MAT 2114 Intro to Linear Algebra

Joe Wells
Virginia Tech

Fall 2021

Contents

Preface

There are many different approaches to linear algebra, and everyone has their preference. This document is compiled from the course I taught starting in the Spring of 2020 at Virginia Tech, where both the book (*Linear Algebra: A Modern Introduction* 4th Ed. by David Poole) and order of topics covers were suggested to me by some others in the department. Although not formally stated anywhere, this class was largely geared towards math-adjacent students (engineering, physics, computer science, etc.) and so these notes and the presentation are at a lower level of abstraction (and occasionally rigor) than what one might experience in another introductory linear algebra course. In hindsight, I probably would have picked both a different text and order in which to introduce the topics. For example, I would delay the coverage of linear systems. Most students are already familiar with them, but they are completely unmotivated and their sole purpose seems to be to introduce a computational tool without any context. On the opposite end of the spectrum, why are linear transformations introduced so late in the text when they are really one of the most central objects of study in the whole of linear algebra? In time, I hope to turn these notes into a book which follows, in my opinion, a more natural and modern treatment of this beatiful subject.

I would like to thank the many students who inadvertently served as my copy editors each semester as these notes evolved.

1.1 The Geometry and Algebra of Vectors

Especially following Descartes' seminal contribution *La Géométrie*, we frequently blur the line between geometry and algebra – the reader is assuredly familiar with thinking about real numbers as points on a number line, or as ordered pairs of real numbers as points in the plane. But the real numbers come equipped with some natural algebraic operations – we can add and multiply them (hence also subtract and divide them). It's not unreasonable to ask whether this algebraic structure continues to ordered pairs of real numbers, but of course doing so requires defining the operations for ordered pairs of real numbers that are analogous to addition and multiplication. As it turns out that the naïve idea for doing so is very close to correct, although we'll see that we have to weaken the notion of multiplication slightly to allow for a meaningful geometric interpretation.

1.1.1 Definitions and Examples

Definition. A ***(real) vector space***, V, is a set of objects (called ***vectors***) with two operations – ***vector addition*** (denoted $+$) and ***scalar multiplication*** (no symbol) – satisfying the following properties: for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and for all real numbers a, b (called ***scalars***),

(a) $\mathbf{u}+\mathbf{v}$ is in V [closure]

(b) $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$ [commutativity]

(c) $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$ [associativity]

(d) There is some vector $\mathbf{0}$, called the ***zero vector***, so that $\mathbf{u}+\mathbf{0} = \mathbf{u}$ for all vectors $\mathbf{u}$. [additive identity]

(e) For each $\mathbf{u}$ in V, there is some vector $-\mathbf{u}$ for which $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$. [additive inverse]

(f) $a\mathbf{u}$ is in V [closure]

(g) $a(\mathbf{u}+\mathbf{v}) = a\mathbf{u}+a\mathbf{v}$ [distributivity]

(h) $(a+b)\mathbf{u} = a\mathbf{u}+b\mathbf{u}$ [distributivity]

(i) $(ab)\mathbf{u} = a(b\mathbf{u})$ [associativity]

(j) $1\mathbf{u} = \mathbf{u}$ [multiplicative identity]

It turns out that vector spaces are very common and you're probably already familiar with many of them without even knowing it.

Example 1.1.1. The real numbers form a real vector space when endowed with the normal addition and multiplication operations.

Example 1.1.2. The set of all ordered pairs of real numbers, (x, y), is a real vector space when endowed with the following operations.

- addition: $(x_1, y_1) + (x_2 + y_2) = (x_1 + x_2, y_1 + y_2)$
- scalar multiplication: $r(x, y) = (rx, ry)$

The pair $(0, 0)$ is the zero vector in this space.

Example 1.1.3. The set of all polynomials with real coefficients and degree at most n, $a_nx^n + \cdots + a_1x + a_0$, is a vector space when considered the usual addition and scalar multiplication.

