# MTH 210: Intro to Linear Algebra — Fall 2019
# Course Notes — Drew Armstrong

Linear algebra is the common denominator of modern mathematics. From the most pure to the most applied, if you use mathematics then you **will** use linear algebra. It is also a relatively new subject. Linear algebra as we know it was first systematized in the 1920s.[1] Since then it has slowly made its way into the school curriculum and is still increasing in importance every year. In the future I expect that introductory linear algebra will be taught as a two-semester sequence, similar to the way that we treat calculus. Until then we have to make do with just one semester. I will do my best to show you the most important ideas.

## Contents



[1] The ingredients of linear algebra were developed in the 1800s but they didn't come together as a coherent subject until the development of quantum physics.

# 1 Vector Arithmetic

Before discussing linear algebra proper, we must first develop the correct language, which is based on *vectors*. These, in turn, are based on the concepts of *coordinate geometry*, so that's where we'll begin.

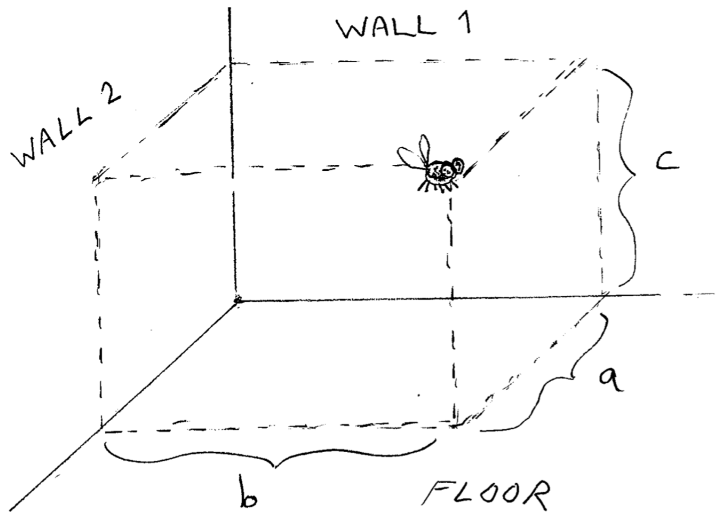
## 1.1 Cartesian Coordinates

The subject of geometry is fundamentally about points in space. But what is "space," and what is a "point"? Our modern understanding is based on the revolutionary ideas of René Descartes and Pierre de Fermat from the early 1600s.

**The Idea of Coordinate Geometry**

A *point* is an ordered list of numbers.

Apparently Descartes was lying in bed one morning when he saw a fly buzzing in the corner:




He realized that the position of the fly could be uniquely specified by measuring the (perpendicular) distance to the two walls and the floor. We can visualize this as a rectangular box with dimensions $a, b, c$. If we arrange these numbers in some fixed order (say $a$ is the distance to wall 1, $b$ is the distance to wall 2 and $c$ is the distance to the floor), then we can say that

$$(a, b, c) = \text{"the (Des)Cartesian coordinates of the fly."}$$

In this course we prefer to express Cartesian coordinates as a vertical column:[2]

$$\mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Such a column $\mathbf{v}$ is called a *vector*, and we say that $a, b, c$ are the *coordinates* of the vector. By allowing negative coordinates we can specify the position of any point in space. The point with all coordinates equal to zero is called the *origin* of the coordinate system:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

///

But this is more than just a notation because it suggests new things that we can **do** with points. For example, we can "add" them.

[2]Remark: In the notes I will use a boldface font for vectors. On the whiteboard I will use the notation $\mathbf{v} = \vec{v}$, since I am not good at drawing boldface letters.

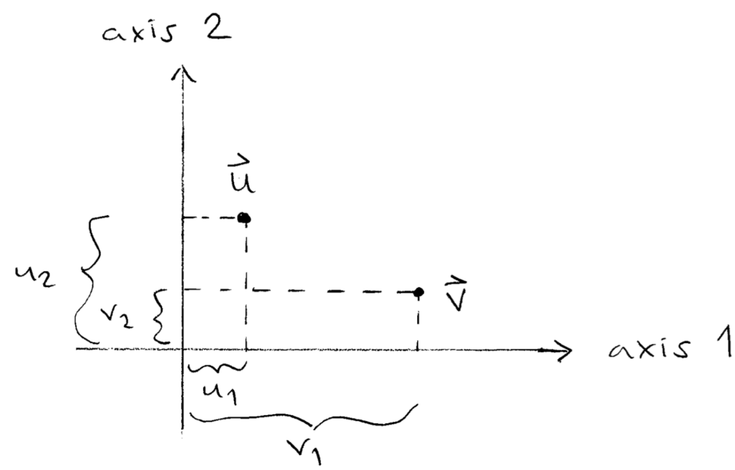
**Addition of Points in the Plane**

Given vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ we define them *sum* as follows:[3]

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} := \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix}.$$

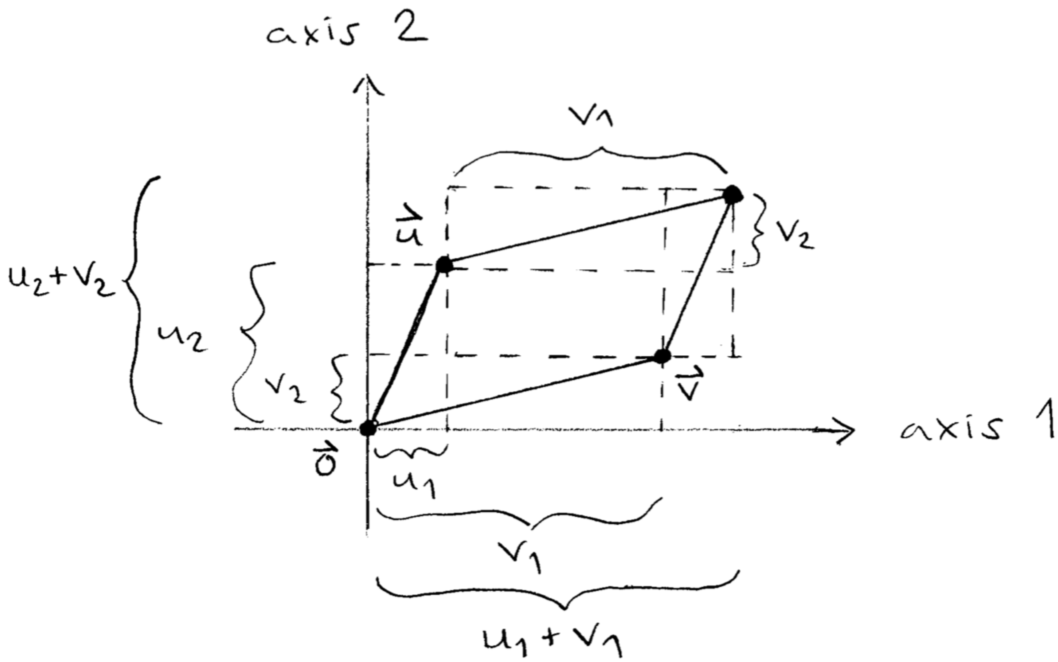
That is, we add two vectors by adding their respective components.

This definition is completely natural in terms of numbers, but what does it mean in terms of geometry? First let us draw the two points in the Cartesian plane (note that I have labeled the two axes arbitrarily):




If we draw the point $\mathbf{u} + \mathbf{v}$ then we observe that the four points $\mathbf{0}, \mathbf{u}, \mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ lie at the four vertices of a parallelogram:

[3]Since column vectors take up lots of vertical space I will sometimes write them as a horizontal list separated by commas.




We will formalize this observation by calling it the "parallelogram law."

**The Parallellogram Law**

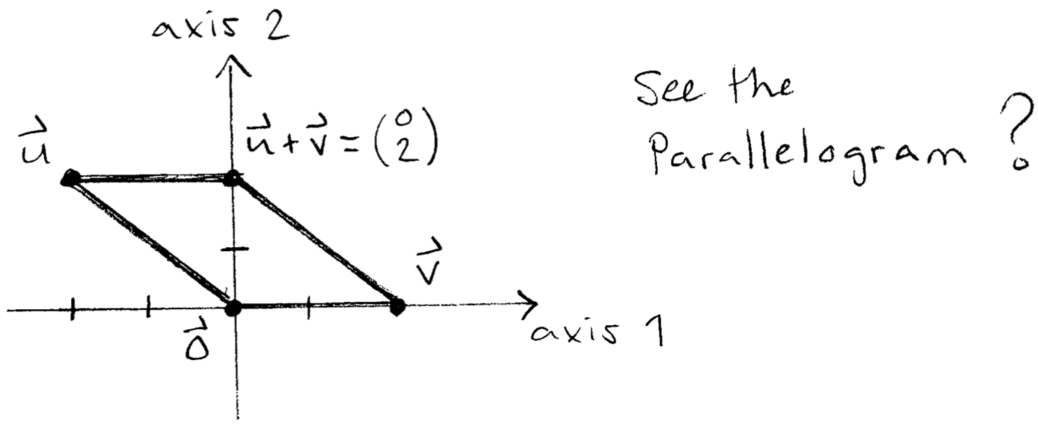
Let $\mathbf{u}$ and $\mathbf{v}$ be any two vectors.[4] Then the four points $\mathbf{0}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$ are the vertices of a parallelogram.

**2D Example:** Let $\mathbf{u} = (-2, 2)$ and $\mathbf{v} = (2, 0)$. We compute the sum as follows:

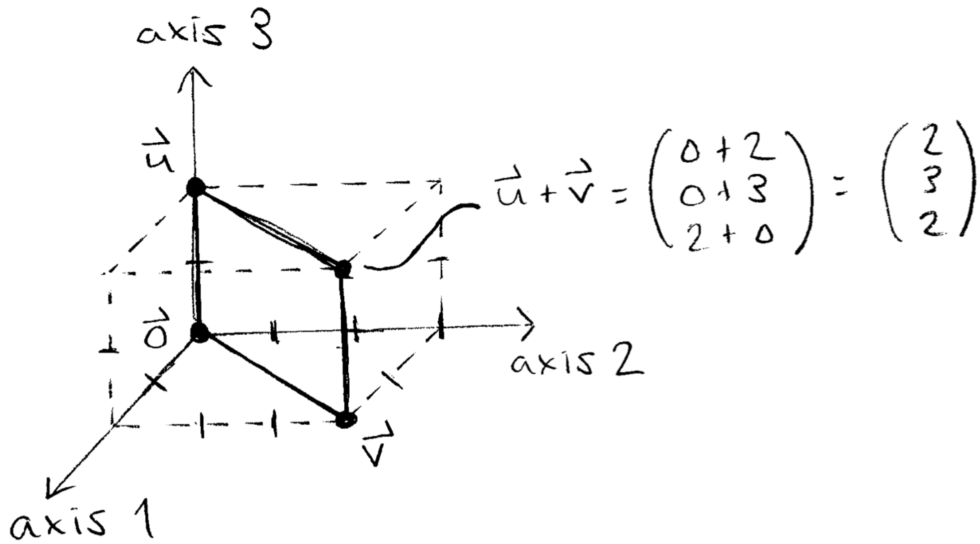
$$\mathbf{u}+\mathbf{v} = \begin{pmatrix}-2\\2\end{pmatrix} + \begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}-2+2\\2+0\end{pmatrix} = \begin{pmatrix}0\\2\end{pmatrix}.$$

And here is the picture:

[4]These vectors could have two or three coordinates. Later we will allow them to have $n$ coordinates and pretend the the parallelogram lives in "$n$-dimensional space."




**3D Example:** The parallelogram law also works in 3D. Let $\mathbf{u} = (0, 0, 2)$ and $\mathbf{v} = (2, 3, 0)$, so that $\mathbf{u} + \mathbf{v} = (2, 3, 2)$. Here is the picture:

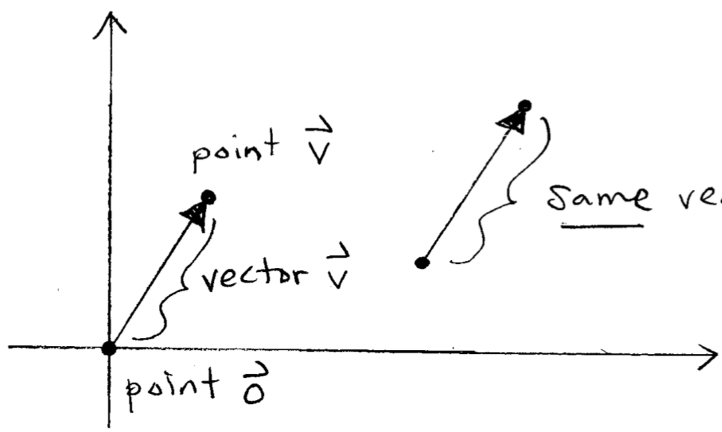



Do you see the 2D parallelogram living in 3D space? (Actually, it is a rectangle.)

We will assume that the parallelogram law holds for vectors with any number $n$ of coordinates, even though we can't draw a picture of $n$-dimensional space when $n \geqslant 4$.

## 1.2 Points vs. Directed Line Segments

The parallelogram law suggests that there is more to a vector than just the point that it represents. Sometimes we choose to view the vector $\mathbf{v} = (v_1, v_2)$ as a *directed line segment*. The subtle thing about this point of view is that **we are allowed to pick up the directed line segment and move it**, as long as we don't change its length or direction:



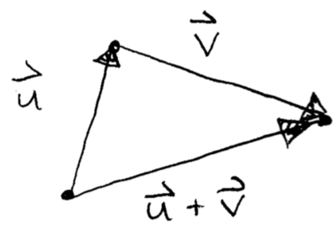

It will take a while for you to internalize this idea, so I will repeat myself frequently. The reason that we want to move vectors is because it allows us to give a purely geometric definition of vector addition.

**Addition of Directed Line Segments**

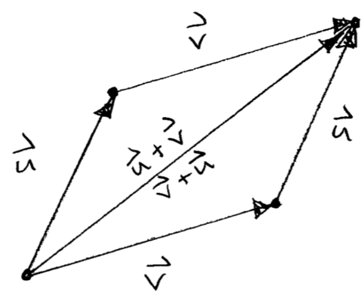
Let $\mathbf{u}$ and $\mathbf{v}$ be directed line segments. To compute $\mathbf{u} + \mathbf{v}$ we first move $\mathbf{v}$ so that its tail coincides with the head of $\mathbf{u}$. Then $\mathbf{u} + \mathbf{v}$ is defined as the directed line segment from the tail of $\mathbf{u}$ to the head of $\mathbf{v}$. In other words:

*directed line segments add head-to-tail.*

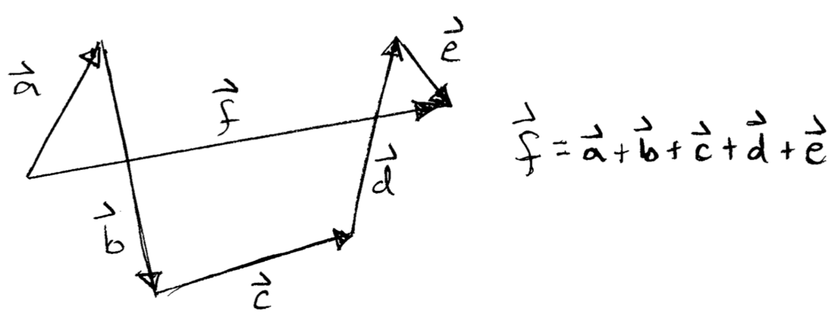
Here is a picture:




It turns out that the geometric definition of vector addition (head-to-tail) agrees with the algebraic definition (componentwise addition) because of the parallelogram law:




This picture also demonstrates that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. In other words, the addition of directed line segments is *commutative*. The advantage of this idea is that we can add many vectors together without worrying about the coordinate system. For example:




Because of commutativity, adding the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}$ in any order (there are $5! = 120$ different ways to do this) always gives the same vector $\mathbf{f}$. [Remark: One must also check that addition of vectors is *associative*: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$. You will do this on the homework.] This picture would be a mess if we had to keep track of all the coordinates.

