Out[1]: (Show / Hide code)

Mathematics Introduction

CSE 4303 / CSE 5365 Computer Graphics

2020 Fall Semester, Version 1.3, 2020 September 15

A certain amount of linear algebra is required for the study of Computer Graphics. As one of **CSE 3380** — *Linear Algebra for CSE* or **MATH 3330** — *Introduction to Matrices and Linear Algebra* is a prerequisite for CSE 4303 / CSE 5365, everything in this introduction should be familiar.

This introduction is *not* a substitute for success in the prerequisite *Linear Algebra* course; it's merely a reminder for what you should have already learned. If anything is unclear in this introduction, review the corresponding material in your prerequisite course's text.

Euclidean Space

$\mathbb{R}^n$ represents the real n-dimensional Euclidean space. Each element of this space is a real n-tuple, or ordered list of n real numbers. We will index the components of the element starting at 0 since that is the more common programming convention. (Be aware that many references — including *Introduction to Computer Graphics* by Foley, van Dam, Feiner, Hughes, and Phillips — follow the mathematical convention of starting indices at 1.)

> $e \in \mathbb{R}^n$ is $\begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{n-1} \end{bmatrix}$ where $e_i \in \mathbb{R},\ 0 \le i < n.$

Many references call any such e a *vector*, but we will be more selective in this class and distinguish *points* from *vectors*.

Note that the n-tuple is presented as a column. This is important as columns and rows are different. If we have to write a point or vector horizontally, we will use the *transpose* operator T to indicate that the row has to be 'stood up' into a column.

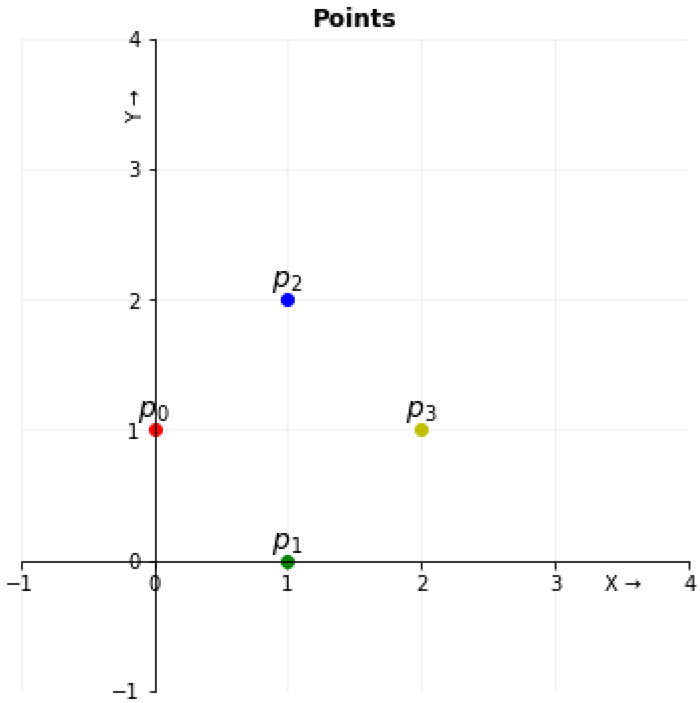
As an example, an element $e \in \mathbb{R}^2 = \begin{bmatrix} e_0, e_1 \end{bmatrix}^T = \begin{bmatrix} e_0 \\ e_1 \end{bmatrix}$

Points

A point $\in \mathbb{R}^n$ is an n-tuple representing a *location*. Points have no other attributes; no length, width, depth, etc.

For example, in $\mathbb{R}^2$ we could have,

$$\mathbf{p}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{p}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{p}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



In the figure, the points are shown as small circles to make them visible. Actual points have no size (or color, for that matter).

Vectors

A vector $\in \mathbb{R}^n$ is an n-tuple representing a *direction* and a *magnitude*. Unlike a point, a vector does *not* have a location. When we draw a vector, we start from some location and move in the vector's direction for its magnitude. However, as far as the vector itself is concerned, that start location is completely arbitrary.

Operations

The two simplest operations for vectors are *multiplication by a scalar* and *addition of two vectors*. Each of these operations has as its result another vector in $\mathbb{R}^n$ (indicated below by the $\in \mathbb{R}^n$ clarification).

Multiplication by a Scalar

Multiplication of a vector $\mathbf{v}$ by a scalar a is defined as

$$a\,\mathbf{v} = \begin{bmatrix} a\,v_0 \\ a\,v_1 \\ \vdots \\ a\,v_{n-1} \end{bmatrix} \in \mathbb{R}^n \quad \forall \mathbf{v} \in \mathbb{R}^n, a \in \mathbb{R}$$

It's pretty simple: just take the scalar and multiply each element of the vector by that value. We can compute *Division by a Scalar* just by multipling by the reciprocal of the scalar, as long as the scalar $\neq 0$.

Scalar muliplication has some interesting properties. When $a > 0$,

- The resulting vector has its *magnitude* scaled by the factor a. The *orientation* and *direction* of the resulting vector are the same as that of the original vector.
- ***Identity***: When $a = 1$ the resulting vector is the same as the original vector. 1 is therefore the *identity* element for scalar multiplication. $1 \cdot \mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^n$.

The scalar a does not have to be greater than 0. Scalar multiplication is well-defined for $a \leq 0$.

- ***Zero vector***: When $a = 0$, the result is the n-component *Zero* or *Null* vector $= \left[0, 0, \ldots, 0\right]^T$, usually written $\mathbf{0}$.
- ***Reversal***: More interesting is when $a < 0$. In this case, the magnitude of the resulting vector is scaled instead by the factor $|a|$. The *orientation* of the vector does not change, but its *direction* is reversed.
- ***Additive inverse***: When $a = -1$, one writes just $-\mathbf{v}$, indicating $\mathbf{v}$'s additive inverse vector (which becomes more meaningful when we consider vector addition below).

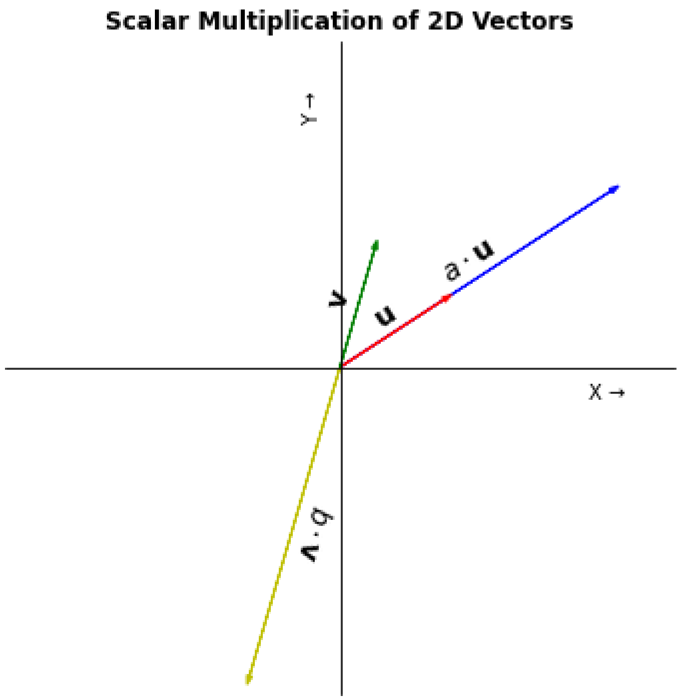
Finally, for any $a, b \in \mathbb{R}$ and any $\mathbf{v} \in \mathbb{R}^n$,

- $(ab)\mathbf{v} = a(b\,\mathbf{v})$

Example

With $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $a = 2.5$, the scalar multiplication $a\,\mathbf{u} = 2.5 \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.5 \cdot 3 \\ 2.5 \cdot 2 \end{bmatrix} = \begin{bmatrix} 7.5 \\ 5 \end{bmatrix}$.

With $\mathbf{v} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}$ and $b = -2.5$, the scalar multiplication $b\,\mathbf{v} = -2.5 \begin{bmatrix} 1 \\ 3.5 \end{bmatrix} = \begin{bmatrix} -2.5 \cdot 1 \\ -2.5 \cdot 3.5 \end{bmatrix} = \begin{bmatrix} -2.5 \\ -8.75 \end{bmatrix}$.