- addition: $(a_nx^n + \cdots + a_0) + (b_nx^n + \cdots + b_0) = (a_n + b_n)x^n + \cdots + (a_0 + b_0)$
- scalar multiplication: $r(a_nx^n + \cdots + a_0) = (ra_n)x^n + (ra_0)$

The number 0 is the zero vector in this space, and this space is sometimes denoted $\mathcal{P}^n$.

Example 1.1.4. The set of all continuous real-valued functions on $\mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$ is a vector space when considered with the usual function addition and scalar multiplication.

- addition: $f_1(x) + f_2(x) = (f_1 + f_2)(x)$
- scalar multiplication: $r(f(x)) = (rf)(x)$

The function $f(x) = 0$ is the zero vector in this space, and this space is denoted $C(\mathbb{R})$.

It is straightforward to show that each of the above is a vector space and we leave it as an exercise to the reader.

1.1.2 Geometric Interpretation of Vector Operations

Now we'll take a geometric interpretation of vectors to help justify the naturality of the operations of vector addition and scalar multiplication. Let $o = (0, 0)$, $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ be some points in the plane. Let $\overrightarrow{op_1}$ be the arrow from o to p_1, and similarly let $\overrightarrow{op_2}$ be the arrow from o to p_2. Furthermore, let $p_3 = p_1 + p_2$ (with addition as described in Example **??**). Since arrows communicate to us a notion of length and direction, the arrow $\overrightarrow{op_3}$ can be described as the total displacement and direction indicated by placing the two arrows $\overrightarrow{op_1}$ and $\overrightarrow{op_2}$ "head-to-tail", as is illustrated in Figure 1.1.1.

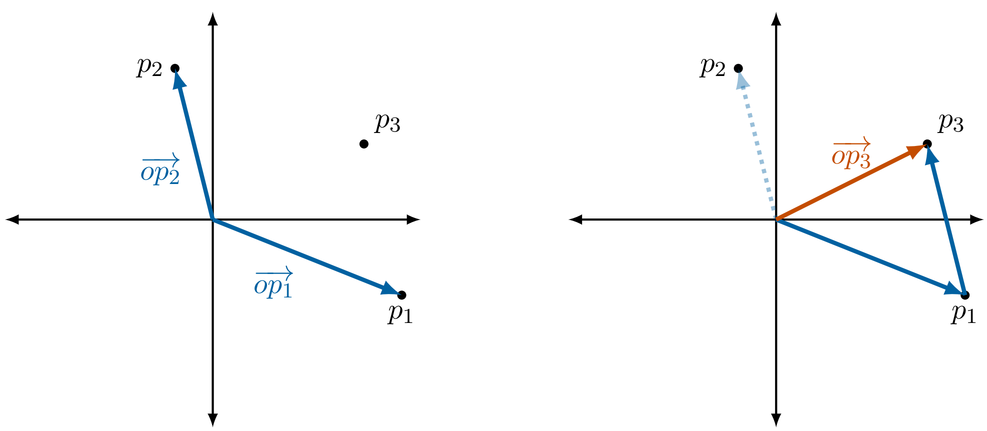


Figure 1.1.1: The original vectors (left) and "head-to-tail" vector addition (right).

With p_1 as before, consider some real number r. By the scalar multiplication operation described in Example **??**, we can consider the point $p_4 = rp_1 = (rx_1, ry_1)$. As the name suggests, scalar multiplication by a real number r has the effect of *scaling* the arrow $\overrightarrow{op_1}$. In the case that $r > 0$, the arrow $\overrightarrow{op_4}$ points in the same direction as $\overrightarrow{op_1}$ and its length is scaled by r. In the case that $r < 0$, the arrow $\overrightarrow{op_4}$ points in the opposite direction of $\overrightarrow{op_1}$ and its length is scaled by $|r|$. (See Figure 1.1.2)