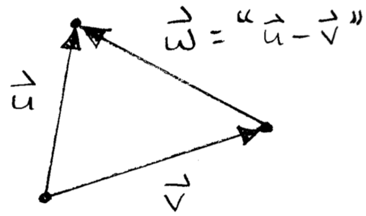
In fact, we can draw the same kind of picture for vectors in any number of dimensions. I didn't mention it at the time, but the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}$ above live in four-dimensional space.

## 1.3 Subtraction of Vectors

What about subtraction of vectors? Let **u** and **v** be vectors with the same number of coordinates. Which vector **w** deserves to be called "**u** − **v**"? Whatever this vector is, it should definitely satisfy the equation

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

Therefore we have the following picture:



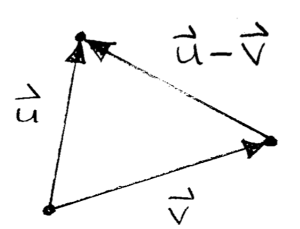

Here is the official definition.

**Subtraction of Vectors**

Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ be vectors of the same dimension. In terms of algebra we define

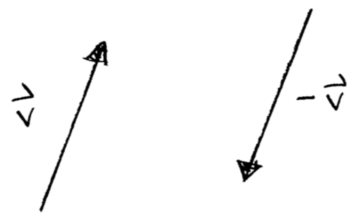
$$\mathbf{u} - \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} - \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}.$$

In terms of geometry, we place the directed line segments **u** and **v** "tail-to-tail." Then **u** − **v** is the directed line segment from the head of **v** to the head of **u**:




These two definitions agree because of the parallelogram law.

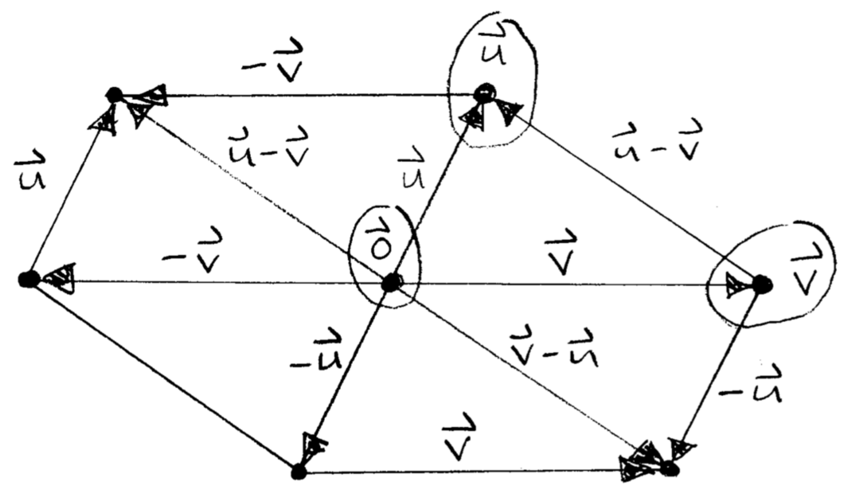
But there is one more way to think about subtraction. For any directed line segment $\mathbf{v}$ we let "$-\mathbf{v}$" denote the directed line segment with the same length, but **opposite direction**:




Then we can think of "$\mathbf{u}$ minus $\mathbf{v}$" as "$\mathbf{u}$ plus the opposite of $\mathbf{v}$":

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

Of course we are allowed to pick up any of these vectors and move them around. We observe that everything fits together very nicely:

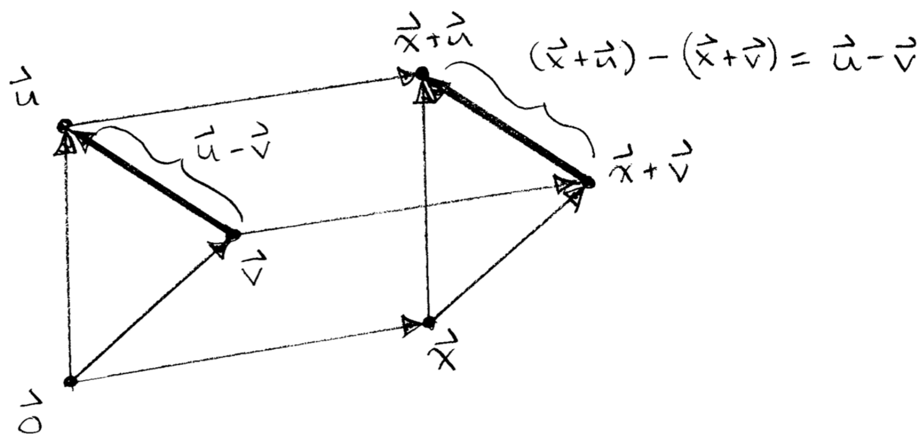



At first it seems confusing that there are so many different ways to think about vectors. But this is actually the reason why linear algebra is so useful. When in doubt, here is a useful mnemonic.

**Mnemonic: Head Minus Tail**

$$\text{(any vector)} = \text{(its head)} - \text{(its tail)}$$

Indeed, the following picture demonstrates that for any three points $\mathbf{u}, \mathbf{v}, \mathbf{x}$, the vector from $\mathbf{v}$ to $\mathbf{u}$ is equal to the vector from $\mathbf{x}+\mathbf{v}$ to $\mathbf{x}+\mathbf{u}$:

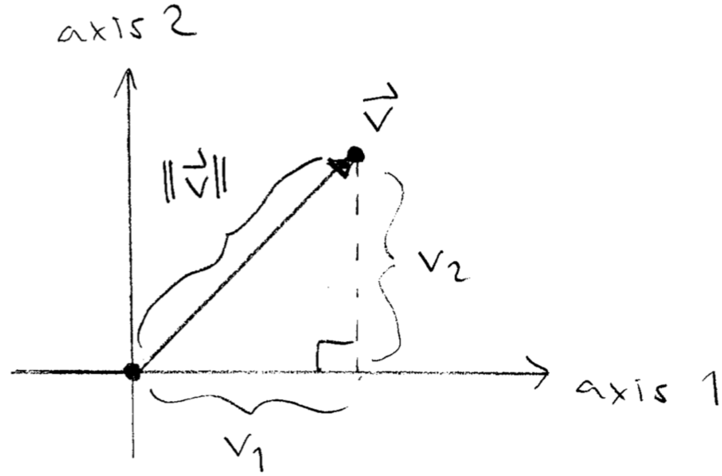



## 1.4 The Pythagorean Theorem

The subject of geometry deals with distances and angles. For any vector $\mathbf{v}$ in the plane we denote its *length* with double[5] absolute value signs:

$$\|\mathbf{v}\| := \text{the length of the directed line segment } \mathbf{v}.$$

In the plane we can easily compute this length using the Pythagorean Theorem. Suppose that $\mathbf{v} = (v_1, v_2)$ in Cartesian coordinates. By definition this means that we have a right triangle:




[5] I guess these are supposed to be "blackboard boldface" absolute value symbols.

Therefore the Pythagorean Theorem tells us that

$$\|\mathbf{v}\|^2 = v_1^2 + v_2^2$$

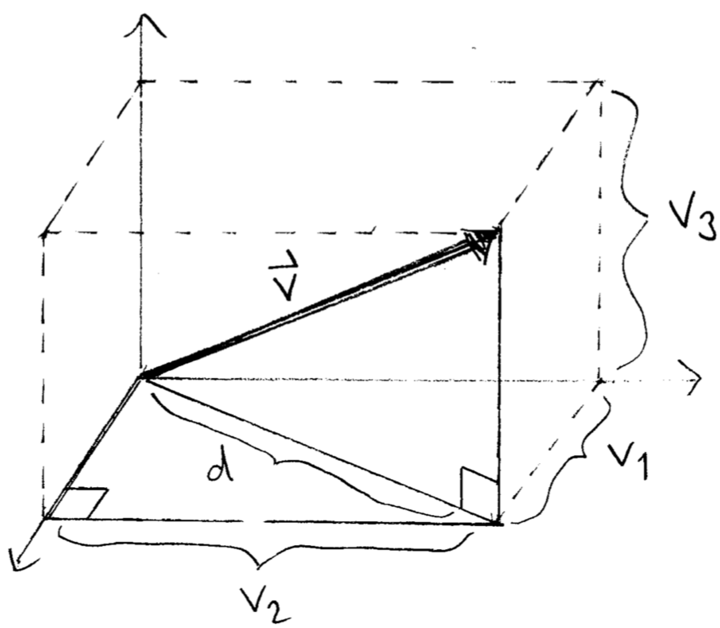
and hence

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}.$$

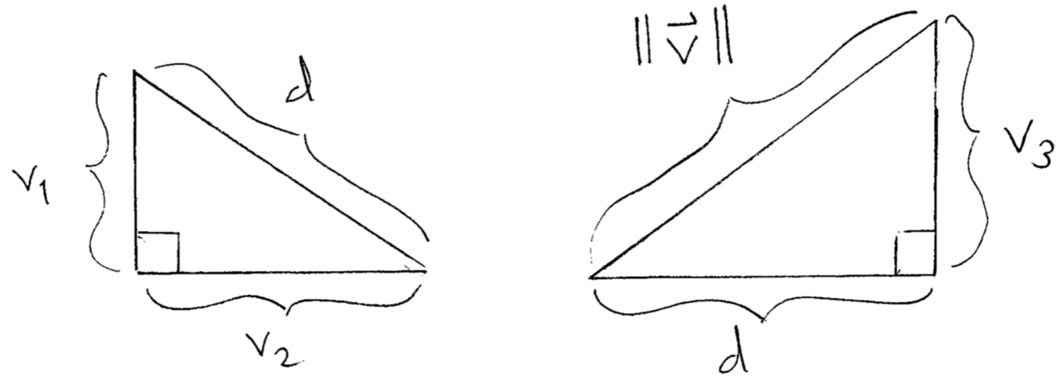
Using vector subtraction, this same formula allows us to compute the *distance* between any two points in the plane. Consider any points $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Then we have

$$\begin{aligned}
(\text{distance between } \mathbf{u} \text{ and } \mathbf{v}) &= (\text{length of the line segment from } \mathbf{u} \text{ to } \mathbf{v}) \\
&= \|\mathbf{v} - \mathbf{u}\| \\
&= \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2}.
\end{aligned}$$

What about in three dimensions? We can think of the vector $\mathbf{v} = (v_1, v_2, v_3)$ as the directed line segment between two opposite corners of a rectangular box:




In order to compute the length $\|\mathbf{v}\|$ we will also draw the line segment from $(0, 0, 0)$ to $(v_1, v_2, 0)$. Then we obtain **two** right triangles:




Applying the Pythagorean Theorem to each triangle separately gives

$$\begin{aligned} d^2 &= v_1^2 + v_2, \\ \|v\|^2 &= d^2 + v_3^2, \end{aligned}$$

and hence

$$\begin{aligned} \|\mathbf{v}\|^2 &= d^2 + v_3 \\ &= (v_1^2 + v_2^2) + v_3^2 \\ &= v_1^2 + v_2^2 + v_3^2. \end{aligned}$$

We can view this as a three dimensional analogue of the Pythagorean Theorem. What about higher dimensions? If $\mathbf{v} = (v_1, v_2, v_3, v_4)$ is a directed line segment in "four dimensional space," is it true that

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2 + v_4^2} \ ?$$

Sure, why not? Whether "four dimensional space" really exists or not, the mathematical concept of distance is easy to work with, so we just go ahead and define it.

**Length and Distance in General**

For any vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ we define its *length* as follows:

$$\|\mathbf{v}\| := \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Then for any points $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ we define the *distance* between them as the length of the line segment $\mathbf{v} - \mathbf{u}$:

$$(\text{distance between } \mathbf{u} \text{ and } \mathbf{v}) := \|\mathbf{v} - \mathbf{u}\| = \sqrt{(v_1 - u_1)^2 + \cdots + (v_n - u_n)^2}.$$

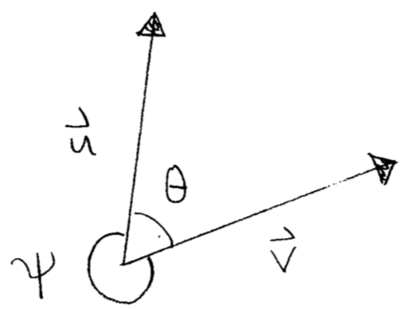
Since the orientation of the line segment doesn't matter, we also have

$$\|\mathbf{v}-\mathbf{u}\| = \|\mathbf{u}-\mathbf{v}\| = \sqrt{(u_1-v_1)^2+\cdots+(u_n-v_n)^2}.$$

We will see later that points in higher dimensional space can be used to represent data sets or sequences of observations. The distance between two data sets represents the amount of difference between them. The technique of *least squares regression* in statistics is based on minimizing the distance between points in higher dimensional space.

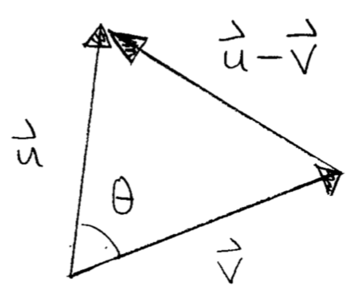
## 1.5 The Dot Product

And what about angles? Let $\mathbf{u}$ and $\mathbf{v}$ be any two vectors in $n$-dimensional space. By placing them tail-to-tail these two vectors determine a 2-dimensional slice of $n$-dimensional space. Here is a picture:




Inside this 2-dimensional slice we can measure the angle between the vectors. Actually, there are two angles $\theta$ and $\psi$ with $\theta+\psi=360°$. It doesn't really matter which angle we choose because the cosines are the same: $\cos\theta=\cos\psi$.

In order to compute the angle $\theta$, let us draw the triangle of vectors with sides $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}-\mathbf{v}$:



The law of cosines gives the following relationship between the side lengths and the angle $\theta$:

$$\boxed{\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta.}$$

On the other hand, we can compute the length $\|\mathbf{u} - \mathbf{v}\|$ in terms of Cartesian coordinates. For example, suppose we are working in 2D space with $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Then applying our formula for the distance between points gives

$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= (u_1 - v_1)^2 + (u_2 - v_2)^2 \\
&= (u_1^2 - 2u_1v_1 + v_1^2) + (u_2^2 - 2u_2v_2 + v_2^2) \\
&= (u_1^2 + u_2^2) + (v_1^2 + v_2^2) - 2(u_1v_1 + u_2v_2).
\end{aligned}$$

We can simplify this formula a bit by substituting $u_1^2 + u_2^2 = \|\mathbf{u}\|^2$ and $v_1^2 + v_2^2 = \|\mathbf{v}\|^2$ to obtain

$$\boxed{\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2(u_1v_1 + u_2v_2).}$$

By comparing the two boxed equations we conclude that

$$\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta = u_1v_1 + u_2v_2.$$

This is a rather strange formula. It relates the geometric concepts of length and angle on the left to a surprising computation on the right involving the Cartesian coordinates. This formula was discovered in the 1840s by the Irish mathematician and physicist William Rowan Hamilton, but his notation was a bit metaphysical. In the 1880s, the scientists Josiah Willard Gibbs and Oliver Heaviside distilled the most useful ideas from Hamilton and they defined the modern language of vectors. Here is their most important definition.