Addition of Two Vectors

The addition of two vectors is defined as

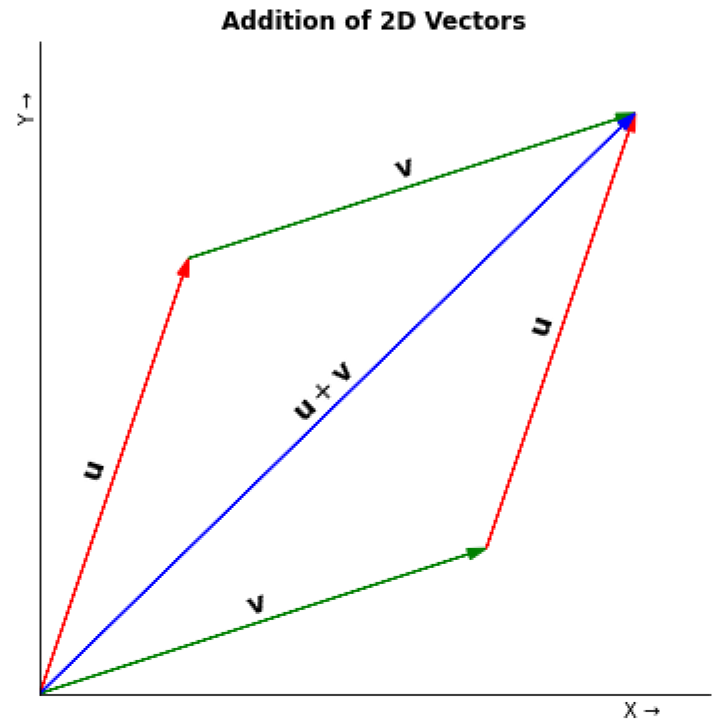
$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_0 + v_0 \\ u_1 + v_1 \\ \vdots \\ u_{n-1} + v_{n-1} \end{bmatrix} \in \mathbb{R}^n \text{ for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

This is pretty simple: just add the corresponding elements of the two vectors.

Vector addition is defined *only* if the two vectors have the same number of components.

Example

With $\mathbf{u} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$, the vector sum $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 2+6 \\ 6+2 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$.



In the *Scalar Multiplication of 2D Vectors* example, all of the vectors were shown with their start point at the graph's origin $(0, 0)$. This was purely for convenience as vectors in and of themselves have no particular origin, being only a direction and a magnitude. This characteristic was clearly shown in the *Addition of 2D Vectors* example as both \Vector$u$ and \Vector$v$ are shown twice, with different origins. Each time, however, the vectors have the same direction (are *parallel*) and magnitudes.

Vector addition has some interesting properties.

- ***Additive Identity***: $\mathbf{0}$ is the vector addition *identity* element, as $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ for any vector $\mathbf{u}$
- ***Commutativity***: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. This is clearly shown in the *Addition of 2D Vectors* example.
- ***Associativity***: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- ***Additive inverse***: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$, which can also be written $\mathbf{u} - \mathbf{u} = \mathbf{0}$

Using the definition of a vector's additive inverse, we can define vector *subtraction* $\mathbf{u} - \mathbf{v} \equiv \mathbf{u} + (-\mathbf{v})$.

However, be careful as vector subtraction is *neither* commutative *nor* associative. That is, in general $\mathbf{u} - \mathbf{v} \neq \mathbf{v} - \mathbf{u}$ and $(\mathbf{u} - \mathbf{v}) - \mathbf{w} \neq \mathbf{u} - (\mathbf{v} - \mathbf{w})$.

This *lack* of properties of vector subtraction is easy to see using its definition. We defined $\mathbf{u} - \mathbf{v} \equiv \mathbf{u} + (-\mathbf{v})$. Therefore, $\mathbf{v} - \mathbf{u}$ must be $\mathbf{v} + (-\mathbf{u})$. So, $\mathbf{u} - \mathbf{v} = \mathbf{v} - \mathbf{u}$ would require $\mathbf{u} + (-\mathbf{v}) = \mathbf{v} + (-\mathbf{u})$, which is clearly not so in the general case.

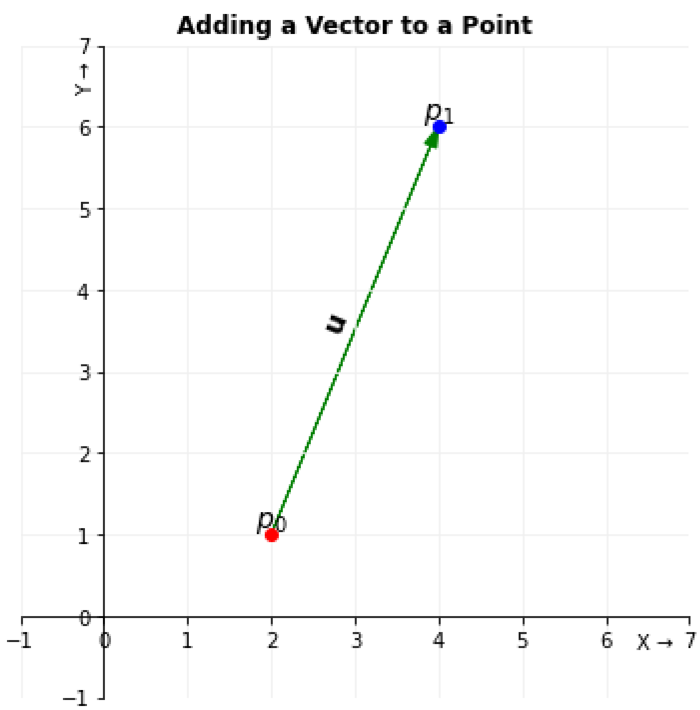
Vectors and Points

Aside from adding two vectors, it is possible to add a vector to a point. In this case, we get another point. The interpretation is that we started at the first point, went some distance in a direction, and ended up at another point.

Thus adding vector $\mathbf{u} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ to point $\mathbf{p}_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ gives point $\mathbf{p}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$.

Conversely, subtracting a point from another point yields a vector.

Thus $\mathbf{p}_1 - \mathbf{p}_0 = \mathbf{u}$, $\begin{bmatrix} 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$. To get from $\mathbf{p}_0$ to $\mathbf{p}_1$, we need to travel according to the direction and magnitude of vector $\mathbf{u}$.



While it's possible to add a vector to a point, it is *not* possible to add a point to a vector — that operation is undefined. Also, it is not possible to add two points.

Combining Operations

Combining scalar multiplication of a vector with vector addition yields some useful and interesting results.

- ***Distribution***: $(a + b)\,\mathbf{u} = a\,\mathbf{u} + b\,\mathbf{u}$
- ***Distribution***: $a\,(\mathbf{u} + \mathbf{v}) = a\,\mathbf{u} + a\,\mathbf{v}$

These results are easy to see from the definitions of the operations. Vector addition is the pairwise adding of the vector components. Multiplication of a vector by a scalar is the multiplication of each vector component by that scalar. Each of these operates at the component level where everything is a scalar.

Therefore, $(a + b)\,\mathbf{u}$ can be thought of as a collection of $(a + b)\,u_i$ operations, where a, b, and u_i are all scalars. At this level, the rules of arithmetic take over and we easily see that $(a + b)\,u_i = au_i + bu_i$. Similarly, $a\,(\mathbf{u} + \mathbf{v})$ can be thought of as $a\,(u_i + v_i) = au_i + bv_i$.

Linear Combination

Going a bit further with scalar multiplication of a vector and vector addition, we can define a *linear combination* $\mathbf{c}$ of an ordered set of m scalar factors $a_i \in \mathbb{R}$ and an ordered set of m vectors $\mathbf{u}_i$ each $\in \mathbb{R}^n$ as

$$\mathbf{c} = \sum_{i=0}^{m-1} a_i \mathbf{u}_i = a_0 \mathbf{u}_0 + a_1 \mathbf{u}_1 + \ldots a_{m-1} \mathbf{u}_{m-1} \text{, where } \mathbf{c} \in \mathbb{R}^n$$

Linear combinations of vectors are used to describe many objects and are especially useful in the representation of curves and surfaces, as we shall see later.

