not mention a requirement that $d(P,Q)=d(Q,P)$, the symmetry property of a distance function. This might seem to be an oversight, but in fact the combination of Postulate 2 and Postulate 3 does guarantee that distance will be symmetric. To see why, let P and Q be any two distinct points. Then there exists a line L containing P and Q. By Postulate 3, there exists a coordinate function for line L, a function that we could call f. (The Postulate does not use this terminology, but we can.) Consider the value of $d(Q,P)$.

$$d(Q,P) = |f(Q) - f(P)| \quad \text{by Postulate 3.3}$$
$$= |f(Q) - f(P)| \quad \text{property of absolute value}$$
$$= d(P,Q) \quad \text{by Postulate 3.3}$$

What we have done, essentially, is to prove a theorem saying that distance is symmetric. So we see that even though the postulates do not explicitly say that distance will be symmetric, they do not need to say so: the fact can be proven as a theorem.

Observe that Postulate 2 does not mention any requirement that for any $P,Q,R \in \mathbb{R}^2$, $d(Q,R) \leq d(P,Q) + d(Q,R)$ (the triangle inequality). But as with the missing statement of the symmetry property, it turns out that other postulates can be used in a theorem that proves that the triangle inequality is always true.

We have just seen two instances (the symmetry property and the triangle inequality) where the SMSG postulates omit a statement that does not absolutely have to be on the list of postulates because it can be proven as a theorem. But there are other instances where the SMSG postulates include statements that could be omitted. Consider Postulate 4. It could be proven as a theorem, using Postulates 2 and 3 as given information. (You already did this when you proved Theorem 13 in exercise [6] of Section 6.3.) So Postulate 4 could be omitted from the list of postulates.

Observe that Postulate 2 does not say exactly what the distance function is. Since we saw four different examples of distance functions for $\mathbb{R}^2$ in Section 5.1, you might reasonably assume that any of these four distance functions could be used, and maybe others. But in fact that is not the case. Remember that in Example #5 in Section 6.2, you saw that if the discrete distance function, *disc*, is used, then it is impossible for a line to have a coordinate function. That would mean that Postulate 3 could not be true. In the exercises, you will re-visit an old homework exercise that will reveal that distance functions d_1 and d_∞ can also be ruled out. The fact is that the only distance function that *will* work is one that is either equal to d_2 or that is a constant multiple of d_2.

7.2. Exercises

[1] In exercise [1] from Section 5.4, you computed $d(A,B)$, $d(B,E)$, and $d(A,E)$ using distance functions d_1 and d_∞. The results contradict one of the theorems of Euclidean geometry. Which theorem? The conclusion that can be drawn is that neither distance function, d_1 nor d_∞, could be used as the distance function for Euclidean geometry.

[2] In Section 7.1, you saw an example of three typical postulates that pertain to distance functions and coordinate functions. In different books, the postulates may take different forms.

a) Find a high school geometry book and locate in it the postulates pertaining to distance functions and coordinate functions.
b) List the postulates from here, along with the book's title, author, publisher, and publication date.
c) Discuss how your books postulates differ from the *SMSG* Postulates 2, 3, and 4.
d) Are the differences merely stylistic, or is there substantial difference in content?
e) If there are substantial differences, why do you think the book's author chose to present the postulates in the way that they did?