**Definition of the Dot Product**

For any vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ we define the *dot product*:

$$\mathbf{u} \bullet \mathbf{v} := u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Note that this operations takes a pair of vectors $\mathbf{u}$ and $\mathbf{v}$ to a single number $\mathbf{u} \bullet \mathbf{v}$ (called a *scalar*). For this reason the dot product is sometimes also called the *scalar product*.[6] It is worth noting that the length of a vector can be expressed in terms of the dot product:

$$\begin{aligned}
\|\mathbf{u}\| &= \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} &&= \sqrt{\mathbf{u} \bullet \mathbf{u}} \\
\|\mathbf{u}\|^2 &= u_1^2 + u_2^2 + \cdots + u_n^2 &&= \mathbf{u} \bullet \mathbf{u}.
\end{aligned}$$

This definition allows us to state the following theorem.

[6] Sometimes it is also called the *inner product* of vectors. We will see why later when we discuss the *outer product* of vectors.

**Theorem (The Angle Between Two Vectors)**

If $\theta$ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$ (measured in the two-dimensional plane that they share), then the same argument as above can be used to show that[7]

$$\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta.$$

It follows from this that the vectors $\mathbf{u}$ and $\mathbf{v}$ are perndicular (written $\mathbf{u} \perp \mathbf{v}$) if and only if the dot product is zero:[8]

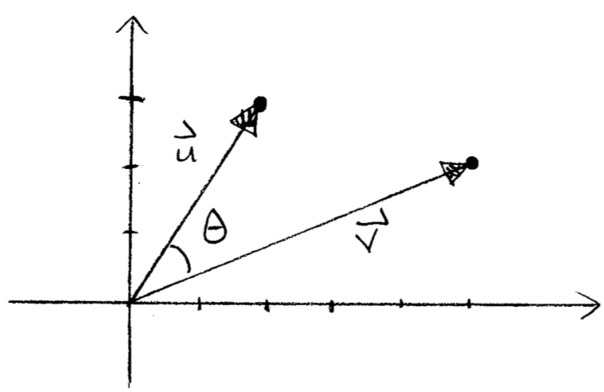
$$\mathbf{u} \perp \mathbf{v} \quad \Longleftrightarrow \quad \cos\theta = 0 \quad \Longleftrightarrow \quad \mathbf{u} \bullet \mathbf{v} = 0.$$

Finally, we can use this formula to solve for (the cosine of) the angle $\theta$ in terms of the Cartesian coordinates of $\mathbf{u}$ and $\mathbf{v}$:

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{u_1v_1 + u_2v_2 + \cdots + u_nv_n}{\sqrt{u_1^2 + \cdots + u_n^2} \cdot \sqrt{v_1^2 + \cdots + v_n^2}}.$$

Let's see some examples.

**2D Example:** Compute the angle $\theta$ between the vectors $\mathbf{u} = (2, 3)$ and $\mathbf{v} = (5, 2)$:




We apply the dot product formula to obtain

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{2 \cdot 5 + 3 \cdot 2}{\sqrt{2^2 + 3^2} \cdot \sqrt{5^2 + 2^2}} = \frac{16}{\sqrt{13} \cdot \sqrt{29}} = 0.824$$

[7]Occasionally you will see a textbook that uses the formula $\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$ as the **definition** of the dot product. I don't like that.

[8]Here we assume that $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ so that $\|\mathbf{u}\|\|\mathbf{v}\| \neq 0$. I guess we could say that the zero vector is perpendicular to every vector.

and hence

$$\theta = \arccos(0.824) = 34.5^\circ \text{ or } 325.5^\circ.$$

Based on our picture we choose the smaller angle. ///

**1D Example:** It might seem a bit silly, but we should make sure that we understand the formula $\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$ in very simple case when $n = 1$.

Let $\mathbf{u} = (u)$ be any "1-dimensional vector," also known as a "real number." Observe that the length of $\mathbf{u}$ is the same as the absolute value of $u$:

$$\begin{aligned} \|\mathbf{u}\|^2 &= u^2 \\ \|\mathbf{u}\| &= \sqrt{u^2} = |u|. \end{aligned}$$

Indeed, this is why we use the absolute value sign for the length of a vector. If $\mathbf{v} = (v)$ is any other number, observe that the dot product of $\mathbf{u}$ and $\mathbf{v}$ is just the product of numbers:

$$\mathbf{u} \bullet \mathbf{v} = uv.$$

If $\theta$ is the angle between $\mathbf{u} = (u)$ and $\mathbf{v} = (v)$ then we conclude that

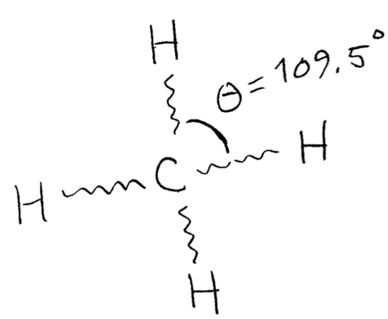
$$uv = |u||v| \cos\theta.$$

Does this make any sense? Sure. The only possible values for $\theta$ are $0^\circ$ (when $u$ and $v$ have the same sign) and $180^\circ$ (when $u$ and $v$ have opposite signs). Since $\cos(0^\circ) = 1$ and $\cos(180^\circ) = -1$ we have the following rule for the product of real numbers:

$$uv = \begin{cases} |u||v| & \text{if } u \text{ and } v \text{ have the same sign,} \\ -|u||v| & \text{if } u \text{ and } v \text{ have opposite signs.} \end{cases}$$

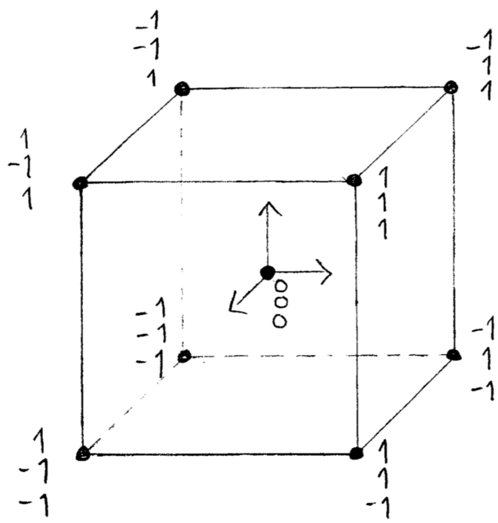
This is correct, and interesting. ///

**3D Example: The Tetrahedral Angle.** Every chemistry textbook quotes $\theta = 109.5^\circ$ as the angle between two hydrogen atoms in a molecule of methane (or any other molecule with tetrahedral symmetry):

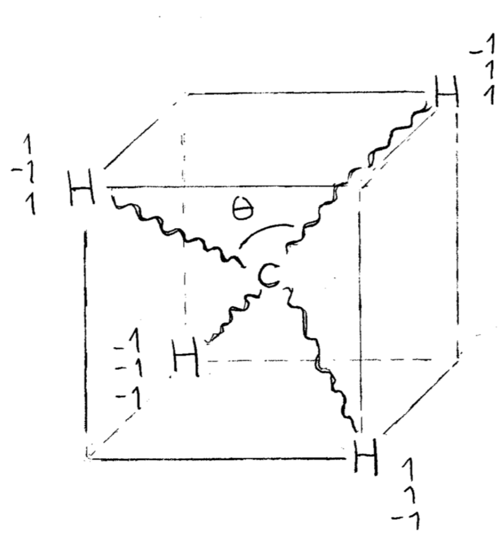


Why is this the correct answer?

In order to minimize energy, the four hydrogen atoms sit at the vertices of a regular tetrahedron centered on the carbon atom. Let's place the carbon atom at the origin $\mathbf{0} = (0, 0, 0)$. Then what are the coordinates of the hydrogen atoms? First let me observe that the eight points $(\pm 1, \pm 1, \pm 1)$ form the vertices of a cube centered at the origin:




By choosing four alternate vertices, we obtain a regular tetrahedron centered at $\mathbf{0}$. There are two choices and it doesn't matter which one we pick. For example, let's place the hydrogen atoms at $(-1, 1, 1)$, $(1, -1, 1)$, $(1, 1, -1)$ and $(-1, -1, -1)$:



Since the angle between any two hydrogen atoms is the same, we can can pick two at random, say $\mathbf{u} = (-1, 1, 1)$ and $\mathbf{v} = (1, -1, 1)$. Then the dot product formula gives

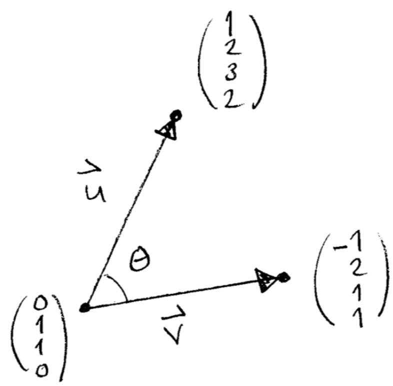
$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{(-1)\cdot 1 + 1 \cdot (-1) + 1 \cdot 1}{\sqrt{(-1)^2 + 1^2 + 1^2} \cdot \sqrt{1^1 + (-1)^2 + 1^2}} = \frac{-1}{\sqrt{3}\cdot\sqrt{3}} = \frac{-1}{3},$$

and it follows that

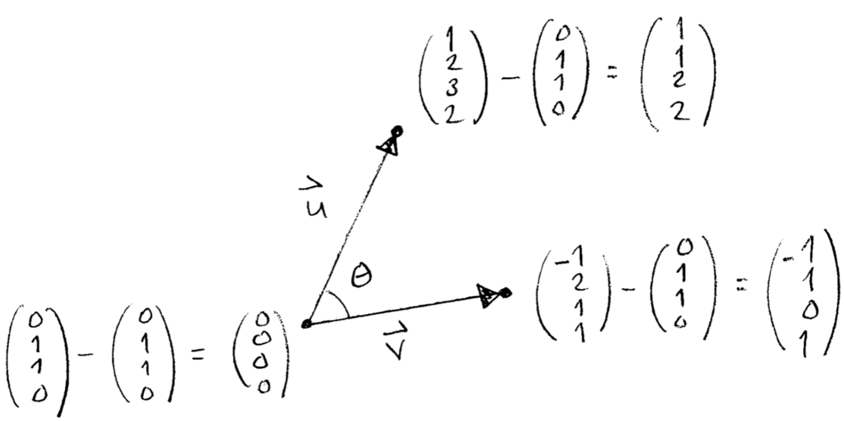
$$\theta = \arccos\left(\frac{-1}{3}\right) = 109.4712206\cdots \text{ degrees.}$$

///

**4D Example:** Compute the angle in the following picture:[9]




In order to compute the dot product $\mathbf{u} \bullet \mathbf{v}$ we must first translate the vectors $\mathbf{u}$ and $\mathbf{v}$ into standard position. For this purpose, we subtract $(0, 1, 1, 0)$ from all three points to obtain




[9]Note that the three points of the triangle live in a 2D slice of 4D space. The angle is defined inside this 2D slice.

Then we compute

$$\mathbf{u} \bullet \mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} \bullet \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = 1 \cdot (-1) + 1 \cdot 1 + 2 \cdot 0 + 2 \cdot 1 = 2,$$

$$\|\mathbf{u}\|^2 = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} = 1^2 + 1^2 + 2^2 + 2^2 = 10,$$

$$\|\mathbf{v}\|^2 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = (-1)^2 + 1^2 + 0^2 + 1^2 = 3.$$

Plugging these into the theorem gives

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{2}{\sqrt{10} \cdot \sqrt{3}} = 0.365$$

and hence

$$\theta = \arccos(0.365) = 68.6^\circ.$$

## 1.6 The Concept of a Vector Space

In the previous sections we developed the basic properties of geometry in terms of Cartersian coordinates. From the modern point of view, we take these formulas as the **definition** of (Euclidean)[10] geometry.

**Definition of Euclidean Space**

Recall that $\mathbb{R}$ denotes the set of real numbers. We define $\mathbb{R}^n$ as the set of ordered $n$-tuples of real numbers, thought of as column vectors:

$$\mathbb{R}^n := \left\{ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}.$$

[10]There exist non-Euclidean geometries, such as the Lorentzian geometry which is used in Einstein's theory of relativity.

For any two column vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ we define their dot product as follows:

$$\mathbf{x} \bullet \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \bullet \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + \cdots + x_n y_n.$$

We define the length $\|\mathbf{x}\| \in \mathbb{R}$ so that the generalized Pythagorean Theorem holds:

$$\|\mathbf{x}\|^2 = \mathbf{x} \bullet \mathbf{x} = x_1^2 + \cdots + x_n^2.$$

Then the angle $\theta$ between $\mathbf{x}$ and $\mathbf{y}$ satisfies the following equation, which is analogous to the law of cosines:

$$\cos\theta = \frac{\mathbf{x} \bullet \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\mathbf{x} \bullet \mathbf{y}}{\sqrt{\mathbf{x} \bullet \mathbf{x}} \cdot \sqrt{\mathbf{y} \bullet \mathbf{y}}}$$

The set $\mathbb{R}^n$ together with the dot product function $\bullet : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called *$n$-dimensional Euclidean space*.

There is one further algebraic operation that we can define on vectors. For any vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ we have defined

$$-\mathbf{v} = -\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} -v_1 \\ \vdots \\ -v_n \end{pmatrix}$$

and we have observed that $-\mathbf{v}$ is the vector with the same length as $\mathbf{v}$, but with **opposite direction**. More generally, we make the following definition.

**Definition of Scalar Multiplication**

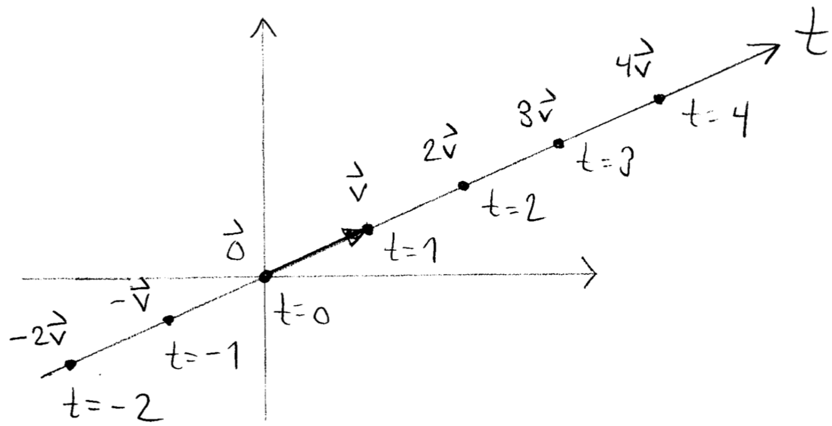
Let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ be any $n$-dimensional vector and let $t \in \mathbb{R}$ be any real number, called a *scalar*. Then we define the vector $t\mathbf{v} \in \mathbb{R}^n$ as follows:

$$t\mathbf{v} = t\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} tv_1 \\ \vdots \\ tv_n \end{pmatrix}.$$

What is the geometric meaning of this operation? We observe that the length of $t\mathbf{v}$ satisfies

$$\|t\mathbf{v}\| = \sqrt{(tv_1)^2 + \cdots + (tv_n)^2} = \sqrt{t^2(v_1^2 + \cdots + v_n^2)} = \sqrt{t^2} \cdot \sqrt{v_1^2 + \cdots + v_n} = |t|\|\mathbf{v}\|.$$

In other words, multiplying a vector $\mathbf{v} \in \mathbb{R}^n$ by a number $t \in \mathbb{R}$ "scales its length" by the factor $|t|$. This is why we call numbers scalars. The direction of $t\mathbf{v}$ is the same as $\mathbf{v}$ when $t > 0$ and it is opposite to $\mathbf{v}$ when $t < 0$. The concept of scalar multiplication is easier to understand if you visualize the set of points $\{t\mathbf{v} : t \in \mathbb{R}\}$ as the infinite line generated by the vector $\mathbf{v}$:




We say that this line is "parametrized by $t$." I would even call it the "$t$-axis." Now we have three different algebraic operations related to vectors:[11]

- *Addition* $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ takes a (vector,vector) pair to a vector.
- *Scalar multiplication* $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ takes (scalar,vector) pair to a vector.
- The *dot product* $\bullet : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ takes a (vector,vector) pair to a scalar.