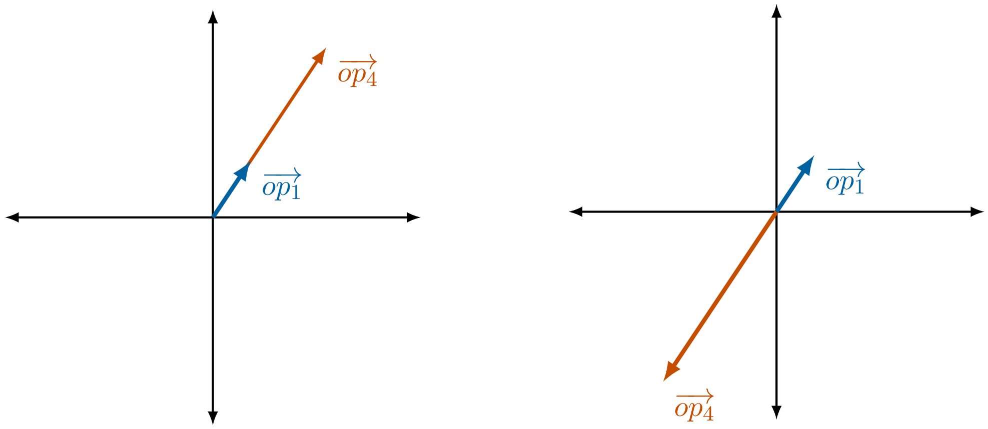


Figure 1.1.2: The original vector scaled by $r > 0$ (left) and $r < 0$ (right).

We can extend this same idea to ordered n-tuples of real numbers $(x_1, x_2, \ldots, x_n)$, associating them with arrows in n-dimensional space (the word "dimension" here should be understood only in an intuitive sense; the definition will be made precise in a later chapter), which leads us to the following definition.

Definition. $\mathbb{R}^n$ is the set of arrays with n real entries of the form

$$\begin{bmatrix} x_1, \ldots, x_n \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The x_i appearing above are called ***components*** of the arrays.

Theorem 1.1.5. $\mathbb{R}^n$ *is a vector space* $\mathbb{R}^n$ *is a vector space with addition given by*

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix},$$

with scalar multiplication given by

$$r \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} rx_1 \\ \vdots \\ rx_n \end{bmatrix},$$

and with zero vector

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Definition. Any vector $\mathbf{v}$ in $\mathbb{R}^n$ may be written as a ***row vector***

$$\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

or as a ***column vector***

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Each of these presentations represents the same object and should be regarded as the same. However, certain computations are very much reliant upon the choice of representation. Throughout this text, we will almost exclusively prefer column vectors and will be very deliberate whenever using row vectors. One could equally well develop the theory of linear algebra using row vectors, so this is merely a stylistic choice on the author's part.

For the sake of concreteness, the remainder of the text will be devoted almost exclusively to developing the theory of linear algebra using $\mathbb{R}^n$. It is a fact that every finite-dimensional vector space can be regarded being "the same" as $\mathbb{R}^n$, and so there is no loss of generality in making this specification. Most of these notions do carry over to infinite-dimensional vector spaces, although there is considerably more prerequisite knowledge and technical detail needed to discuss such things with any sort of rigor.

1.1.3 Linear combinations

With the operations of addition and scalar multiplication, the fundamental building blocks of any vector space are linear combinations.

Definition. A vector $\mathbf{u}$ in $\mathbb{R}^n$ is a ***linear combination*** of the vectors $\mathbf{v}_1, \ldots \mathbf{v}_n$ if there are scalars $r_1, \ldots, r_n$ so that

$$\mathbf{u} = r_1\mathbf{v}_1 + \cdots + r_n\mathbf{v}_n.$$

We say that the linear combination is *trivial* if $r_1 = r_2 = \cdots = r_n = 0$.

You can think of a linear combination as some sort of recipe - the $\mathbf{v}_i$'s are the ingredients, the r_i's are the quantities of those ingredients, and $\mathbf{u}$ is the finished product.