Demonstration

In the first following demonstration, we have $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

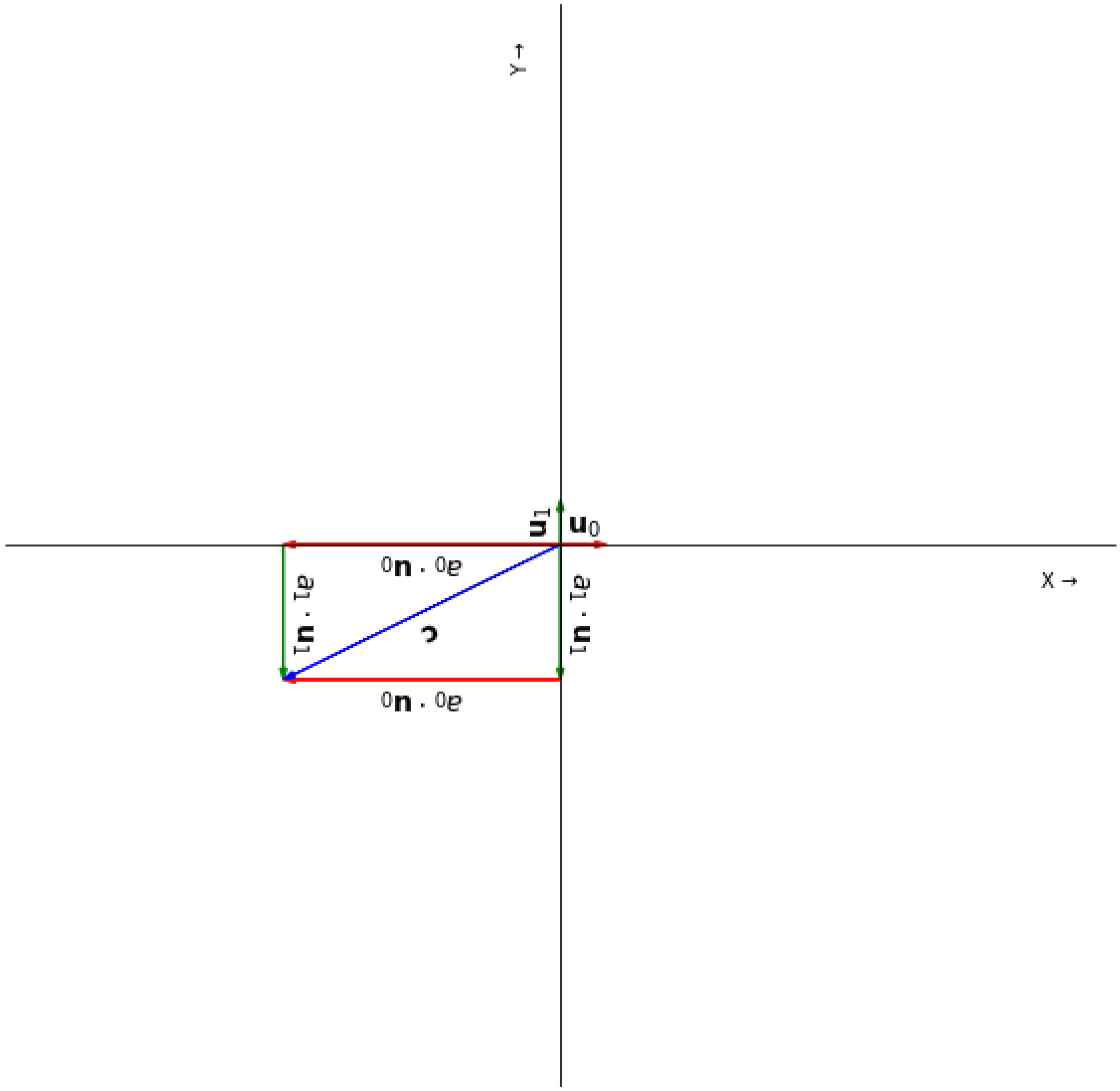
a_0 and a_1 are each adjustable in the range $[-10 \ldots 10]$.

a0 -6.0

a1 -3.0

- [x] Show u0, u1

Linear Combination of 2D Vectors



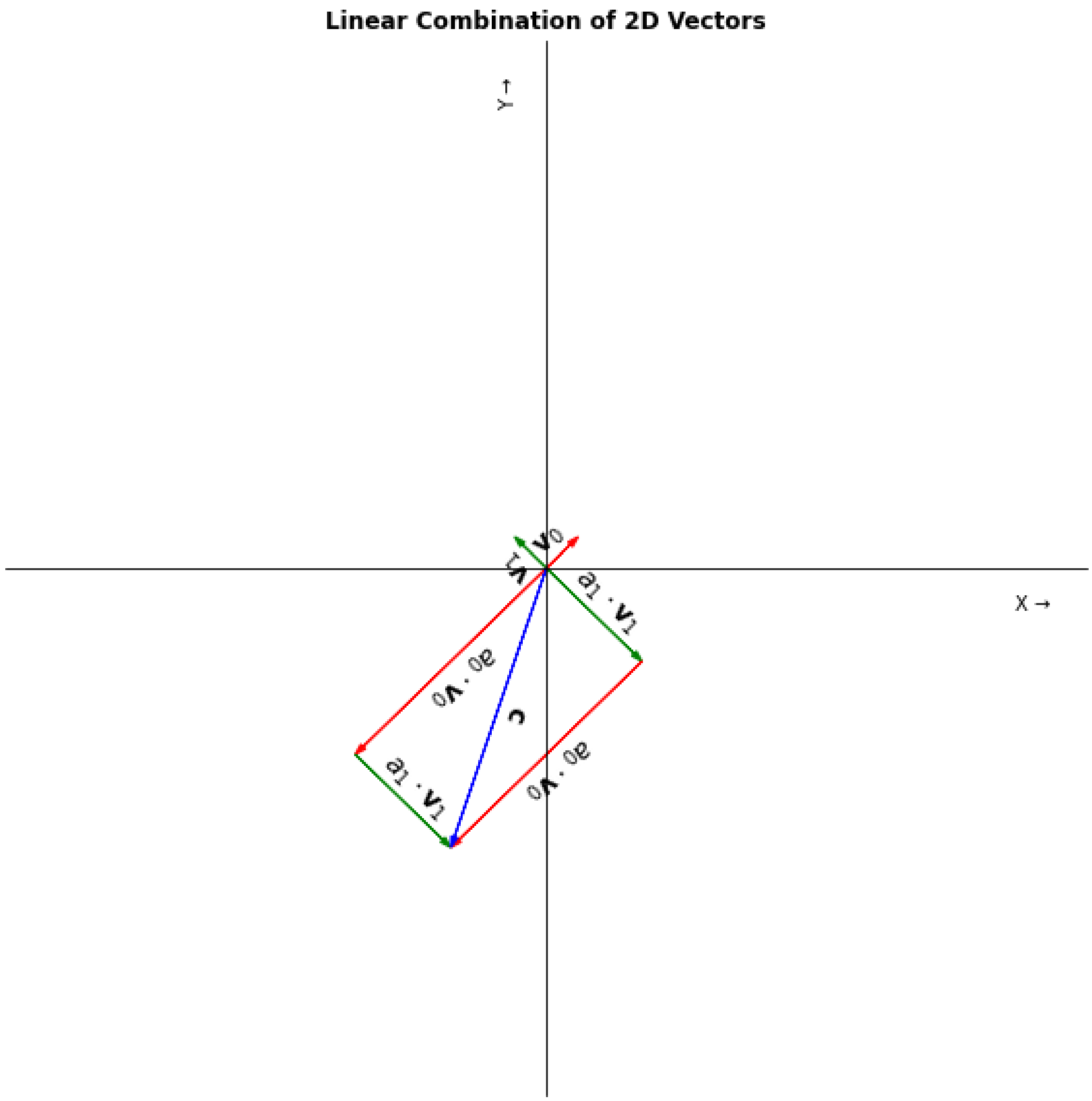
In the second demonstration, let's make it a bit more interesting by rotating $\mathbf{u}_0$ and $\mathbf{u}_1$ by $\frac{\pi}{4}, 45°$ counterclockwise so that they are no longer aligned with the axes. We'll call these two new vectors $\mathbf{v}_0$ and $\mathbf{v}_1$ respectively. We therefore now have $\mathbf{v}_0 = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$ and $\mathbf{v}_1 = \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$.

a_0 and a_1 are still each adjustable in the range $[-10 \ldots 10]$.

a0 -6.0

a1 -3.0

☑ Show v0, v1



In either case, we can direct $\mathbf{c}$ in any direction and magnitude (within the limits of the display), but the required scalar coefficients a_0, a_1 depend on the particular values of the base vectors $\mathbf{u}_0, \mathbf{u}_1$ and $\mathbf{v}_0, \mathbf{v}_1$. (Not surprising.)