It is worth recording all of the abstract properties that these operations must satisfy.

**Properties of Vector Arithmetic**

For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and for any scalars $a, b \in \mathbb{R}$ we have

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- $\mathbf{u} + \mathbf{0} = \mathbf{u}$
- $0\mathbf{u} = \mathbf{0}$
- $1\mathbf{u} = \mathbf{u}$
- $a(b\mathbf{u}) = (ab)\mathbf{u}$

[11] For any sets $S$ and $T$, the notation $S \times T$ denotes the set of ordered pairs $(s, t)$ with $s \in S$ and $t \in T$. It has nothing to do with vector multiplication.

- $a(\mathbf{u}+\mathbf{v}) = a\mathbf{u}+a\mathbf{v}$
- $(a+b)\mathbf{u} = a\mathbf{u}+b\mathbf{u}$
- $\mathbf{u}\bullet\mathbf{v} = \mathbf{v}\bullet\mathbf{u}$
- $\mathbf{u}\bullet(\mathbf{v}+\mathbf{w}) = \mathbf{u}\bullet\mathbf{v}+\mathbf{u}\bullet\mathbf{w}$
- $a(\mathbf{u}\bullet\mathbf{v}) = (a\mathbf{u})\bullet\mathbf{v} = \mathbf{u}\bullet(a\mathbf{v})$

These properties are mostly obvious so we won't bother proving them.[12] Instead, here is an example showing how to use the rules.

**?D Example:** Let $\mathbf{x}$ and $\mathbf{y}$ be any two vectors satisfying

$$\mathbf{x}\bullet\mathbf{x} = \mathbf{y}\bullet\mathbf{y} = 1 \quad \text{and} \quad \mathbf{x}\bullet\mathbf{y} = 0.$$

(I won't tell you the dimension of these vectors because it doesn't matter.) Compute the angle between the vectors $\mathbf{u} = \mathbf{x}+2\mathbf{y}$ and $\mathbf{v} = 3\mathbf{x}+\mathbf{y}$.

**Solution 1:** We can solve this problem without thinking by applying the rules of vector arithmetic. First we compute the dot products $\mathbf{u}\bullet\mathbf{u}$, $\mathbf{v}\bullet\mathbf{v}$ and $\mathbf{u}\bullet\mathbf{v}$:

$$\begin{aligned}
\mathbf{u}\bullet\mathbf{u} &= (\mathbf{x}+2\mathbf{y})\bullet(\mathbf{x}+2\mathbf{y}) = \mathbf{x}\bullet\mathbf{x}+4\mathbf{x}\bullet\mathbf{y}+4\mathbf{y}\bullet\mathbf{y} = 1+4\cdot 0+4\cdot 1 = 5,\\
\mathbf{v}\bullet\mathbf{v} &= (3\mathbf{x}+\mathbf{y})\bullet(3\mathbf{x}+\mathbf{y}) = 9\mathbf{x}\bullet\mathbf{x}+6\mathbf{x}\bullet\mathbf{y}+\mathbf{y}\bullet\mathbf{y} = 9\cdot 1+6\cdot 0+1 = 10,\\
\mathbf{u}\bullet\mathbf{v} &= (\mathbf{x}+2\mathbf{y})\bullet(3\mathbf{x}+\mathbf{y}) = 3\mathbf{x}\bullet\mathbf{x}+7\mathbf{x}\bullet\mathbf{y}+2\mathbf{y}\bullet\mathbf{y} = 3\cdot 1+7\cdot 0+2\cdot 1 = 5.
\end{aligned}$$

If $\theta$ is the angle between $\mathbf{u}$ and $\mathbf{v}$ then it follows that

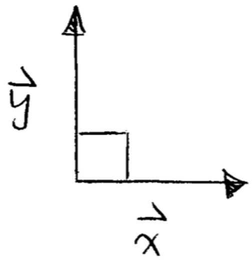
$$\cos\theta = \frac{\mathbf{u}\bullet\mathbf{v}}{\sqrt{\mathbf{u}\bullet\mathbf{u}}\cdot\sqrt{\mathbf{v}\bullet\mathbf{v}}} = \frac{5}{\sqrt{5}\cdot\sqrt{10}} = \frac{1}{\sqrt{2}}$$

and hence

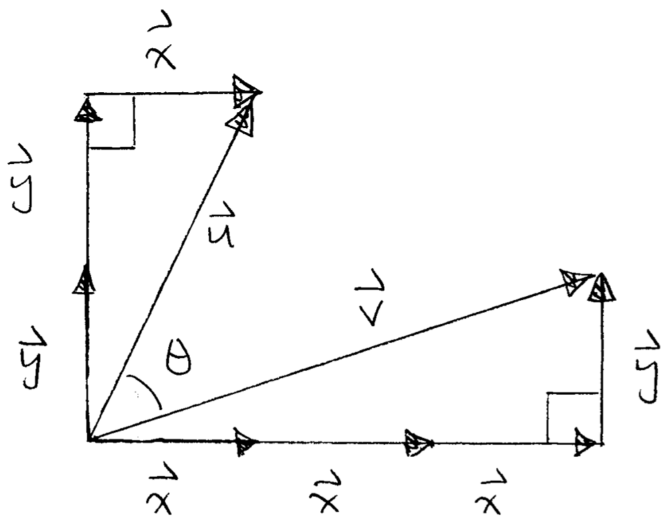
$$\theta = \arccos\left(\frac{1}{\sqrt{2}}\right) = 45^\circ.$$

**Solution 2:** Or we can draw a picture. The identities $\mathbf{x}\bullet\mathbf{x} = \mathbf{y}\bullet\mathbf{y} = 1$ tell me that $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ and the identity $\mathbf{x}\bullet\mathbf{y} = 0$ tells me that $\mathbf{x}$ and $\mathbf{y}$ are perpendicular. We say that $\mathbf{x}$ and $\mathbf{y}$ are perpendicular unit vectors. Here is a picture:

[12] You will prove the identity $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$ on the homework.




(This picture might live in 10-dimensional space. It doesn't matter.) Now we can add the vectors $\mathbf{u} = \mathbf{x} + 2\mathbf{y}$ and $\mathbf{v} = 3\mathbf{x} + \mathbf{y}$ to our picture:




From this point of view we see that the angle between $\mathbf{u}$ and $\mathbf{v}$ is the same as the angle between the vectors $(1,2)$ and $(3,1)$ in the Cartesian plane. It follows that

$$\cos\theta = \frac{(1,2)\bullet(3,1)}{\|(1,2)\|\|(3,1)\|} = \frac{1\cdot 3 + 2\cdot 1}{\sqrt{1^2+2^2}\cdot\sqrt{3^2+1^1}} = \frac{5}{\sqrt{5}\cdot\sqrt{10}} = \frac{1}{\sqrt{2}},$$

and we get the same answer $\theta = 45^\circ$. The key idea here is to treat the vectors $\mathbf{x}$ and $\mathbf{y}$ as the "basis for a coordinate system." Then we might as well express any vector $a\mathbf{x} + b\mathbf{y}$ using the following notation:

$$a\mathbf{x} + b\mathbf{y} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Since the basis vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular and have length 1 (we say that they form an "ortho-normal basis") it turns out that any computation using this language will give the correct answer. For example:[13]

$$\mathbf{u} \bullet \mathbf{v} = (\mathbf{x} + 2\mathbf{y}) \bullet (3\mathbf{x} + \mathbf{y}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \bullet \begin{pmatrix} 3 \\ 1 \end{pmatrix} = 1 \cdot 3 + 2 \cdot 1 = 5.$$

///

The point of view in this example is rather abstract. There are two reasons that we might want to work with an abstract coordinate system such as $\mathbf{x}$ and $\mathbf{y}$ instead of the usual Cartesian coordinates:

- Some objects, such as a 1D line or a 2D plane in 3D space, do not come with a standard coordinate system. In this case we have to introduce our own coordinate system. We will see many examples of this in the next chapter.
- In the twentieth century, mathematicians observed that there are structures other than Cartesian space that obey the same rules of vector arithmetic. (We call these *abstract vector spaces*.) For example, in the theory of probability we can think of random variables as "vectors" and we can think of covariance as the "dot product." For another example, the subject of signal processing is based on the idea of treating a signal as some kind of "vector." If $f(t)$ represents the amplitude of a signal at time $t$, then the "dot product" between two signals $f(t)$ and $g(t)$ is defined by the integral

  $$\int f(t)g(t)\,dt.$$

  These kinds of abstract vector spaces do not come with a built-in coordinate system; if you want to work in coordinates then you have to define your own.[14] We won't purse exotic kinds of vector spaces in this course, but you should be aware that they exist and that they follow the same general theory.

## 1.7 Exercises with Solutions

**1.A.** Define the standard basis vectors $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$ and $\mathbf{e}_3 = (0, 0, 1)$.

(a) Draw the cube with the following 8 vertices:

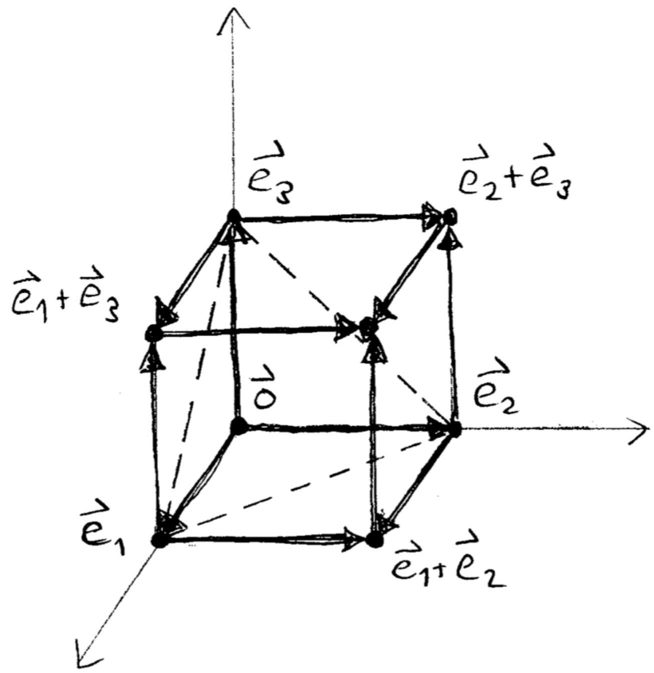
$$\mathbf{0},\ \mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3,\ \mathbf{e}_1 + \mathbf{e}_2,\ \mathbf{e}_1 + \mathbf{e}_3,\ \mathbf{e}_2 + \mathbf{e}_3,\ \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3.$$

(b) Draw the triangle in 3D with corners at $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Compute the side lengths and the angles of this triangle by using the dot product.

[13] Remark: If the basis vectors $\mathbf{x}$ and $\mathbf{y}$ are **not** ortho-normal then addition and scalar multiplication are still correct, but the dot product will give the wrong answer.

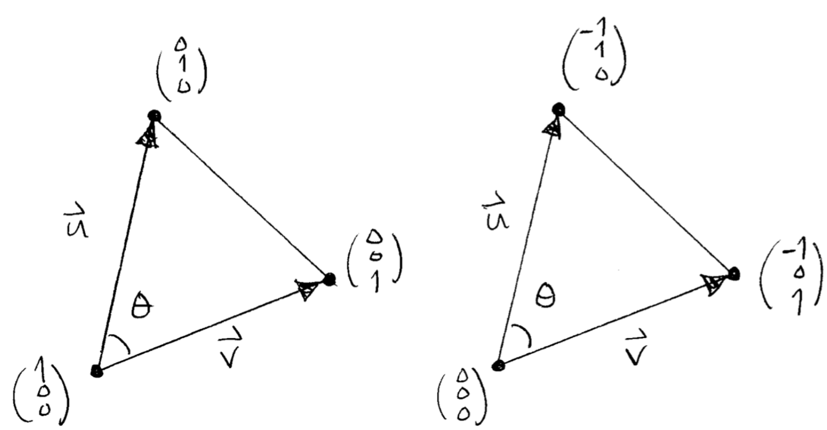
[14] In the theory of signal processing it is common to use the "pure sine waves" $\sin(t), \sin(2t), \sin(3t), \ldots$ and $\cos(t), \cos(2t), \cos(3t), \ldots$ as a coodinate system.

(a) Observe that the cube is made out of many parallelograms (actually, they are squares):




(b) Now consider the dotted triangle with vertices $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Since each side of the triangle is the diagonal of a unit square, we conclude from the Pythagorean Theorem that each side has length $\sqrt{2}$. Then since all three sides of the triangle have the same length, we conclude that the three interior angles are equal, i.e., each is $60°$. Let's check this result using the dot product.

Let $\mathbf{u}$ and $\mathbf{v}$ be the vectors with tails at $\mathbf{e}_1$ and heads at $\mathbf{e}_2$ and $\mathbf{e}_3$, respectively. In order to compute the dot products $\mathbf{u} \bullet \mathbf{u}$, $\mathbf{v} \bullet \mathbf{v}$ and $\mathbf{u} \bullet \mathbf{v}$ we translate these vectors into standard position by subtracting $\mathbf{e}_1$ from all three points:



Using the fact that $\mathbf{u}(-1,1,0)$ and $\mathbf{v}=(-1,0,1)$ gives

$$\begin{aligned}
\mathbf{u}\bullet\mathbf{u} &= (-1)^2+1^2+0^2=2,\\
\mathbf{v}\bullet\mathbf{v} &= (-1)^2+0^2+1^2=2,\\
\mathbf{u}\bullet\mathbf{v} &= (-1)(-1)+1\cdot 0+0\cdot 1=1.
\end{aligned}$$

We conclude that $\|\mathbf{u}\|=\|\mathbf{v}\|=\sqrt{2}$ and

$$\cos\theta=\frac{\mathbf{u}\bullet\mathbf{v}}{\sqrt{\mathbf{u}\bullet\mathbf{u}}\cdot\sqrt{\mathbf{v}\bullet\mathbf{v}}}=\frac{1}{\sqrt{2}\cdot\sqrt{2}}=\frac{1}{2},$$

which implies $\theta=\arccos(1/2)=60^\circ$. Alternatively, we can use vector arithmetic and the fact that the standard basis vectors satisfy

$$\mathbf{e}_i\bullet\mathbf{e}_j=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{if } i\neq j.\end{cases}$$

Then since $\mathbf{u}=\mathbf{e}_2-\mathbf{e}_1$ and $\mathbf{v}=\mathbf{e}_3-\mathbf{e}_1$ we have

$$\begin{aligned}
\mathbf{u}\bullet\mathbf{u} &= (\mathbf{e}_2-\mathbf{e}_1)\bullet(\mathbf{e}_2-\mathbf{e}_1)=\mathbf{e}_2\bullet\mathbf{e}_2-2\mathbf{e}_1\bullet\mathbf{e}_2+\mathbf{e}_1\bullet\mathbf{e}_1=1-2\cdot 0+1=2,\\
\mathbf{v}\bullet\mathbf{v} &= (\mathbf{e}_3-\mathbf{e}_1)\bullet(\mathbf{e}_3-\mathbf{e}_1)=\mathbf{e}_3\bullet\mathbf{e}_3-2\mathbf{e}_1\bullet\mathbf{e}_3+\mathbf{e}_1\bullet\mathbf{e}_1=1-2\cdot 0+1=2,\\
\mathbf{u}\bullet\mathbf{v} &= (\mathbf{e}_2-\mathbf{e}_1)\bullet(\mathbf{e}_3-\mathbf{e}_1)=\mathbf{e}_2\bullet\mathbf{e}_3-\mathbf{e}_1\bullet\mathbf{e}_2-\mathbf{e}_1\bullet\mathbf{e}_3+\mathbf{e}_1\bullet\mathbf{e}_1=0-0-0+1=1.
\end{aligned}$$

The computations for the other two angles are similar, so we omit them.