Definition. In $\mathbb{R}^n$, there are n vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \cdots \qquad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

which we call the ***standard basis vectors for $\mathbb{R}^n$***.

For now, ignore the word *basis* above; we will give technical meaning to that later. The reason these are standard is because, when looking to decompose a vector $\mathbf{u}$ into a linear combination of vectors, then simply picking apart the components is probably the most natural thing to try first.

Example 1.1.6. The vector $\mathbf{u} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$ is a linear combination of the standard basis vectors in the following way:

$$\mathbf{u} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 5\mathbf{e_1} + 6\mathbf{e_2} + 7\mathbf{e_3}$$

With the standard basis vectors above, one can be convinced that the linear combination that appears is the unique such combination. However, in general, linear combinations need not be unique.

Example 1.1.7. The vector $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ is a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ in multiple ways:

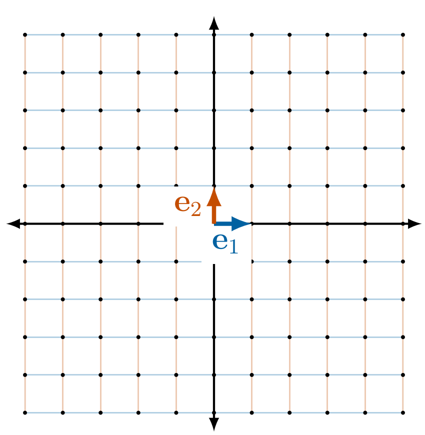
$$\begin{aligned} \mathbf{u} &= 1\mathbf{v}_1 + 0\mathbf{v}_2 + (-1)\mathbf{v}_3 \\ &= 0\mathbf{v}_1 + (-1)\mathbf{v}_2 + 1\mathbf{v}_3 \\ &= (-2)\mathbf{v}_1 + (-3)\mathbf{v}_2 + 5\mathbf{v}_3 \end{aligned}$$

The reader may be wondering precisely *when* a given vector admits a unique linear combination. This is a very important discussion with important implications, and so we will postpone this discussion for a later chapter.

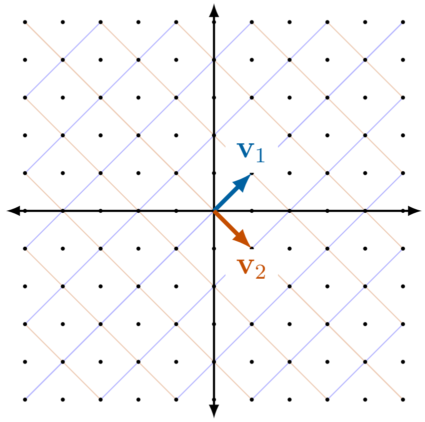
1.1.4 Geometry of Linear Combinations

The reader is probably familiar with the Cartesian grid, which provides a useful geometric depiction of the algebra. We similarly want to construct a grid that is uniquely suited to a given set of vectors in $\mathbb{F}^n$. We'll call this a ***coordinate grid*** (which is nonstandard terminology), and its construction is simple: the lines of the grids should be parallel to the vectors (in standard position) and the intersections of these grid lines correspond to integer linear combinations of vectors.

Example 1.1.8. The coordinate grid for $\mathbb{R}^2$ formed from the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ is the usual Cartesian grid.



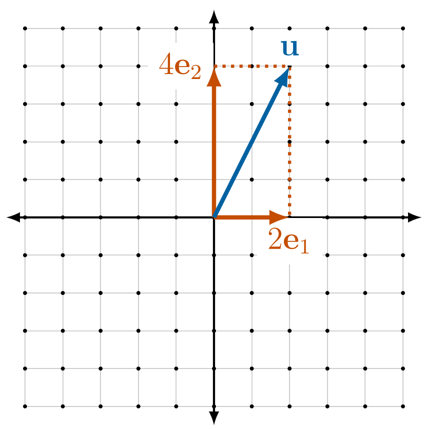
Example 1.1.9. The coordinate grid for $\mathbb{R}^2$ formed from the vectors $\mathbf{v}_1 = [1, 1]^T$ and $\mathbf{v}_2 = [1, -1]^T$ is below.