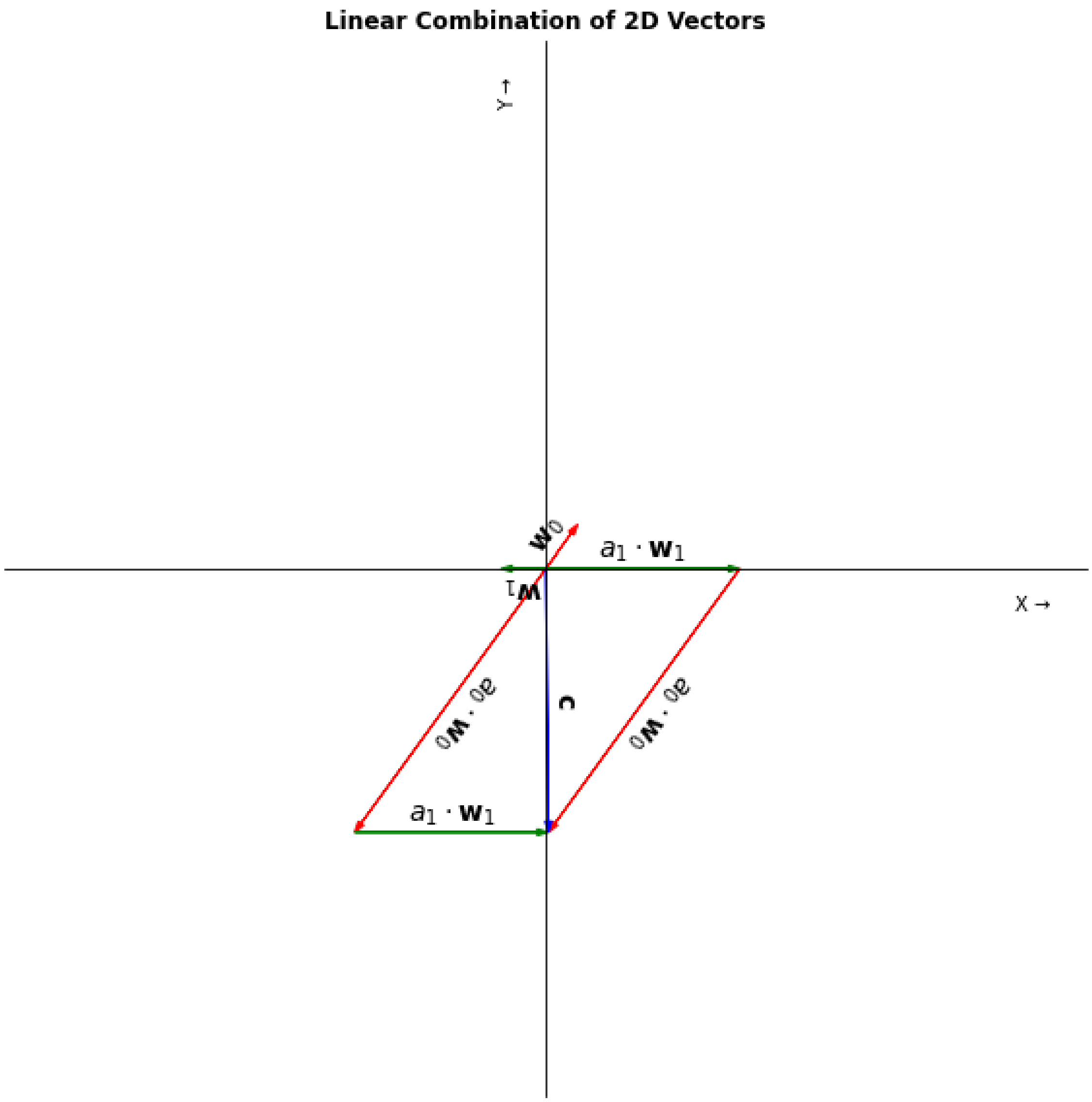
Interestingly enough, the base vectors do *not* have to be orthogonal for this to work; they just have to be non-parallel. As long as that's true, we can direct $\mathbf{c}$ in any direction and magnitude (within the limits of the display). As before, the required scalar coefficients a_0, a_1 depend on the particular values of the base vectors. Let's try this with $\mathbf{w}_0 = \left[\frac{\sqrt{2}}{2}, 1\right]^T$ and $\mathbf{w}_1 = \left[-1, 0\right]^T$. These two vectors are clearly non-parallel, but we have no difficulty in getting $\mathbf{c}$ to point in any direction we want.

Here, we have $\mathbf{c}$ pointing straight down the $-y$ axis.

a0 -6.0

a1 -4.3

Show w0, w1



“Reverse” Linear Combination

So far we have applied a given a_0 and a_1 and computed the linear combination $\mathbf{c}$ from the given basis vectors. Suppose, however, we have a particular $\mathbf{c}$ and want to compute the required a_0 and a_1 for given basis vectors? Is this possible?

Mais oui!

Let's say we have a particular vector in mind and we want to know the required values of a_0 and a_1 such that $a_0\mathbf{w}_0 + a_1\mathbf{w}_1$ gives us that vector.

Well, vectors have a direction θ and a magnitude r. We first convert (r, θ) to (x, y) using basic trigonometry,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

and realize we need to solve this equation set,

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= a_0\mathbf{w}_0 + a_1\mathbf{w}_1 \\ &= a_0 \begin{bmatrix} w_{0x} \\ w_{0y} \end{bmatrix} + a_1 \begin{bmatrix} w_{1x} \\ w_{1y} \end{bmatrix} \\ &= \begin{bmatrix} a_0 w_{0x} \\ a_0 w_{0y} \end{bmatrix} + \begin{bmatrix} a_1 w_{1x} \\ a_1 w_{1y} \end{bmatrix} \\ &= \begin{bmatrix} a_0 w_{0x} + a_1 w_{1x} \\ a_0 w_{0y} + a_1 w_{1y} \end{bmatrix} \end{aligned}$$

which is two equations in the two unknowns a_0, and a_1. The values $x, y, w_{0x}, w_{0y}, w_{1x}$, and w_{1y} are all known. We can solve this system using standard methods. (Everyone remembers how to solve the n equations with n unknowns problem using *Linear Algebra*, right? [*Sigh*. Multiply the inverse of the "equation" matrix by the x, y vector.]

Let's demonstrate this with $\mathbf{w}_0 = \left[\frac{\sqrt{2}}{2}, 1\right]^T$ and $\mathbf{w}_1 = [-1, 0]^T$ as the basis vectors.

> By the way, we're using w_0 and w_1 as our basis vectors because the process is absolutely trivial (by inspection) for $\mathbf{u}_0 = [1, 0]^T$ and $\mathbf{u}_1 = [0, 1]^T$. The process isn't much harder when the basis vectors are $\mathbf{v}_0 = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$ and $\mathbf{v}_1 = \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$) either.
>
> Do you see why this is so?