**1.B.** Let $\mathbf{u}=(1,2)$ and $\mathbf{v}=(3,1)$.

(a) Draw the points $\mathbf{u}$ and $\mathbf{v}$ together with the points

$$\frac{1}{2}\mathbf{u}+\frac{1}{2}\mathbf{v},\quad \frac{1}{4}\mathbf{u}+\frac{3}{4}\mathbf{v},\quad \frac{1}{4}\mathbf{u}+\frac{1}{4}\mathbf{v},\quad \mathbf{u}+\mathbf{v}.$$

(b) Draw the infinite line $\{t\mathbf{v}+t\mathbf{u}\}$ where $t$ is any real number. [Hint: It is enough to draw two points on this line and then use a ruler.]

(c) Draw the infinite line $\{s\mathbf{u}+(1-s)\mathbf{v}\}$ where $s$ is any real number. [Hint: Same as (b).]

(d) Shade the finite region of the plane defined by $\{s\mathbf{u}+t\mathbf{v}: 0\leqslant s\leqslant 1, 0\leqslant t\leqslant 1\}$.

(e) Shade the infinite region of the plane defined by $\{s\mathbf{u}+t\mathbf{v}: 0\leqslant s, 0\leqslant t\}$.

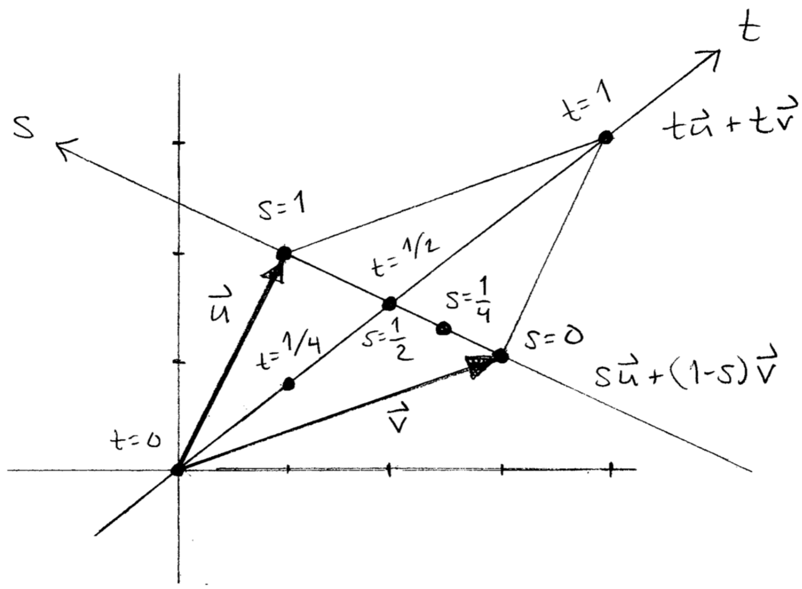
(a) First we observe that the points $\mathbf{0}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$ are the four vertices of a parallelogram. Next let me observe that

$$\frac{1}{2}\mathbf{u}+\frac{1}{2}\mathbf{v}=\frac{1}{2}(\mathbf{u}+\mathbf{v})=\frac{\mathbf{u}+\mathbf{v}}{2}.$$

We can view this as the fourth vertex of the parallelogram with vertices $\mathbf{0}$, $\mathbf{u}/2$ and $\mathbf{v}/2$. Or we can view it as half of the vector $\mathbf{u}+\mathbf{v}$. Or we can view it at the midpoint (the center of mass) of the two points $\mathbf{u}$ and $\mathbf{v}$. More generally, the midpoint of any two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is $(\mathbf{x}+\mathbf{y})/2$. (Can you explain why?) It follows that $\frac{1}{4}\mathbf{u} + \frac{1}{4}\mathbf{v} = \frac{1}{4}(\mathbf{u}+\mathbf{v})$ is the midpoint between $\mathbf{0}$ and $\frac{1}{2}(\mathbf{u}+\mathbf{v})$, and $\frac{1}{4}\mathbf{u} + \frac{3}{4}\mathbf{v}$ is the midpoint between $\frac{1}{2}\mathbf{u} + \frac{1}{2}\mathbf{v}$ and $\mathbf{v}$:

$$\frac{(\mathbf{u}/2 + \mathbf{v}/2) + \mathbf{v}}{2} = \frac{\mathbf{u}/2 + 3\mathbf{v}/2}{2} = \frac{1}{2}\left(\frac{1}{2}\mathbf{u} + \frac{3}{2}\mathbf{v}\right) = \frac{1}{4}\mathbf{u} + \frac{3}{4}\mathbf{v}.$$

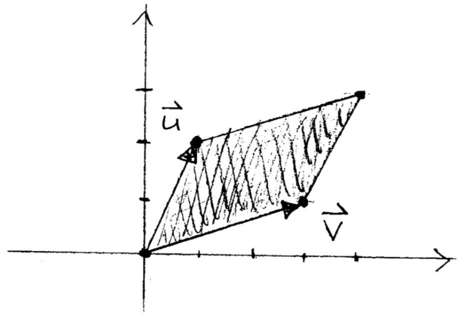
Here is the picture:




(b) I have drawn the line $t\mathbf{u} + t\mathbf{v} = t(\mathbf{u}+\mathbf{v})$ in the picture above. The line passes through the points $\mathbf{0}$, $(\mathbf{u}+\mathbf{v})/4$, $(\mathbf{u}+\mathbf{v})/2$ and $\mathbf{u}+v$, in that order.

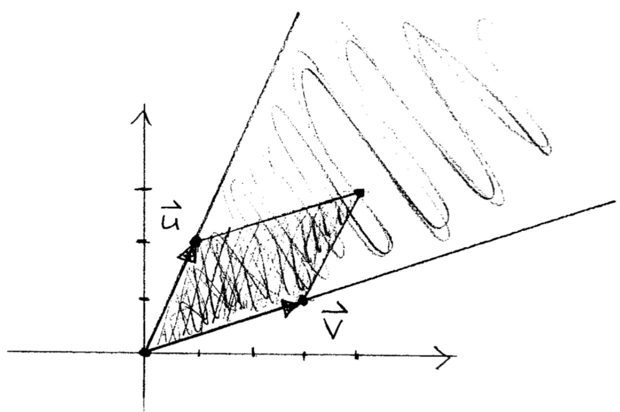
(c) I have drawn the line $s\mathbf{u} + (1-s)\mathbf{v}$ in the picture above. This line passes through the points $\mathbf{v}$, $\mathbf{u}/4 + 3\mathbf{v}/4$, $\mathbf{u}/2 + \mathbf{v}/2$ and $\mathbf{u}$, in that order. We can also think of this as the line $\mathbf{v} + s(\mathbf{u}-\mathbf{v})$ that starts at the point $\mathbf{v}$ when $s = 0$ and then moves in the direction of the vector $\mathbf{u}-\mathbf{v}$ as $s$ increases.

(d) The set of points $\{s\mathbf{u} + t\mathbf{v} : 0 \leqslant s \leqslant 1, 0 \leqslant t \leqslant 1\}$ is the filled parallelogram with vertices at $\mathbf{0}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$:




We call this the "parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$."

(e) The set of points $\{s\mathbf{u} + t\mathbf{v} : 0 \leqslant s, 0 \leqslant t\}$ is the "cone" spanned by $\mathbf{u}$ and $\mathbf{v}$:




We can think of this as the "first quadrant" of the coordinate system defined by $\mathbf{u}$ and $\mathbf{v}$. This coordinate system is a bit squished.

**1.C.** Let $\mathbf{u}$ and $\mathbf{v}$ be any two vectors of length 2. Compute the following dot products:

(a) $\mathbf{u} \bullet (-\mathbf{u})$

(b) $(\mathbf{u} + \mathbf{v}) \bullet (\mathbf{u} - \mathbf{v})$

(c) $(\mathbf{u} + 2\mathbf{v}) \bullet (\mathbf{u} - 2\mathbf{v})$

(d) Explain why you do not have enough information to compute $(\mathbf{u} - \mathbf{v}) \bullet (\mathbf{u} - \mathbf{v})$.

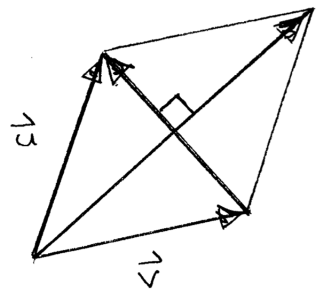
(a) Since $\mathbf{u}$ has length 2 we know that $\mathbf{u} \bullet \mathbf{u} = \|\mathbf{u}\|^2 = 4$ and hence

$$\mathbf{u} \bullet (-\mathbf{u}) = \mathbf{u} \bullet ((-1)\mathbf{u}) = (-1)\mathbf{u} \bullet \mathbf{u} = -4.$$

(b) Since $\mathbf{v}$ has length 2 we also know that $\mathbf{v} \bullet \mathbf{v} = \|\mathbf{v}\|^2 = 4$ and hence

$$(\mathbf{u}+\mathbf{v}) \bullet (\mathbf{u}-\mathbf{v}) = \mathbf{u} \bullet \mathbf{u} - \cancel{\mathbf{u} \bullet \mathbf{v}} + \cancel{\mathbf{u} \bullet \mathbf{v}} - \mathbf{v} \bullet \mathbf{v} = \mathbf{u} \bullet \mathbf{u} - \mathbf{v} \bullet \mathbf{v} = 4 - 4 = 0.$$

Evidently the vectors $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}-\mathbf{v}$ are perpendicular. I do not know the angle between $\mathbf{u}$ and $\mathbf{v}$, but here is a picture:



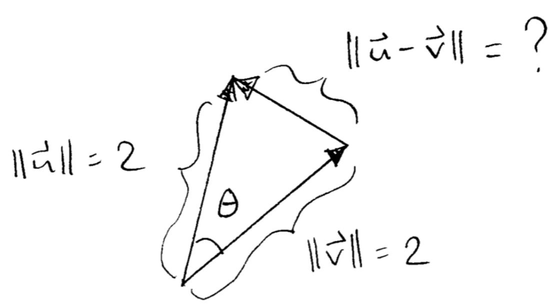

We can regard this as a theorem that the two diagonals of a rhombus are perpendicular. [If $\mathbf{u}$ and $\mathbf{v}$ did not have the same length, this parallelogram would **not** be a rhombus and its diagonals would **not** be perpendicular.]

(c) We have

$$(\mathbf{u}+2\mathbf{v}) \bullet (\mathbf{u}-2\mathbf{v}) = \mathbf{u} \bullet \mathbf{u} - \cancel{2\mathbf{u} \bullet \mathbf{v}} + \cancel{2\mathbf{v} \bullet \mathbf{u}} - 4\mathbf{v} \bullet \mathbf{v} = \mathbf{u} \bullet \mathbf{u} - 4\mathbf{v} \bullet \mathbf{v} = -12.$$

I cannot draw a picture of this.

(d) $(\mathbf{u}-\mathbf{v}) \bullet (\mathbf{u}-\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|^2$, where $\|\mathbf{u}-\mathbf{v}|$ is the length of the third side of an isoceles triangle:




In terms of algebra, we have

$$\begin{aligned}\|\mathbf{u}-\mathbf{v}\|^2 &= \mathbf{u} \bullet \mathbf{u} - \mathbf{u} \bullet \mathbf{v} - \mathbf{u} \bullet \mathbf{v} + \mathbf{v} \bullet \mathbf{v}, \\ &= \mathbf{u} \bullet \mathbf{u} - 2\mathbf{u} \bullet \mathbf{v} + \mathbf{v} \bullet \mathbf{v}, \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \bullet \mathbf{v}\end{aligned}$$

$$= 8 - 2\mathbf{u} \bullet \mathbf{v}.$$

We cannot compute this because were not given the dot product $\mathbf{u} \bullet \mathbf{v}$. Equivalently, we were not given the angle $\theta$ between $\mathbf{u}$ and $\mathbf{v}$.

**1.D. Lines and Planes.**

(a) Let $\mathbf{u} = (2, 1)$. The following set of points is a line:

$$\{\mathbf{x} = (x, y) : \mathbf{x} \bullet \mathbf{u} = 0\}.$$

Draw this line and find its equation.

(b) Now let $\mathbf{x}_0 = (3, 1)$ and $\mathbf{u} = (2, 1)$. The following set of points is also a line:

$$\{\mathbf{x}_0 + t\mathbf{u} : t \in \mathbb{R}\}.$$

Draw this line and find its equation.

(c) Now let $\mathbf{u} = (1, 1, 1)$. Describe the shape formed by the following set of points:

$$\{\mathbf{x} = (x, y, z) : \mathbf{x} \bullet \mathbf{u} = 0\}.$$

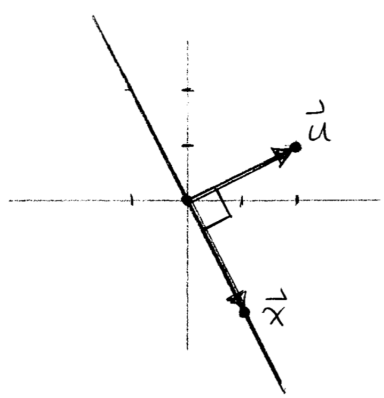
Try to draw a picture.

(d) Let $\mathbf{u} = (1, 1, 1)$ and $\mathbf{v} = (1, 2, 3)$. Describe the shape of the following set of points:

$$\{\mathbf{x} = (x, y, z) : \mathbf{x} \bullet \mathbf{u} = 0 \text{ and } \mathbf{x} \bullet \mathbf{v} = 0\}.$$

Try to draw a picture.

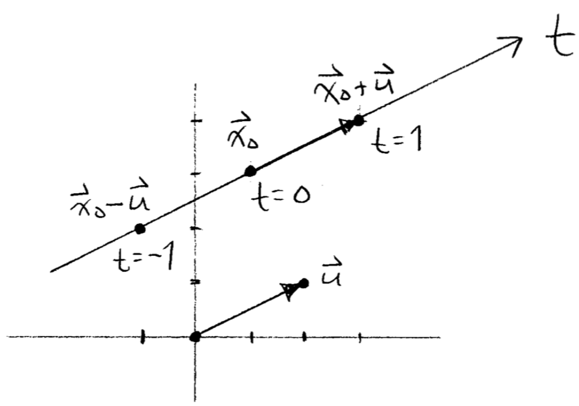
(a) The equation $\mathbf{x} \bullet \mathbf{u} = 0$ means that the vector $\mathbf{x}$ is perpendicular to the vector $\mathbf{u}$. The set of all such points $\mathbf{x}$ forms a line through the origin:



The equation of this line is

$$\begin{aligned}
\mathbf{x} \bullet \mathbf{u} &= 0 \\
\begin{pmatrix} x \\ y \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 1 \end{pmatrix} &= 0 \\
2x + y &= 0.
\end{aligned}$$

(b) The line $\{\mathbf{x}_0 + t\mathbf{u} : t \in \mathbb{R}\}$ contains the point $\mathbf{x}_0 = (3, 1)$ and is parallel to the vector $\mathbf{u} = (2, 1)$. We can think of $\mathbf{x}_0 + t\mathbf{u}$ as the position of a particle at time $t$. When $t = 0$ the particle is at the point $\mathbf{x}_0$. Then as $t$ increases the particle travels in the direction of $\mathbf{u}$ at constant velocity $\|\mathbf{u}\| = \sqrt{5}$ units per second:




The general point on the line has the form $(x, y) = (1, 3) + t(2, 1) = (1 + 2t, 3 + t)$. If you insist on expressing this line as a single equation, we can solve for $t$ to obtain

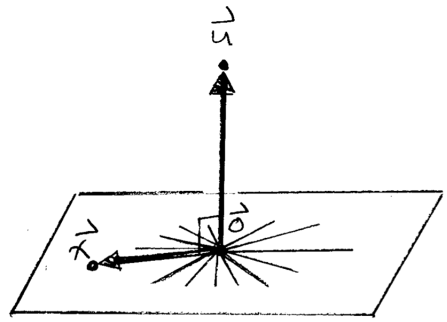
$$y - 3 = t = (x - 1)/2$$

and hence

$$\begin{aligned}
y - 3 &= (x - 1)/2 \\
y &= \frac{1}{2}x + \frac{5}{2}.
\end{aligned}$$

Indeed, this is the line with slope $1/2$ and $y$-intercept $5/2$.

(c) The equation $\mathbf{x} \bullet \mathbf{u} = 0$ means that the vector $\mathbf{x}$ is perpendicular to the vector $\mathbf{u}$. The set of all such points $\mathbf{x}$ forms a 2D plane in 3D space:

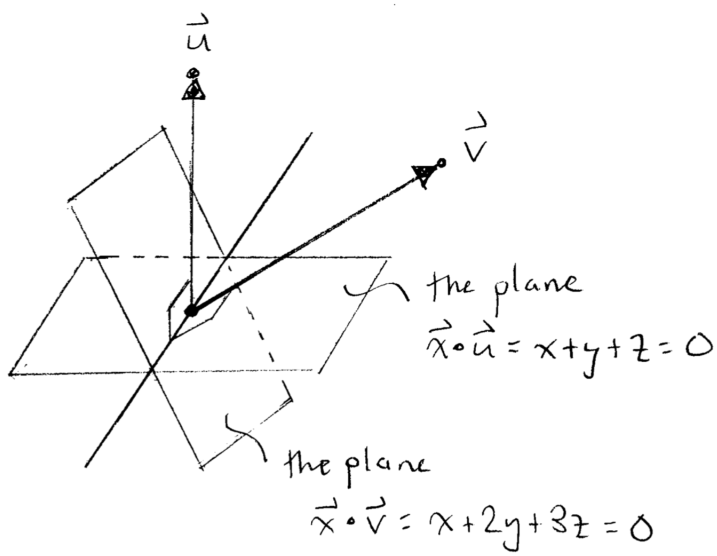



The equation of the plane is

$$\mathbf{x} \bullet \mathbf{u} = 0$$
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$
$$x + y + z = 0.$$

This is the unique plane that is perpendicular to the vector $\mathbf{u}$ and contains the point $\mathbf{0}$.

(d) Now consider the vector $\mathbf{v} = (1, 2, 3)$. From part (c) we know that $\mathbf{x} \bullet \mathbf{u} = 0$ and $\mathbf{x} \bullet \mathbf{v} = 0$ are the equations of the two planes that are perpendicular to $\mathbf{u}$ and $\mathbf{v}$, respectively, and pass through the origin $\mathbf{0}$. The set $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \bullet \mathbf{u} = 0 \text{ and } \mathbf{x} \bullet \mathbf{v} = 0\}$ is the intersection the planes:




In other words, the set is a line. What more can we say about this line? **Implicitly**, we can describe the line as the solution to the following system of two linear equations in three unknowns:

$$\begin{cases} x + \phantom{2}y + \phantom{3}z = 0, \\ x + 2y + 3z = 0. \end{cases}$$

**Explicitly**, we can describe this line as the set of points $\{t\mathbf{w} : t \in \mathbb{R}\}$, where $\mathbf{w}$ is some specific vector that is simultaneously perpendicular to $\mathbf{u}$ and $\mathbf{v}$. How can we find such a vector? We will discuss this in the next chapter.[15]

**1.E. Associativity of Vector Addition.** Consider three vectors in the plane:

$$\mathbf{u} = (u_1, u_2), \quad \mathbf{v} = (v_1, v_2) \quad \text{and} \quad \mathbf{w} = (w_1, w_2).$$

(a) Use algebra to prove that $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$. [Hint: You may assume that addition of numbers is associative.]

(b) Draw a picture to demonstrate that $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.

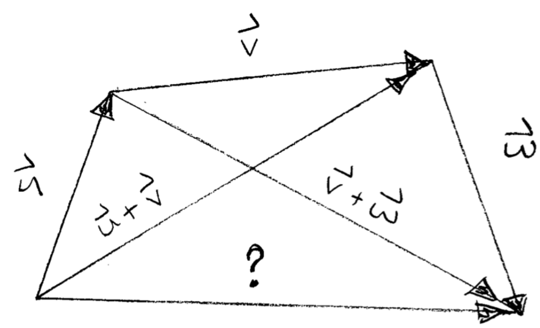
(a) We may assume that $u_1 + (v_1 + w_1) = (u_1 + v_1) + w_1$ and $u_2 + (v_2 + w_2) = (u_2 + v_2) + w_2$. Then it follows that

$$\begin{aligned}
\mathbf{u} + (\mathbf{v} + \mathbf{w}) &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \left( \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \right) \\
&= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix} \\
&= \begin{pmatrix} u_1 + (v_1 + w_2) \\ u_2 + (v_2 + w_2) \end{pmatrix} \\
&= \begin{pmatrix} (u_1 + v_1) + w_2 \\ (u_2 + v_2) + w_2 \end{pmatrix} \\
&= \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \\
&= \left( \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \\
&= (\mathbf{u} + \mathbf{v}) + \mathbf{w}.
\end{aligned}$$

This proof was supposed to be humorous. It doesn't really explain anything.

(b) The following proof is much more enlightening:

[15]If you learned about the "cross product" in your physics course, you can check that $\mathbf{w} = \mathbf{u} \times \mathbf{v} = (1, -2, 1)$ is one such vector. The cross product is a useful trick for 3D calculations that was discovered by Hamilton in 1843. I will not emphasize the cross product in this course because it does not generalize easily to higher dimensions.




The unlabeled arrow has two different names. What are they?

# 2 Systems of Linear Equations

## 2.1 Simultaneous of Equations

At the beginning of Chapter 1 we discussed the basic idea of coordinate geometry. By identifying a "point" with an "ordered list of numbers" we obtain a dictionary between geometric concepts and algebraic concepts. In this section we will discuss the correspondence between "equations" and "shapes."[16]

**The Idea of Analytic Geometry**

An ordered list of $n$ real numbers $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ corresponds to a point in $n$-dimensional space. If $f : \mathbb{R}^n \to \mathbb{R}$ is any function with $n$ inputs and 1 output and if $c \in \mathbb{R}$ is any real constant then the set of points $(x_1, x_2, \ldots, x_n)$ satisfying the equation

$$f(x_1, x_2, \ldots, x_n) = c$$

forms a "shape in space."[17] The set of points satisfying a system of simultaneous equations corresponds to the "intersection of the shapes." That is, if $H_i \subseteq \mathbb{R}^n$ is the shape corresponding to the equation $f_i(x_1, x_2, \ldots, x_n) = c_i$, then the system of simultaneous

[16]The most basic version of this correspondence is called *analytic geometry*. The terms *arithmetic geometry* and *algebraic geometry* would also be appropriate, but sadly these terms have technical meanings that are beyond the scope of our course.

equations

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n) &= c_1, \\ f_2(x_1, x_2, \ldots, x_n) &= c_2, \\ &\vdots \\ f_m(x_1, x_2, \ldots, x_n) &= c_m, \end{cases}$$

corresponds to the shape

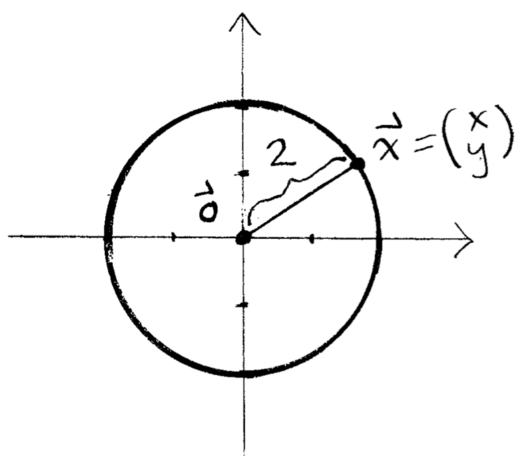
$$H_1 \cap H_2 \cap \cdots \cap H_m \subseteq \mathbb{R}^n.$$ [18]

That definition looks fancy because I expressed it in complete generality. The examples are not so fancy.

**2D Example:** Consider the following system of two equations in two unknowns:

$$\begin{cases} x^2 + y^2 &= 4, \\ 2x + y &= 0. \end{cases}$$

Recall that $x^2 + y^2 = 4$ is the equation of the circle with radius 2 centered at the origin:




Indeed, this circle is **defined** as the set of points $\mathbf{x} \in \mathbb{R}^2$ that have distance 2 from the origin:

$$\|\mathbf{x}\| = 2$$
$$\|\mathbf{x}\|^2 = 4$$
$$\mathbf{x} \bullet \mathbf{x} = 4$$

[17]Later we will observe that the shape determined by one equation in $n$ variables is "$(n-1)$-dimensional."

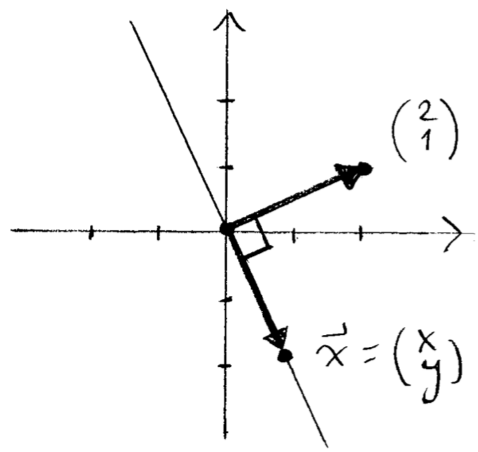
[18]The symbol "$\cap$" stands for *intersection* of sets, and the symbol "$S \subseteq T$" means that $S$ is a *subset* of $T$. The symbol "$x \in S$" means that the thing $x$ is a *member* of the set $S$.

$$x^2 + y^2 = 4.$$

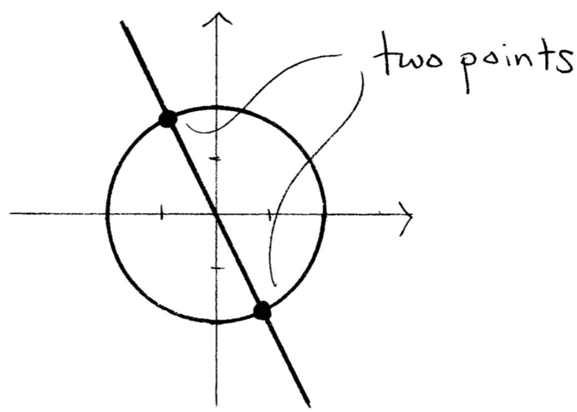
And we observed on the previous homework that the equation $2x + y = 0$ is the equivalent to

$$\begin{pmatrix} x \\ y \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0.$$

This represents the line that contains the origin and is perpendicular to the vector $(2, 1)$:




By plotting the two shapes on the same set of axes, we observe that the intersection consists of two distinct points:




To "solve the system" means to find the coordinates of these points. Since the equation $x^2 + y^2 = 4$ is "nonlinear" (i.e., the variables $x$ and $y$ occur with exponents larger than 1) there is no general algorithm for this. Luckily, in this example we can solve the second equation for $y$ to obtain $y = -2x$. Then we can substitute this into the first equation to obtain

$$x^2 + y^2 = 4$$

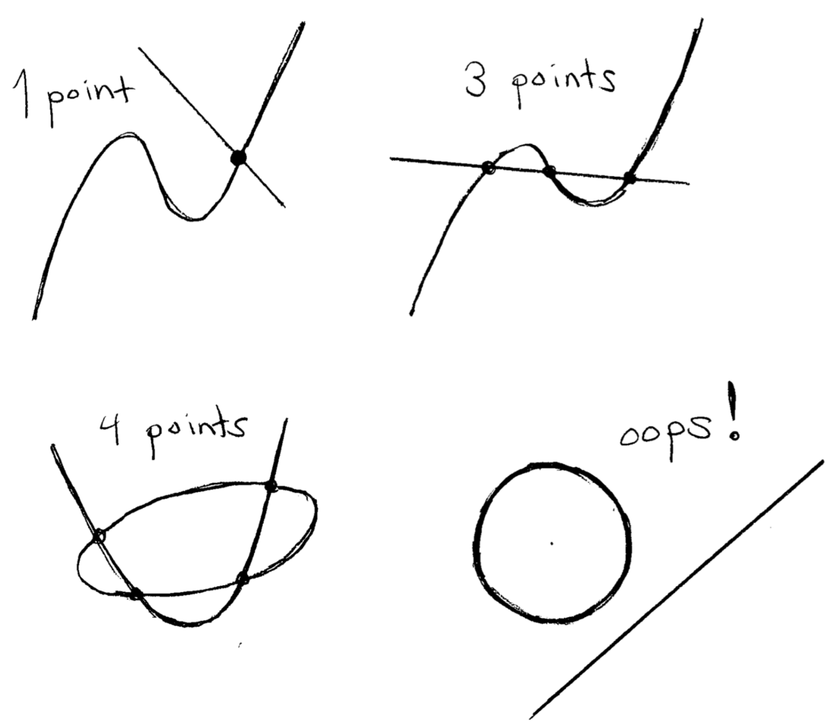
$$
\begin{aligned}
x^2 + (-2x)^2 &= 4 \\
x^2 + 4x^2 &= 4 \\
5x^2 &= 4 \\
x^2 &= 4/5 \\
x &= \pm\sqrt{4/5} \\
&= \pm 2/\sqrt{5}
\end{aligned}
$$

Note that there are two possible $x$ values, as expected. Finally, we substitute these two values back into the first equation to obtain the two points of intersection:

$$
\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} \\ -4/\sqrt{5} \end{pmatrix} = \frac{2}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2\sqrt{5} \\ 4/\sqrt{5} \end{pmatrix} = \frac{-2}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}
$$

///

Let me observe that each of the equations $x^2 + y^2 = 4$ and $2x + y = 0$ (a circle and a line, respectively) represents a "1-dimensional shape" in the plane, and their intersection represents a pair of points, which is a "0-dimensional shape." In general, any two equations $f(x, y) = c$ and $g(x, y) = d$ in two variables should represent a pair of curves in the plane. Unless these two curves are accidentally disjoint then they will intersect in a set of points, which is a "0-dimensional shape":



To be specific:

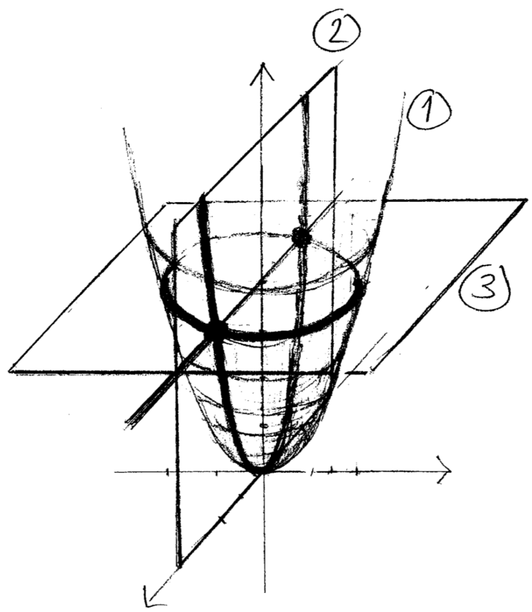
- A system of 0 equations in 2 unknowns has a 2-dimensional solution (the whole plane).
- A system of 1 equation in 2 unknowns has a 1-dimensional solution (some kind of curve).
- A system of 2 equations in 2 unknowns probably has a 0-dimensional solution (a collection of points).
- A system of 3 or more equations in 2 unknowns probably has no solution, because the third curve probably does not contain any of the intersection points of the first two curves.

Let's see what happens when there are three variables.