r 6.0

theta -90.0

☑ Show w0, w1

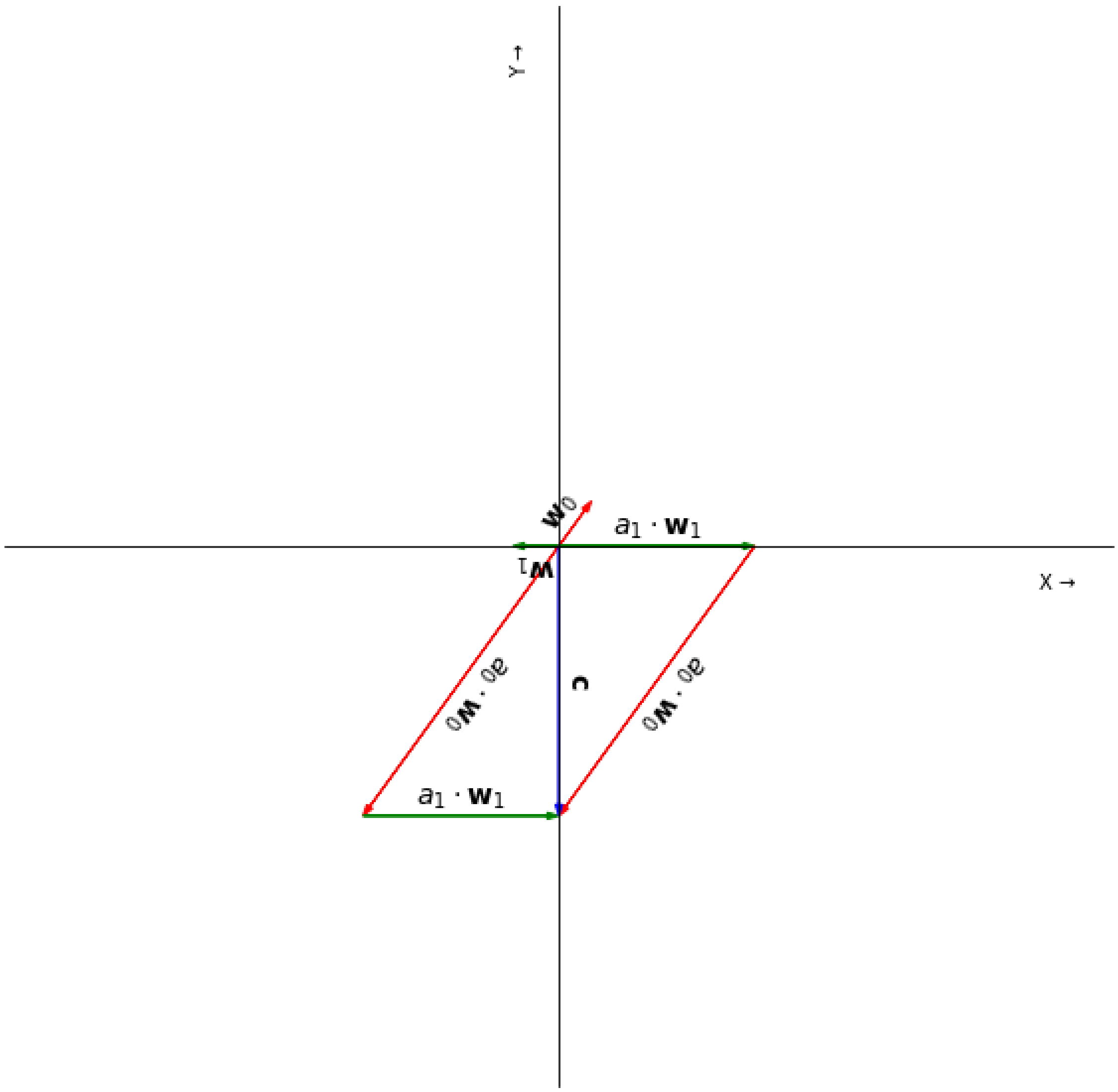
```
w0 ( 0.707, 1.000 ), w1 ( -1.000, 0.000 )
r 6.0, theta -90.0 degrees -> x 0.000, y -6.000

Eqn matrix: [ [ 0.707, -1.000 ], [ 1.000, 0.000 ] ]
Inverse: [ [ 0.000, 1.000 ], [ -1.000, 0.707 ] ]

Result: a0 =  0.000 ·  0.000 +  1.000 · -6.000 = -6.000
Result: a1 = -1.000 ·  0.000 +  0.707 · -6.000 = -4.243

Check: a0w0 ( -4.243, -6.000 ), a1w1 ( 4.243, -0.000 )
Check: a0w0 + a1w1 ( 0.000, -6.000 )
```

"Reverse" Linear Combination of 2D Vectors



Dot Product

The next operation to consider is the *dot product* of two vectors, denoted for two vectors $\mathbf{u}, \mathbf{v}$ as $\mathbf{u} \cdot \mathbf{v}$.

> *Be very careful with your notation!* There are four kinds of common operations applicable to vectors that could be considered "multiplication", denoted by $\cdot$ (Dot Product), $\times$ (Cross Product), $\otimes$ (Outer Product), and $*$ (Convolution).
>
> Granted, it's not likely you would write $\otimes$ by accident, but it's very easy to casually write one of $\cdot$, $\times$, or $*$ when you really mean one of the others. After all, they all mean the same thing — arithmetic multiplication — when applied to scalars.

Dot product is defined as

$$\mathbf{u} \cdot \mathbf{v} = d = \sum_{i=0}^{n-1} u_i v_i \text{, for } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, d \in \mathbb{R}$$

In words, the dot product is the sum of multiplying the corresponding components of the two vectors. Therefore, as with vector addition, the dot product is defined only if the two vectors have the same number of components. However, unlike vector addition, the result of the dot product operation is a *scalar* not a *vector*.

The dot product has some useful and interesting properties.

- $\mathbf{u} \cdot \mathbf{u} \geq 0$, with $\mathbf{u} \cdot \mathbf{u} = 0 \iff \mathbf{u} = \mathbf{0}$
- ***Commutivity***: $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- ***Distribution***: $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
- $(a\,\mathbf{u}) \cdot (b\,\mathbf{v}) = ab(\mathbf{u} \cdot \mathbf{v})$

And finally an incredibly useful property,

- $\mathbf{u} \cdot \mathbf{v} = 0 \iff \mathbf{u} \perp \mathbf{v}$, for $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$

In words, the dot product of two non-zero vectors is equal to zero if and only if the two vectors are orthogonal, that is, perpendicular to each other, $\perp$. (Technically, the vectors do not have to be non-zero as mathematicians consider the zero vector orthogonal to all vectors, including itself.)

(This property isn't difficult to show using basic trignometry. Think about it. :)

Above we had two vectors $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Are they perpendicular? Let's see!

$$\begin{aligned} \mathbf{u}_0 \cdot \mathbf{u}_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 1 \cdot 0 + 0 \cdot 1 \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

That could hardly have been easier. An useful observation here is that if one has vectors that are aligned with different coordinate axes, they will *always* be perpendicular to each other simply because one or the other of the corresponding components will always be zero. (E.g., if a vector is aligned with the x axis, its y component will be zero. If it's aligned with the y axis, its x component will be zero. And so forth — this argument works for any number of dimensions.)

What about the second linear combination example? Well, since the only change was to rotate $\mathbf{u}_0$ and $\mathbf{u}_1$ by $\frac{\pi}{4}, 45^\circ$ counterclockwise to get $\mathbf{v}_0$ and $\mathbf{v}_1$, those two vectors should still be perpendicular. We have $\mathbf{v}_0 = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right]^T$ and $\mathbf{v}_1 = \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right]^T$. Are these two vectors perpendicular? Let's do the math!

$$\begin{aligned} \mathbf{v}_0 \cdot \mathbf{v}_1 &= \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \cdot \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \\ &= \frac{\sqrt{2}}{2} \cdot \left(-\frac{\sqrt{2}}{2} \right) + \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2}}{2} \\ &= \frac{\sqrt{2} \cdot (-\sqrt{2})}{2 \cdot 2} + \frac{\sqrt{2} \cdot \sqrt{2}}{2 \cdot 2} \\ &= -\frac{2}{4} + \frac{2}{4} \\ &= 0 \end{aligned}$$

Ta-da!

That was easy, wasn't it?

How about the third linear combination example? We specifically made the basis vectors $\mathbf{w}_0$ and $\mathbf{w}_1$ non-orthogonal in that case. Let's see what our dot product check reveals.

$$
\begin{aligned}
\mathbf{w}_0 \cdot \mathbf{w}_1 &= \begin{bmatrix} \frac{\sqrt{2}}{2} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix} \\
&= \frac{\sqrt{2}}{2} \cdot (-1) + 1 \cdot 0 \\
&= -\frac{\sqrt{2}}{2} + 0 \\
&= -\frac{\sqrt{2}}{2}
\end{aligned}
$$

OK, clearly $-\frac{\sqrt{2}}{2} \neq 0$ so $\mathbf{w}_0$ and $\mathbf{w}_1$ are not orthogonal according to our check, which is exactly what we expected.

Et voila, ça marche!

Norm

In the 2D plane defined by $\mathbb{R}^2$, the distance from any point (x, y) to the origin is $\sqrt{x^2 + y^2}$. This can be extended to n dimensions so that we have the distance of point $\mathbf{p} \in \mathbb{R}^n$ to the origin $= \sqrt{\sum_{i=0}^{n-1} p_i^2}$.

Given the definition of the dot product, it's clear that the magnitude of a vector $\mathbf{u}$ is given by $\sqrt{\mathbf{u} \cdot \mathbf{u}}$. This is known as the *norm* of the vector and is written as $||\mathbf{u}||$.

Some properties of the norm include,

- $||\mathbf{u}|| = 0 \iff \mathbf{u} = \mathbf{0}$
- $||a\mathbf{u}|| = |a| \; ||\mathbf{u}||$
- ***Triangle inequality***
  $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
  $||\mathbf{u} + \mathbf{v}|| = ||\mathbf{u}|| + ||\mathbf{v}|| \iff \mathbf{u}$ and $\mathbf{v}$ have the same direction for $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$
- ***Cauchy-Schwarz inequality***
  $|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| \; ||\mathbf{v}||$
  $|\mathbf{u} \cdot \mathbf{v}| = ||\mathbf{u}|| \; ||\mathbf{v}|| \iff \mathbf{u} || \mathbf{v}$ for $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$

For the Triangle inequality, the two sides are equal if and only if the (non-zero) vectors $\mathbf{u}, \mathbf{v}$ have the same direction. For the Cauchy-Schwarz inequality, the two sides are equal if and only if the (non-zero) vectors $\mathbf{u}, \mathbf{v}$ are parallel, $||$ (have the same *orientation*). (Technically, the vectors do not have to be non-zero as mathematicians consider the zero vector parallel to all vectors, including itself.)

(These properties aren't difficult to show using basic trigonometry. Think about it. :)

Normalization

To *normalize* a non-zero vector means to scale its magnitude to 1. We compute the normalized version of a vector thusly,

$$\hat{\mathbf{u}} = \frac{1}{||\mathbf{u}||}\mathbf{u} = \frac{\mathbf{u}}{||\mathbf{u}||}, \text{ for } \mathbf{u} \neq \mathbf{0}$$

Any vector with a magnitude of 1 is known as a *unit vector*. The $\hat{\mathbf{u}}$ notation means “the unit vector in the direction of $\mathbf{u}$”.

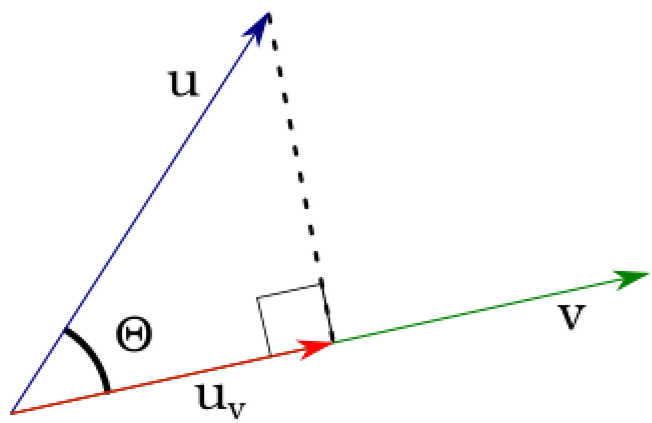
Geometrical Interpretation of the Dot Product

As observed above, the dot product $\mathbf{u} \cdot \mathbf{v}$ is equal to 0 when the vectors $\mathbf{u}, \mathbf{v}$ are perpendicular to each other and is equal to $||\mathbf{u}||\ ||\mathbf{v}||$ when the vectors are parallel.

If we take the dot product of the normalized versions of two vectors, we get the cosine of the angle between them.

$$\cos\theta_{uv} = \frac{\mathbf{u}}{||\mathbf{u}||} \cdot \frac{\mathbf{v}}{||\mathbf{v}||} = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||\ ||\mathbf{v}||}$$

Wow. Think about that for a moment. We can compute the trigonometric function $\cos$ just by doing some multiplications, additions, and a couple of square roots.



We can use this property of the dot product to find the *projection* of one of the vectors onto the other. For example, to find the projection of $\mathbf{u}$ onto $\mathbf{v}$ we first compute the magnitude of the projection

$$\begin{aligned} u_v &= ||\mathbf{u}|| \cos\theta_{uv} \\ &= ||\mathbf{u}|| \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||\ ||\mathbf{v}||} \\ &= \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||} \\ &= \mathbf{u} \cdot \hat{\mathbf{v}} \end{aligned}$$

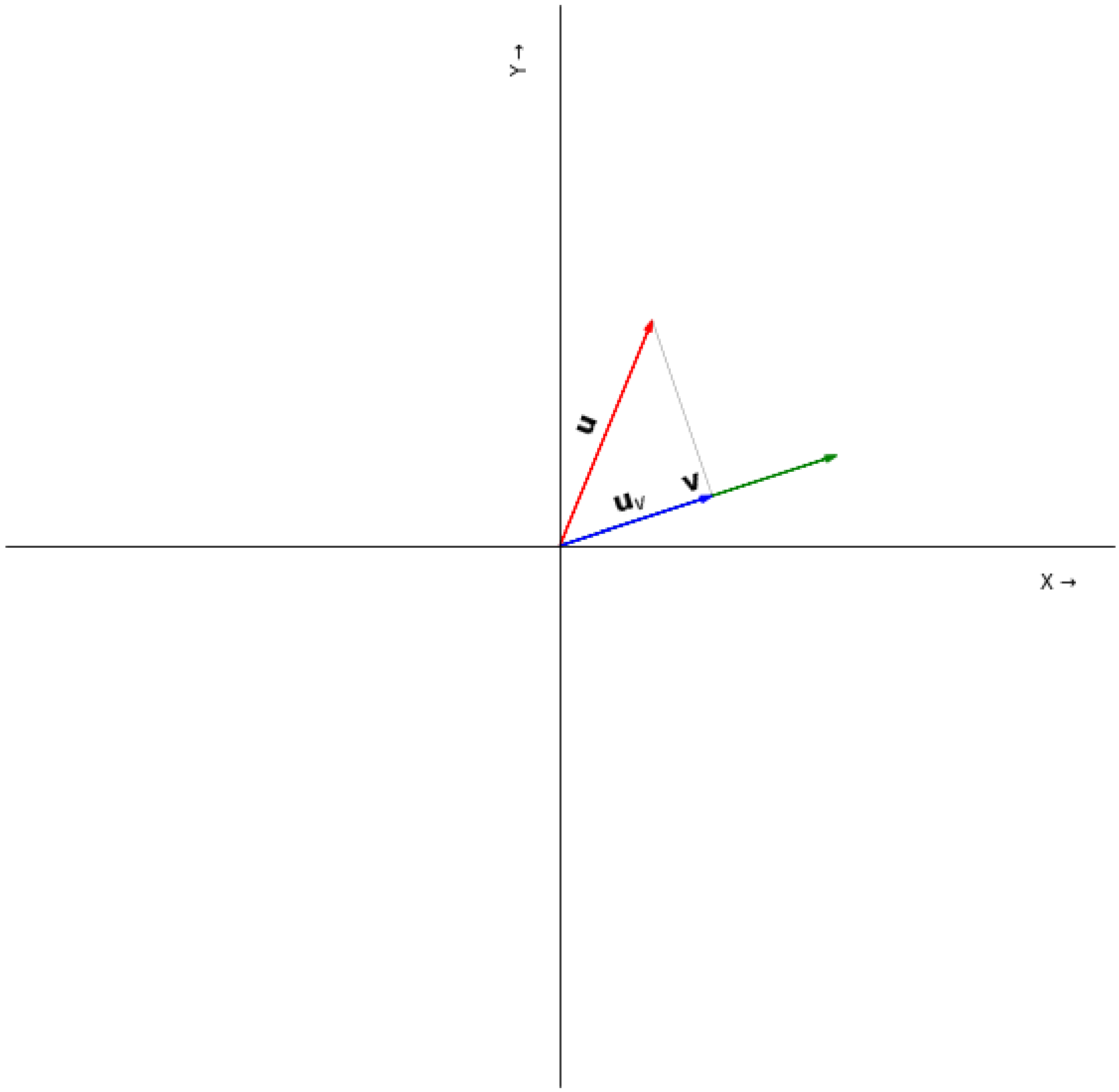
We can then form the projected vector $\mathbf{u_v} = u_v \hat{\mathbf{v}}$.

Similarly, the projection of $\mathbf{v}$ onto $\mathbf{u}$ is $\mathbf{v_u} = (\mathbf{v} \cdot \hat{\mathbf{u}})\,\hat{\mathbf{u}}$.

u_x	2.0	u_y	5.0
v_x	6.0	v_y	2.0

```
u dot v 22.000, norm u 5.385, norm v 6.325
cos theta 0.646, theta 49.764 degrees
norm uv 3.479, hat v ( 0.949, 0.316 )
uv ( 3.300, 1.100 )
```

Dot Product Projection of 2D Vectors



Matrices

For us a *matrix* is a two-dimensional arrangement of numbers organized in rows and columns. A matrix $\mathbf{M}$ with p rows and q columns is called a $p \times q$ matrix and appears thus,

$$\mathbf{M} = \begin{bmatrix} m_{0,0} & m_{0,1} & \cdots & m_{0,q-1} \\ m_{1,0} & m_{1,1} & \cdots & m_{1,q-1} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p-1,0} & m_{p-1,1} & \cdots & m_{p-1,q-1} \end{bmatrix} \in \mathbb{R}^{p \times q}$$

As with vectors, we start the indices at 0.

A vector $\vec{\mathbf{u}} \in \mathbb{R}^n$ is an $n \times 1$ matrix. Mathematically, it is called a *column vector*, though we use the term vector in a more restricted sense. Similarly, an $1 \times n$ matrix is called a *row vector*.