**3D Example:** Consider the following system of three equations in three unknowns:

$$\begin{array}{c} (1) \\ (2) \\ (3) \end{array} \begin{cases} x^2 + y^2 - z &= 0, \\ y &= 0, \\ z &= 4. \end{cases}$$

The equation (1) defines a paraboloid, while each of (2) and (3) defines a plane. The intersection of (1) and (2) is the parabola $z = y^2$ in the $yz$-plane. The intersection of (1) and (3) is the circle $x^2 + y^2 = 4$ in the plane $z = 4$ (which is parallel to the $xy$-plane), and the intersection of (2) and (3) is the line $(0, 0, 4) + t(1, 0, 0)$ (which is parallel to the $x$-axis). Finally, the intersection of all three shapes is the pair of points $(2, 0, 4)$ and $(-2, 0, 4)$. Here is a picture:



Thus we make the following observations:

- A system of 0 equations in 3 unknowns has a 3-dimensional solution (the whole space).
- A system of 1 equation in 3 unknowns has a 2-dimensional solution (some kind of surface).
- A system of 2 equations in 3 unknowns probably has a 1-dimensional solution (a curve or a collection of curves).
- A system of 3 equations in 3 unknowns probably has a 0-dimensional solution (a collection of points).
- A system of 4 or more equations in 3 unknowns probably has no solution, because the fourth surface probably does not contain any of the intersection points of the first three surfaces.

These examples lead us to the following guiding principle.

**The Dimension Principle**

In a "generic" system of equations, each new equation should reduce the dimension of the solution by one. If the number of equations is greater than the number of variables then there should be no solution.

Unfortunately, this principle has many subtle exceptions. In order to make things work out as cleanly as possible, we will restrict our attention in this course to the easiest kind of equations. Namely: linear equations.

## 2.2 What is a Hyperplane?

It turns out that solving general systems of equations is extremely difficult. In this course we will focus the easiest kind of equations.

**Definition of a Linear Equation**

A *linear equation* in $n$ unknowns has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $x_1, x_2, \ldots, x_n \in \mathbb{R}$ are real variables and $a_1, a_2, \ldots, a_n, b \in \mathbb{R}$ are real constants.

What kind of shape does a linear equation represent?

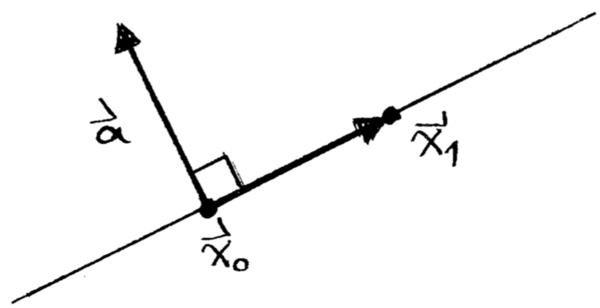
**2D Example:** The general linear equation in two unknowns has the form

$$ax + by = c.$$

I claim that this is the equation of a line that is perpendicular to the vector $\mathbf{a} = (a, b)$. Indeed, note that any point $\mathbf{x} = (x, y)$ on the line satisfies $\mathbf{a} \bullet \mathbf{x} = c$. If $\mathbf{x}_0 = (x_0, y_0)$ and $\mathbf{x}_1 = (x_1, y_1)$ are any two points on the line then we find that

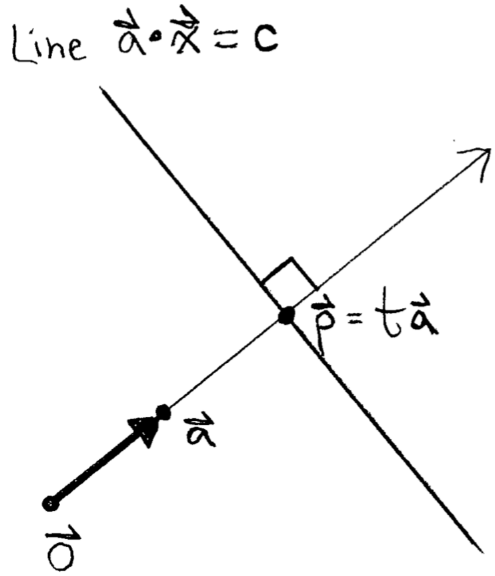
$$\mathbf{a} \bullet (\mathbf{x}_1 - \mathbf{x}_0) = \mathbf{a} \bullet \mathbf{x}_1 - \mathbf{a} \bullet \mathbf{x}_0 = c - c = 0,$$

which implies that the difference $\mathbf{x}_1 - \mathbf{x}_0$ is perpendicular to $\mathbf{a}$, as in the following picture:




But how can we actually **find** a point on this line? First let's try to set $x = 0$ so that $by = c$. If $b \neq 0$ then we find that $y = c/b$ and hence $(0, c/b)$ is a point on the line. But what if $b = 0$? In this case we can try to set $y = 0$, so that $ax = c$. If $a \neq 0$ then we conclude that $x = c/a$ and hence $(c/a, 0)$ is a point on the line. Since at least one of $a$ and $b$ is nonzero (otherwise the equation $ax + by = 0x + 0y = c$ is pretty boring), this method will always find a point on the line.

But I want to find a **good** point on the line. Suppose that $\mathbf{p}$ is the point on the line that is **closest to the origin 0**. The for geometric reasons we know that the line segment from $\mathbf{0}$ to $\mathbf{p}$ is perpendicular to the line. Equivalently, we know that the vectors $\mathbf{p}$ and $\mathbf{a}$ are parallel, hence $\mathbf{p} = t\mathbf{a} = t(a, b, c)$ for some scalar $t$. Here is a picture:




To solve for $t$ we let $\mathbf{p} = (x, y, z)$ and then substitute to get

$$\begin{aligned} ax + by &= c \\ a(ta) + b(tb) &= c \\ t(a^2 + b^2) &= c \\ t &= c/(a^2 + b^2). \end{aligned}$$

We conclude that the closest point to the origin is[19]

$$\mathbf{p} = t\begin{pmatrix} a \\ b \end{pmatrix} = \frac{c}{a^2 + b^2}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ac/(a^2 + b^2) \\ bc/(a^2 + b^2) \end{pmatrix}.$$

Furthermore, the minimum distance between the line and the origin is

$$\|\mathbf{p}\| = \left\| t\begin{pmatrix} a \\ b \end{pmatrix} \right\| = |t| \left\| \begin{pmatrix} a \\ b \end{pmatrix} \right\| = \frac{|c|}{a^2 + b^2}\sqrt{a^2 + b^2} = \frac{|c|}{\sqrt{a^2 + b^2}}.$$

Here is a summary.

**The Equation of a Line**

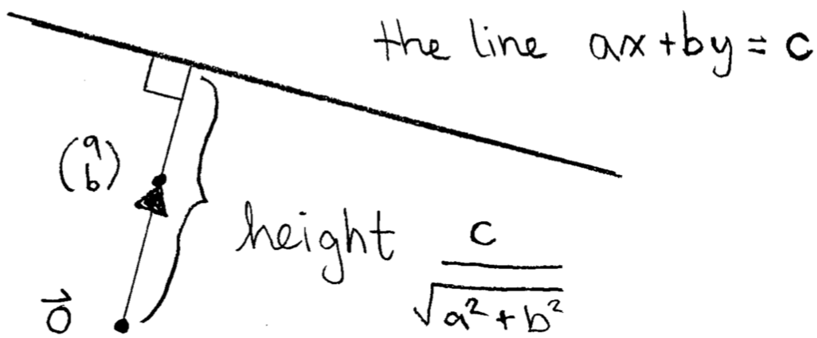
Let $a, b, c \in \mathbb{R}$ be any real numbers with $\mathbf{a} = (a, b) \neq (0, 0)$, and let $\mathbf{x} = (x, y)$ be a general point in the $xy$-plane. Then the equation

$$\mathbf{a} \bullet \mathbf{x} = c$$

[19] Later we will say that the point $\mathbf{p}$ is the "orthogonal projection of $\mathbf{0}$ onto the line $ax + by = 0$."

$$ax + by = c$$

represents the line in the $xy$-plane that is perpendicular to the vector $\mathbf{a}$ and has "height $c/\|\mathbf{a}\|$ above the origin" in the direction of $\mathbf{a}$ :




In particular, since $\|\mathbf{a}\| \neq 0$ we observe that this line passes through the origin (i.e. has zero height above the origin) if and only if $c = 0$.

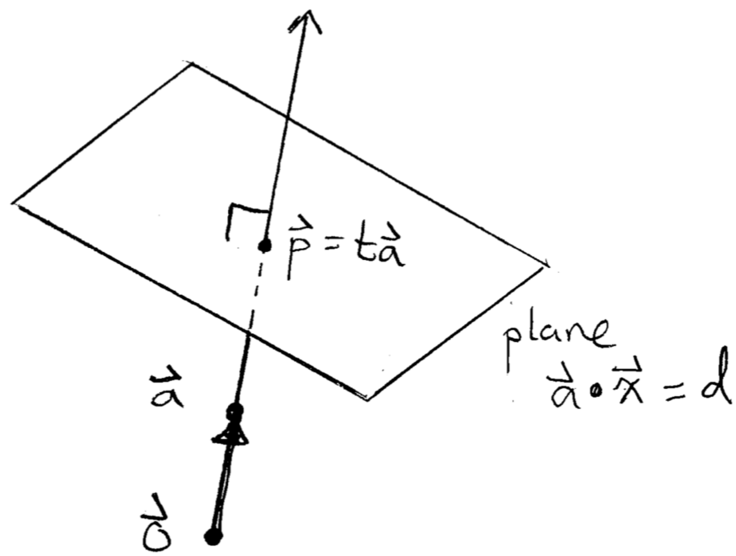
**3D Example:** The general linear equation in three unknowns has the form

$$ax + by + cz = d.$$

I claim that this is the equation of a plane that is perpendicular to the vector $\mathbf{a} = (a, b, c)$. Indeed, we can rewrite the equation as $\mathbf{a} \bullet \mathbf{x} = d$, where $\mathbf{x} = (x, y, z)$ is a general point on the line. Then for any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ on the plane we have

$$\mathbf{a} \bullet (\mathbf{x}_1 - \mathbf{x}_0) = \mathbf{a} \bullet \mathbf{x}_1 - \mathbf{a} \bullet \mathbf{x}_0 = d - d = 0,$$

which implies that the vector $\mathbf{x}_1 - \mathbf{x}_2$ is perpendicular to $\mathbf{a}$. Since this is true for any two points on the plane it follows that the whole plane is perpendicular to $\mathbf{a}$. Now let's try to find a good point on the plane. The closest point on the plane to the origin must have the form $\mathbf{p} = t\mathbf{a}$ since the line segment from $\mathbf{0}$ to $\mathbf{p}$ must be perpendicular to the plane, and hence must be parallel to $\mathbf{a}$. Here is a picture:




Since $\mathbf{p} = t\mathbf{a}$ is on the plane we know by definition that

$$\begin{aligned} \mathbf{a} \bullet \mathbf{p} &= d \\ \mathbf{a} \bullet (t\mathbf{a}) &= d \\ t(\mathbf{a} \bullet \mathbf{a}) &= d \\ t\|\mathbf{a}\|^2 &= d \\ t &= d/\|\mathbf{a}\|^2. \end{aligned}$$

We conclude that

$$\mathbf{p} = t\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{d}{\|\mathbf{a}\|^2}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} ad/\|\mathbf{a}\|^2 \\ bd/\|\mathbf{a}\|^2 \\ cd/\|\mathbf{a}\|^2 \end{pmatrix},$$

hence the minimum distance between the plane and the origin is

$$\|\mathbf{p}\| = \left\|\frac{d}{\|\mathbf{a}\|^2}\mathbf{a}\right\| = \frac{|d|}{\|\mathbf{a}\|^2}\|\mathbf{a}\| = \frac{|d|}{\|\mathbf{a}\|}.$$

Here is the general situation.

**The Equation of a Hyperplane**

Let $a_1, a_2, \ldots, a_n, b \in \mathbb{R}$ be any real constants with $\mathbf{a} = (a_1, a_2, \ldots, a_n) \neq \mathbf{0}$ and let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ be a general point in Euclidean $n$-dimensional space. Then the equation

$$\begin{aligned} \mathbf{a} \bullet \mathbf{x} &= b \\ a_1x_1 + a_2x_2 + \cdots + a_nx_n &= b \end{aligned}$$

represents a hyperplane (i.e., a flat $(n-1)$-dimensional shape). For any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ on the hyperplane we observe that

$$\mathbf{a} \bullet (\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{a} \bullet \mathbf{x}_1 - \mathbf{a} \bullet \mathbf{x}_2 = d - d = 0.$$

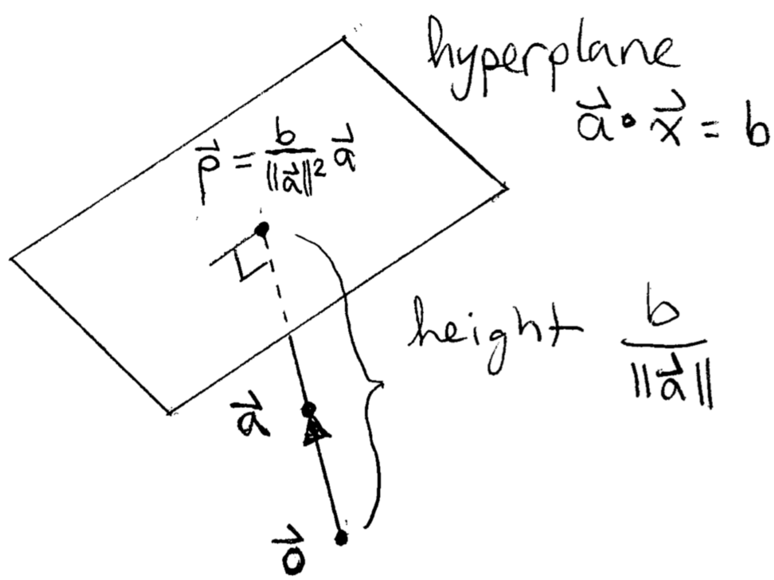
It follows that the entire hyperplane is perpendicular to $\mathbf{a}$. If $\mathbf{p}$ is the point on the hyperplane that is closest to the origin then we must have $\mathbf{p} = t\mathbf{a}$ for some scalar $t \in \mathbb{R}$ and hence

$$t\|\mathbf{a}\|^2 = t(\mathbf{a} \bullet \mathbf{a}) = \mathbf{a} \bullet (t\mathbf{a}) = \mathbf{a} \bullet \mathbf{p} = b.$$

It follows that $t = b/\|\mathbf{a}\|^2$ and hence the minimum distance between the hyperplane and the origin is

$$\|\mathbf{p}\| = \left\| \frac{b}{\|\mathbf{a}\|^2}\mathbf{a} \right\| = \frac{|b|}{\|\mathbf{a}\|^2}\|\mathbf{a}\| = \frac{|b|}{\|\mathbf{a}\|}.$$

More specifically, we observe that the hyperplane has "height $b/\|\mathbf{a}\|$ above the origin in the direction of $\mathbf{a}$."[20] Here is a picture:




[20] If $b < 0$ then the height is negative. If $\mathbf{a}$ is "up" then this means that the hyperplane is "below the origin."