If the number of rows of a matrix is equal to its number of columns, the matrix is known as a *square* matrix. Many matrix operations and processes are applicable only to square matrices. We will note this requirement as necessary.

As might be expected, a *zero* or *null* matrix is a matrix whose entries are all zero. We represent this matrix as $\mathbf{0}$, when its dimensions are clear from the context, or $\mathbf{0}_n$ for the square zero matrix, or $\mathbf{0}_{p,q}$ when not square.

The *identity* matrix $\mathbf{I}_n$ is the $n \times n$ (and therefore *square*) matrix whose entries are

$$m_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \text{ for } 0 \le i, j < n.$$

For example,

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \ldots$$

Operations

There are a number of elementary operations on matrices, some analgous to vector operations and some a bit more complex.

Transpose

The *transpose* of the matrix $\mathbf{M} \in \mathbb{R}^{p \times q}$ is the $q \times p$ matrix (notice the swapping of p and q) built by taking the rows of $\mathbf{M}$ as its columns, or equivalently the columns of $\mathbf{M}$ as its rows. As with vectors, we use T as the transpose operator and denote the transpose of $\mathbf{M}$ as $\mathbf{M}^{\mathbf{T}}$. More precisely,

$$m_{ij}^{T} = m_{ji}, \text{ where } 0 \le i < q,\ 0 \le j < p$$

For example,

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}^T = \begin{bmatrix} 0 & 2 & 4 & 6 \\ 1 & 3 & 5 & 7 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}^T = \begin{bmatrix} 0 & 3 & 6 \\ 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix}$$

Notice that the matrix in the second example above is a square matrix. When a square matrix is transposed, the elements on the diagonal do not change position. Also, the elements in the upper left and lower right “triangles” are mirrored about the diagonal.

Properties of the transpose operation include

- $(a\,\mathbf{M})^T = a\,\mathbf{M}^T$
- $(\mathbf{M} + \mathbf{N})^T = \mathbf{M}^T + \mathbf{N}^T$
- $(\mathbf{M}^T)^T = \mathbf{M}$
- $(\mathbf{M}\mathbf{N})^T = \mathbf{N}^T\mathbf{M}^T$

A matrix is called *symmetric* if it is equal to its transpose, $\mathbf{M} = \mathbf{M}^T$. All symmetric matrices are square.

Trace

The *trace* of a matrix is the sum of its diagonal elements. The trace is defined only for square matrices. It is denoted $\mathrm{tr}(\mathbf{M})$ and is defined as

$$\mathrm{tr}(\mathbf{M}) = \sum_{i=0}^{n-1} m_{ii} \text{ for } \mathbf{M} \in \mathfrak{R}^{n\times n}$$

For example,

$$\mathrm{tr}\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix} = 12$$

Properties of the trace operation include

- $\mathrm{tr}(\mathbf{M} + \mathbf{N}) = \mathrm{tr}(\mathbf{M}) + \mathrm{tr}(\mathbf{N})$
- $\mathrm{tr}(a\mathbf{M}) = a\,\mathrm{tr}(\mathbf{M})$
- $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}\left(\mathbf{M}^T\right)$

The trace of the identity matrix $\mathbf{I}_n$ is n since its diagonal has n elements all of which are 1.

Multiplication by a Scalar

The multiplication of a matrix $\mathbf{B} \in \mathbb{R}^{p\times q}$ by a scalar a is denoted $\mathbf{M} = a\,\mathbf{B}$ and is defined as

$$m_{ij} = ab_{ij}, \text{ where } 0 \le i < p,\ 0 \le j < q$$

As in the vector case, each element of the matrix is multiplied by the scalar.

For example,

$$10 \begin{bmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 0 & 10 \\ 20 & 30 \\ 40 & 50 \\ 60 & 70 \end{bmatrix} \qquad -2 \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 0 & -2 & -4 \\ -6 & -8 & -10 \\ -12 & -14 & -16 \end{bmatrix}$$

Properties of scalar multiplication include some trivial observations.

- $0\,\mathbf{M} = \mathbf{0}$
- $a\,\mathbf{0} = \mathbf{0}$
- $1\,\mathbf{M} = \mathbf{M}$

Analagous to the vector scalar multiplication operation, matrix scalar multiplication has these properties.

- $a(b\mathbf{M}) = (ab)\mathbf{M}$
- ***Distribution***: $(a + b)\mathbf{M} = a\mathbf{M} + b\mathbf{M}$
- ***Distribution***: $a(\mathbf{M} + \mathbf{N}) = a\mathbf{M} + a\mathbf{N}$

Addition of two Matrices

Two matrices may be added if they are the same size (have the same number of rows and columns). The addition of two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p \times q}$ is denoted $\mathbf{S} = \mathbf{A} + \mathbf{B}$ and is defined as

$$s_{ij} = a_{ij} + b_{ij}, \text{ where } 0 \le i < p,\ 0 \le j < q$$

For example,

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} + \begin{bmatrix} 0 & -2 & -4 \\ -6 & -8 & -10 \\ -12 & -14 & -16 \end{bmatrix} \Longrightarrow \begin{bmatrix} 0 & -1 & -2 \\ -3 & -4 & -5 \\ -6 & -7 & -8 \end{bmatrix}$$

Analogous to the vector addition operation, matrix addition has these properties.

- ***Additive Identity***: $\mathbf{M} + \mathbf{0} = \mathbf{M}$
- ***Commutivity***: $\mathbf{M} + \mathbf{N} = \mathbf{N} + \mathbf{M}$
- ***Associtivity***: $(\mathbf{L} + \mathbf{M}) + \mathbf{N} = \mathbf{L} + (\mathbf{M} + \mathbf{N})$
- ***Additive Inverse***: $\mathbf{M} + (-\mathbf{M}) = \mathbf{0}$, which can also be written $\mathbf{M} - \mathbf{M} = \mathbf{0}$

We can define matrix subtraction $\mathbf{M} - \mathbf{N} = \mathbf{M} + (-\mathbf{N})$, but be careful as matrix subtraction is neither *commutative* nor *associative*.

Multiplication

Matrix multiplication is intricate. There are requirements on the sizes of the two matrices and the process of computing the elements of the product is involved and lengthy. The straightforward algorithm is $\Theta\left(n^3\right)$ though if the matrices are large enough, there are clever algorithms that can reduce that to $\Theta\left(n^{\log_2 7}\right)$ and even lower. Here we will be concerned with very small matrices ($2 \times 2, 3 \times 3, 4 \times 4, \ldots$) so that cleverness is not required.

Two matrices may be multiplied if and only if the number of columns of the left matrix is equal to the number of rows of the right matrix. The product has the same number of rows as the left matrix and the same number of columns as the right matrix.

More precisely, for $\mathbf{A} \in \mathbb{R}^{p \times q}$ and $\mathbf{B} \in \mathbb{R}^{q \times r}$, the product $\mathbf{P} = \mathbf{AB}$ will be $\in \mathbb{R}^{p \times r}$. The elements of $\mathbf{P}$ are

$$p_{ij} = \sum_{k=0}^{q-1} a_{ik} b_{kj} \qquad \text{where } 0 \leq i < p, 0 \leq j < r$$

For example,

$$\begin{bmatrix} 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 30 & 36 & 42 \\ 39 & 48 & 57 \end{bmatrix}$$

Out[11]:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} ag + bi + ck & ah + bj + cl \\ dg + ei + fk & dh + ej + fl \end{bmatrix}$$

Out[12]:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{k} & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & m \\ -\sin(\theta) & 0 & \cos(\theta) & n \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & m \\ 0 & 0 & 0 & 0 \\ -\frac{\sin(\theta)}{k} & 0 & \frac{\cos(\theta)}{k} & 1 + \frac{n}{k} \end{bmatrix}$$

Out[13]:

$$\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_x x \\ s_y y \end{bmatrix}$$

Out[14]:

$$\begin{bmatrix} 1 & h_x \\ h_y & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} h_x y + x \\ h_y x + y \end{bmatrix}$$

Out[15]:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos(\theta) - y\sin(\theta) \\ x\sin(\theta) + y\cos(\theta) \end{bmatrix}$$

Out[16]:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & h_x \\ h_y & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$$

$$= \begin{bmatrix} s_x\left(-h_y\sin(\theta) + \cos(\theta)\right) & s_y\left(h_x\cos(\theta) - \sin(\theta)\right) \\ s_x\left(h_y\cos(\theta) + \sin(\theta)\right) & s_y\left(h_x\sin(\theta) + \cos(\theta)\right) \end{bmatrix}$$

Out[17]:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & h_x \\ h_y & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= \begin{bmatrix} s_x x\left(-h_y\sin(\theta) + \cos(\theta)\right) + s_y y\left(h_x\cos(\theta) - \sin(\theta)\right) \\ s_x x\left(h_y\cos(\theta) + \sin(\theta)\right) + s_y y\left(h_x\sin(\theta) + \cos(\theta)\right) \end{bmatrix}$$

Out[18]:

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 & t_x \\ -\sin(\theta) & \cos(\theta) & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} t_x + x\cos(\theta) + y\sin(\theta) \\ t_y - x\sin(\theta) + y\cos(\theta) \\ t_z + z \\ 1 \end{bmatrix}$$

Properties of matrix multiplication include

- ***Associativity***: $(\mathbf{LM})\mathbf{N} = \mathbf{L}(\mathbf{MN})$

- ***Distribution***: $\mathbf{L}(\mathbf{M}+\mathbf{N}) = \mathbf{LM}+\mathbf{LN}$
- ***Distribution***: $(\mathbf{L}+\mathbf{M})\mathbf{N} = \mathbf{LN}+\mathbf{MN}$
- ***Identity***: $\mathbf{IM} = \mathbf{MI} = \mathbf{M}$

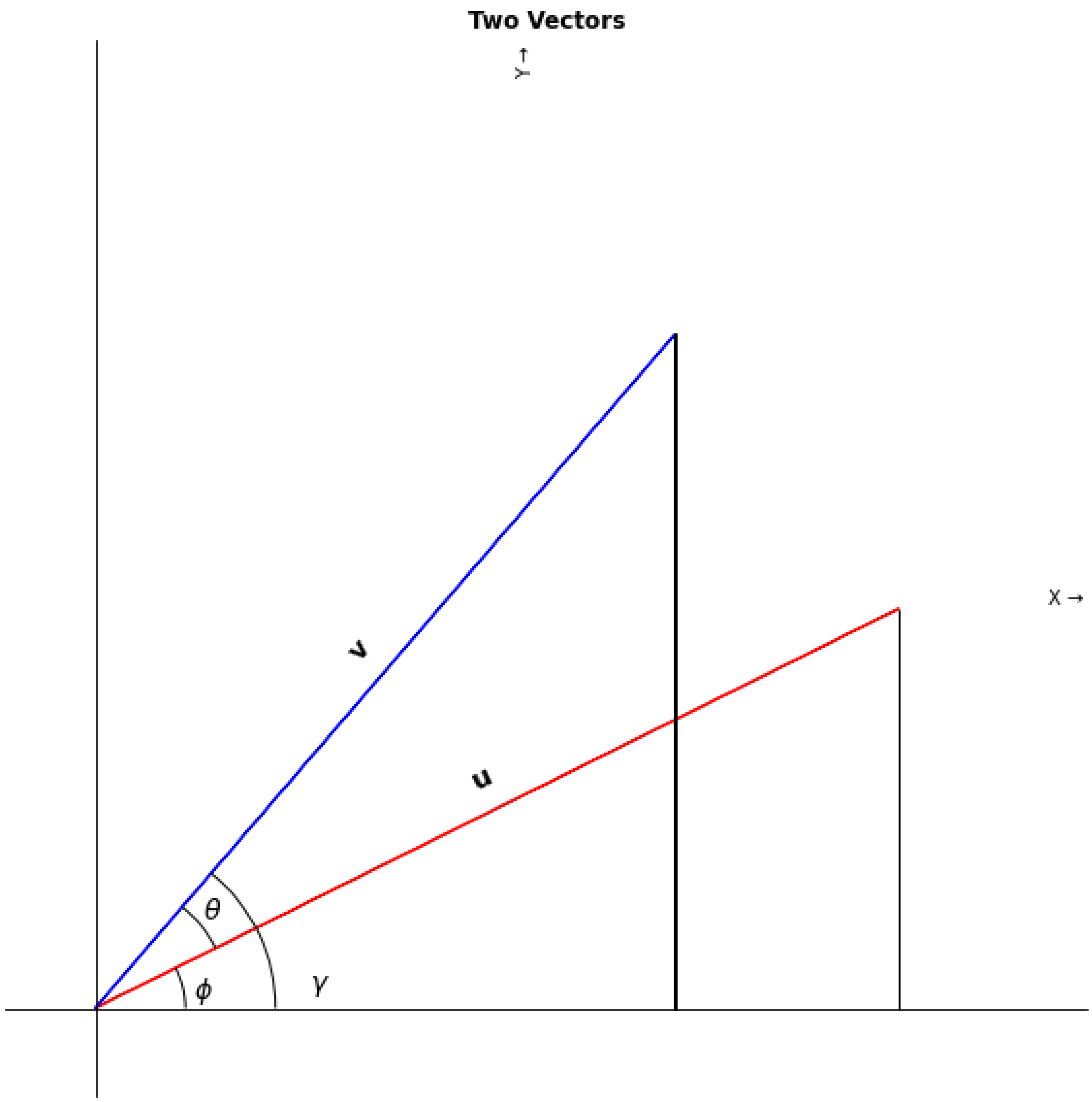
One property that we do *not* have for matrix multiplication is *commutivity*. In general, $\mathbf{MN} \neq \mathbf{NM}$.

Appendix: Showing $\mathbf{u} \cdot \mathbf{v} = 0 \iff \mathbf{u} \perp \mathbf{v}$

Above, I said this was easy to show using basic trigonometry.

> ***Don't read this until you have thought about it for a bit.*** Really, you should be trying to develop your own insights rather than just reading what someone else has written.

Anyway, take a look at the *Two Vectors* diagram below. It shows $\mathbf{u}$ and $\mathbf{v}$, with $\mathbf{u}$ making the angle ϕ with the x axis and $\mathbf{v}$ making the angle γ. The angle between the two vectors is therefore $\theta = \gamma - \phi$.



The dot product $\mathbf{u} \cdot \mathbf{v}$ can be computed as follows,

$$
\begin{aligned}
\mathbf{u}\cdot\mathbf{v} &= \begin{bmatrix} u_x \\ u_y \end{bmatrix} \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix} && \text{(vector definition)} \\
&= \begin{bmatrix} \|\mathbf{u}\|\cos\phi \\ \|\mathbf{u}\|\sin\phi \end{bmatrix} \cdot \begin{bmatrix} \|\mathbf{v}\|\cos\gamma \\ \|\mathbf{v}\|\sin\gamma \end{bmatrix} && \text{(trigonometry)} \\
&= \|\mathbf{u}\|\cos\phi\cdot\|\mathbf{v}\|\cos\gamma + \|\mathbf{u}\|\sin\phi\cdot\|\mathbf{v}\|\sin\gamma && \text{(dot product definition)} \\
&= \|\mathbf{u}\|\,\|\mathbf{v}\|\,(\cos\phi\cos\gamma + \sin\phi\sin\gamma) && \text{(algebra)} \\
&= \|\mathbf{u}\|\,\|\mathbf{v}\|\cos(\gamma-\phi) && \text{(trigonometry identity)} \\
&= \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta && \text{(definition from diagram)}
\end{aligned}
$$

At this point, we see that $\mathbf{u}\cdot\mathbf{v}$ will be zero in one of three ways, either $\|\mathbf{u}\| = 0$, $\|\mathbf{v}\| = 0$, or $\cos\theta = 0$ (or some combination of these three ways). Well, we are interested only in those cases where $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$, so for the dot product to be zero, we must have $\cos\theta = 0$, which occurs only when $\theta = \frac{\pi}{2} + n\pi, n \in \mathbb{Z}$. This is just another way of saying that the two vectors are *orthogonal*.

Ta-da!

That was easy, wasn't it?

By the way, this argument works also to show that $|\mathbf{u}\cdot\mathbf{v}| = \|\mathbf{u}\|\,\|\mathbf{v}\| \iff \mathbf{u} \parallel \mathbf{v}$. (This is a specialization of the *Cauchy-Schwarz Inequality*.)

Since $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$, the only way $|\mathbf{u}\cdot\mathbf{v}| = \|\mathbf{u}\|\,\|\mathbf{v}\|$ is $\cos\theta = 1$ or -1, which occurs only when $\theta = 0 + n\pi, n \in \mathbb{Z}$. This is just another way of saying that the two vectors are *parallel*.

Strictly speaking, *parallel* means pointing in the *same* direction — that is, $\cos\theta = 1$ — and *anti-parallel* means pointing in the *opposite* direction — $\cos\theta = -1$. However, common usage is to use the term *parallel* for either case. (The strict term for either *parallel* or *anti-parallel* is *colinear* or *having the same orientation*.)