## 2.3 What is a Line?

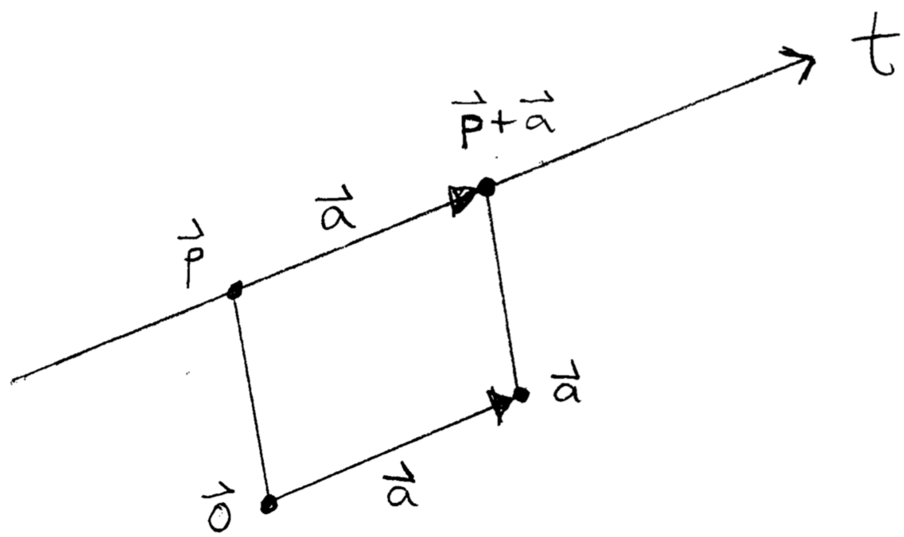
We have seen that one linear equation defines a hyperplane; it does not define a line. So what is the definition of a line?

**The Definition of a Line**

Let $\mathbf{p} \in \mathbb{R}^n$ be any point and let $\mathbf{a} \in \mathbb{R}^n$ be any vector in $n$-dimensional space. Then the following set is called a *line*:

$$\{\mathbf{p} + t\mathbf{a} : t \in \mathbb{R}\} \subseteq \mathbb{R}^n.$$

Here is a picture:




We call this the line through $\mathbf{p}$ "in the direction of $\mathbf{a}$" or ("parallel to $\mathbf{a}$"). Note that this representation is not unique because we can replace $\mathbf{p}$ by any point on the line and we can replace $\mathbf{a}$ by any scalar multiple of $\mathbf{a}$.

Unless $n = 2$ there is no such this as "the equation of a line." In fact, **at least $n-1$ equations are necessary to define a line in $n$-dimensional space**. For example, consider the line $\{\mathbf{p} + t\mathbf{a} : t \in \mathbb{R}\}$ where

$$\mathbf{p} = (p_1, p_2, \ldots, p_n) \qquad \text{and} \qquad \mathbf{a} = (a_1, a_2, \ldots, a_n).$$

Then the general point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ on the line satisfies

$$\mathbf{x} = \mathbf{p} + t\mathbf{a}$$

$$(x_1, x_2, \ldots, x_n) = (p_1 + ta_1, p_2 + ta_2, \ldots, p_n + ta_n).$$

We can think of this as a system of $n$ linear equations in the $n+1$ unknowns $x_1, x_2, \ldots, x_n, t$. Assuming that $a_i \neq 0$ for all $i$, we can eliminate $t$ from the system to obtain a sequence of $n-1$ linear equations in the $n$ unknowns $x_1, x_2, \ldots, x_n$:

$$\frac{x_1 - p_1}{a_1} = \frac{x_2 - p_2}{a_2} = \cdots = \frac{x_n - p_n}{a_n}.$$

Note that each of these equations defines a hyperplane. Thus we have expressed the parametrized line $\mathbf{p} + t\mathbf{a}$ as the intersection of $n-1$ hyperplanes. I do not want to draw a picture of this.

## 2.4 What is a $d$-Plane?

We say that a line is "one dimensional" because it can be described by one free parameter. We also say that a plane is "two dimensional," but why? Consider the following plane:

$$x + 2y + 3z = 5.$$

I claim that this plane can be "parametrized with two free parameters." There are infinitely many ways to do this so we have to make some arbitrary choices. For example, let's arbitrarily let $s = y$ and $t = z$ be free. (By doing this we are declaring that we never try to "solve for" $y$ or $z$.) Then we can solve for $x$ in terms of the two free parameters:

$$x = 5 - 2y - 3z = 5 - 2s - 3t.$$

But I prefer to express this in terms of vectors:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 - 2s - 3t \\ s \\ t \end{pmatrix} = \begin{pmatrix} 5 - 2s - 3t \\ 0 + 1s + 0t \\ 0 + 0s + 1t \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}.$$

What does this mean?

Parametrization $\mathbf{p} + s\mathbf{u} + t\mathbf{v}$. Draw a picture. Define $\mathbf{u} = (-2, 1, 0)$, $\mathbf{v} = (-3, 0, 1)$ and talk about the $\mathbf{u}, \mathbf{v}$ coordinate system.

**The Idea of Dimension**

The dimension of a shape is the minimum number of free parameters needed to describe it.

Examples: Let $\mathbf{p}$ be any point.

- The set $\{\mathbf{p}\}$ is 0-dimensional because it does have any free parameters.
- If $\mathbf{u} \neq \mathbf{0}$ then the line $\{t\mathbf{u} : t \in \mathbb{R}\}$ is 1-dimensional.
- If the vectors $\mathbf{u}, \mathbf{v}$ are non-zero and non-parallel then the plane $\{\mathbf{p} + s\mathbf{u} + t\mathbf{v} : s, t \in \mathbb{R}\}$ is 2-dimensional.
- If the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are non-zero and pairwise non-parallel then the set $\{\mathbf{p}+r\mathbf{u}+s\mathbf{u}+t\mathbf{w} : r, s, t \in \mathbb{R}\}$ **might** be a 3-dimensional space. But suppose that $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Then the equation ...

Do I want to define "linear independence" right now???

**Definition of a $d$-Plane**

Let $\mathbf{p} \in \mathbb{R}^n$ any point and let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_d \in \mathbb{R}^n$ be a set of "linearly independent"[21] vectors. Then the set of points

$$\{\mathbf{p} + t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_d\mathbf{u}_d : t_1, t_2, \ldots, t_d \in \mathbb{R}\} \subseteq \mathbb{R}^n$$

forms a flat $d$-dimensional shape, called a *$d$-plane*. For example, a 1-plane is called a line and a 2-plane is called a plane. An $(n-1)$-plane living in $n$-dimensional space is called a hyperplane.

## 2.5 Systems of Linear Equations

Qualitative pictures for lines and planes. The number of solutions is 0, 1 or $\infty$. For example, it is impossible for a linear system to have exactly two solutions. (You will prove this on a future homework.) More generally:

**The Dimension Principle for Linear Systems**

The solution to a linear system is always a $d$-plane for some $d$. (In order to include the possibility of no solution we sometimes say that the empty solution is a "$(-1)$-plane.") Now suppose that we have a system of $m$ linear equations in $n$ unknowns. If $m \leqslant n$ then the solution is probably an $(n-m)$-plane. If $m < n$ then the solution is probably empty.

[21] This means that no one vector in the set can be expressed as a linear combination of the other vectors. We will see later that this restriction implies $d \leqslant n$.

Much of the theory of linear algebra has to do with quantifying the exceptions to this general principle.[22] If $n = 2$ or $n = 3$ then we can explain everything with pictures. If $n \geqslant 4$ it is impossible to visualize the system. Nevertheless, humans have a complete and satisfactory way to work with linear systems of any dimension. How do they do it?

### 2.6 The Idea of Elimination

You may be familiar with the method of substitution. The method of elimination is more general.

First example:

$$\begin{cases} x + \phantom{2}y + \phantom{3}z = 0, \\ x + 2y + 3z = 0. \end{cases}$$

System of three planes meeting in a line.

System of three hyperplanes in 4D meeting in a 2-plane.

### 2.7 Gaussian Elimination and RREF

Time to get specific. Here is how I would teach the algorithm to my computer.

**AT THIS POINT I GOT TENDONITIS FROM TYPING, SO WHAT FOLLOWS IS ONLY A BARE OUTLINE OF THE COURSE. I WILL POST OLD HAND-WRITTEN NOTES ON THE WEBPAGE.**

## 3 Matrix Arithmetic

### 3.1 Harmless Formalism

We can simplify a linear system so it looks like $A\mathbf{x} = \mathbf{b}$.

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

[22] The deviations are encoded by two numbers, called the *rank* and *nullity* of the system, which are related by a result called the *Fundamental Theorem*. We will discuss this in Chapter 4.

## 3.2 Concept of a Linear Function

What kind of operation takes the (matrix,vector) pair $A, \mathbf{x}$ to the vector $A\mathbf{x}$. Note that it generalizes multiplication of numbers and the dot product of vectors. Notation:

$$\mathbf{x} \bullet \mathbf{y} = \mathbf{x}^T \mathbf{y} \qquad \text{(concept of matrix transpose)}$$

Thus we might think of it as some kind of super general "multiplication." But then what about division? It is tempting to "solve" the linear system $A\mathbf{x} = \mathbf{b}$ by writing $\mathbf{x} = \frac{\mathbf{b}}{A}$. In fact, we will do something like this. The key is to think of a matrix $A$ as a kind of function that sends a vector in $n$-dimensional space ($\mathbf{x} \in \mathbb{R}^n$) to a vector in $m$-dimensional space ($A\mathbf{x} \in \mathbb{R}^m$). But this is not just any function. Matrix functions satisfy the following special properties:

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} \qquad \text{and} \qquad A(\alpha\mathbf{x}) = \alpha(A\mathbf{x}).$$

Jargon: We say that $\mathbf{x} \mapsto A\mathbf{x}$ is a *linear function*. Conversely, let $f : \mathbb{R}^m \to \mathbb{R}^n$ be any function satisfying

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \qquad \text{and} \qquad f(\alpha\mathbf{x}) = \alpha f(\mathbf{x}).$$

Then I claim that we can find an $m \times n$ matrix $A$ such that $f(\mathbf{x}) = A\mathbf{x}$. In other words: Matrix functions and linear functions are the same concept.

Proof: Let $f(\mathbf{e}_j)$ be the $j$th column of $A$. Check that it works. ///

Here is an example of a non-linear function $f(x, y) = (xy, x^2 + y^2)$.

Thinking of matrices as (linear) functions raises some interesting questions.

- Some functions can be composed. How does this apply to matrix functions?
- Some functions can be inverted. How does this apply to matrix functions?

We will deal with these in the following sections.

## 3.3 Matrix Multiplication

Draw a diagram. Compose the arrows. If $A$ is $k \times \ell$ and $B$ is $\ell \times m$ then the composite function $A \circ B$ goes from $\mathbb{R}^m$ to $\mathbb{R}^k$. But observe that the composition of linear functions is linear.

Proof: $(A \circ B)(\mathbf{x} + \alpha\mathbf{y}) = A(B\mathbf{x} + \alpha B\mathbf{y}) = A(B\mathbf{x}) + \alpha A(B\mathbf{y}) = (A \circ B)\mathbf{x} + \alpha(A \circ B)\mathbf{y}$.

It follows that $A \circ B$ is a matrix function. In other words, there exists a $k \times m$ matrix $C$ such that

$$C\mathbf{x} = (A \circ B)\mathbf{x} = A(B\mathbf{x}) \qquad \text{for all } \mathbf{x} \in \mathbb{R}^m.$$

Following the multiplication analogy we will call this matrix $C = AB =$ "$A$ times $B$." How can we compute it?

$$(j\text{th column of } AB) = (AB)\mathbf{e}_j = A(B\mathbf{e}_j) = A(j\text{th column of } B).$$

It follows that

$$(i, j \text{ entry of } AB) = (i\text{th row of } A)(j\text{th column of } B).$$

But we can also think of matrix multiplication in terms of rows. Use the transposition operator:

$$\begin{aligned}(i, j \text{ entry of } (AB)^T) &= (j, i \text{ entry of } AB)\\ &= (j\text{th row of } A)(i\text{th column of } B)\\ &= (i\text{th row of } B^T)(j\text{th column of } A^T)\\ &= (i, j \text{ entry of } B^T A^T).\end{aligned}$$

This implies that $(AB)^T = B^T A^T$, hence

$$(i\text{th row of } AB) = (i\text{th row of } A)B.$$

The concept of matrix multiplication is quite deep because it compactifies many different ideas into the concise notation $AB$. Furthermore, we can work with notation in a rather mindless way because it satisfies many reasonable looking identities:

- $A(BC) = (AB)C$
- $A(B + C) = AB + AC$
- $(A + B)C = AC + BC$

[Note that these imply all our previous identities for vector arithmetic.] The main point of departure from pre-1850 mathematics is the fact that matrix multiplication is not commutative:

$$AB \neq BA.$$

Actually this is a strength because it allows matrices to represent operations such as the composition of space rotations, which is not commutative. Example: Rotate the textbook twice.

## 3.4 Some Examples

Identity, Reflection, Rotation, Projection. Mention determinant of $2 \times 2$ while we're at it.

Which of these functions are invertible? What are the inverses?

## 3.5 Matrix Inversion

Say $m \times n$ matrix $A$ is invertible if there exists $n \times m$ matrix $B$ such that $AB = I_m$ and $BA = I_n$. Draw a diagram.

A projection is not invertible because $P\mathbf{x} = P\mathbf{y}$ for $\mathbf{x} \neq \mathbf{y}$. The same holds for any short and wide matrix. Example: $2 \times 3$ matrix. Intersection of two 2-planes through the origin in 3D contains a line.[23] /// Indeed, this principle tells us that every non-square matrix is non-invertible. How about $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$? We are looking for $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $AB = I$ and $BA = I$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This is 8 linear equations in 4 unknowns. You might guess that this system has no solution, but it turns out that the left four equations are equivalent to the right four equations. (This is mysterious; see the next section.) Solve the left 4 equations by two rounds of RREF. Consolidate them into one round of $RREF$:

$$(A|I) \to (I|B).$$

If it works then $AB = I$. One can then check that $BA = I$ (miracle!). Example where is doesn't work :$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. Why doesn't it work? Column relation $A \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \mathbf{0}$ shows there is no left inverse.

Relation to the determinant:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} -a & c \\ b & -d \end{pmatrix}.$$

What happens in higher dimensions? There does exist such a thing as the determinant of an $n \times n$ matrix but I see no need to define it today.

### 3.6 Secret: The Fundamental Theorem

We saw that $A\mathbf{x} = A\mathbf{y}$ with $\mathbf{x} \neq \mathbf{y}$ implies $A$ has no left inverse. More concisely: If $\mathbf{x} \neq \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ then $A$ has no left inverse. Let us formalize this with the concept of a nullspace.

Nullspace / Kernel / Space of column relations

Nullity = dimension of nullspace = number of non-pivot columns in the RREF. What about right inverses?

Column Space / Image

Rank = dimensiona of column space. Note that rank is the number of pivot columns in the RREF. Thus we have rank(A)=# cols(A) if and only if RREF(A)=I. In this case, the equation $AC = I$ has a unique solution $C$ because each system $A\mathbf{c}_j = \mathbf{e}_j$ has a unique solution. We conclude that:

- $A$ has a left inverse if and only if nullity of $A$ is zero.

[23] What about two 2-planes in 4D?

- $A$ has a right inverse if and only if rank equals the number of columns.

The Fundamental Theorem: rank + nullity = number of columns.

Proof: nullity = number of non-pivot columns.

Conclusion: Consider square matrices $A, B$.

$$\begin{aligned} AB = I &\Rightarrow A \text{ has right inverse} \\ &\Rightarrow A \text{ has rank} = \# \text{ columns} \\ &\Rightarrow A \text{ has nullity} = 0 && \text{[Fundamental Theorem]} \\ &\Rightarrow A \text{ left inverse} \\ &\Rightarrow CA = I \text{ for some matrix } C. \end{aligned}$$

Finally,

$$C = CI = C(AB) = (CA)B = IB = B.$$

We conclude that $BA = CA = I$. This is the simplest proof that I know that $AB = I$ implies $BA = I$ for square matrices. Most textbooks give the false impression that this is a straightforward result. It is not.

Obligatory list of equivalent conditions for invertibility. (Maybe not.)

# 4 Least Squares Regression

## 4.1 Projection onto a line or plane

## 4.2 Projection onto a subspace

## 4.3 The normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$

# 5 Spectral Analysis