Combined with the geometric intuition about vector addition and scalar multiplication, these coordinate grids provide us with a way to visually identify the linear combination.

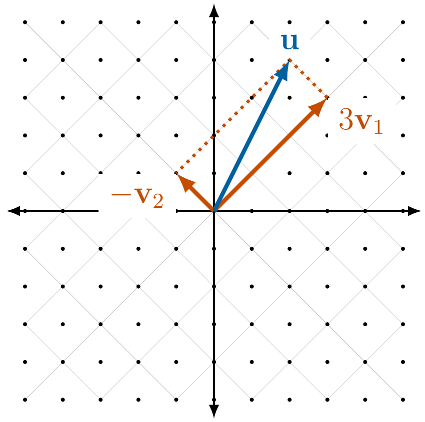
Example 1.1.10. The vector $\mathbf{u} = [2, 4]^T$ is clearly seen to be a linear combination of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$:

$$\mathbf{u} = 2\mathbf{e}_1 + 5\mathbf{e}_2$$



Example 1.1.11. The vector $\mathbf{u} = [2, 4]^T$ is clearly seen to be a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$:

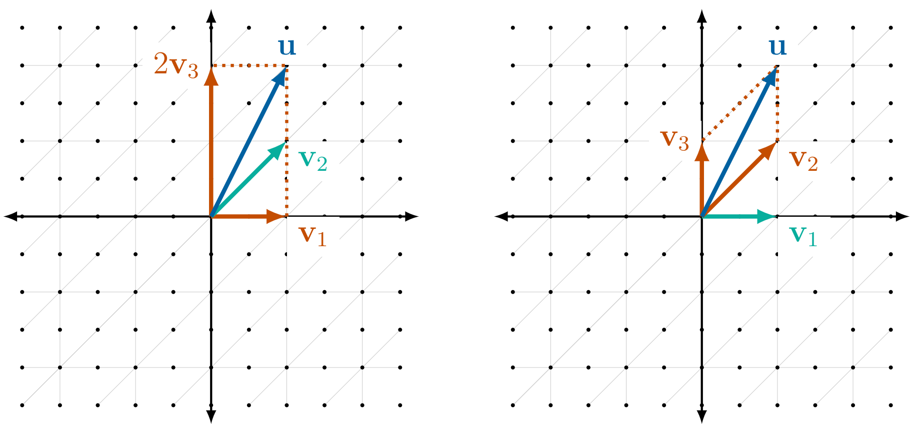
$$\mathbf{u} = 3\mathbf{v}_1 - \mathbf{v}_2.$$



Of course, this coordinate grid can also help to show us when linear combinations are not unique.

Example 1.1.12. The vector $\mathbf{u} = [2, 4]^T$ is clearly seen to be a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ in multiple different ways:

$$\begin{aligned} \mathbf{u} &= \mathbf{v}_1 + 2\mathbf{v}_3 \\ &= \mathbf{v}_2 + \mathbf{v}_3 \end{aligned}$$



1.2 Length and Angle: The Dot Product

Definition. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the ***dot product*** of $\mathbf{u}$ and $\mathbf{v}$, denoted $\mathbf{u} \cdot \mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n.$$

Remark. Note that the dot product of two vectors *is a scalar.*

The dot product has the following nice properties.

Theorem 1.2.1 (*Poole* Theorem 1.2)**.** *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *and let* k *be some scalar. Then*

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \cdot \mathbf{v}) + (\mathbf{u} \cdot \mathbf{w})$
3. $(\mathbf{v} + \mathbf{w}) \cdot \mathbf{u} = (\mathbf{v} \cdot \mathbf{u}) + (\mathbf{w} \cdot \mathbf{u})$
4. $(k\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (k\mathbf{v}) = k(\mathbf{u} \cdot \mathbf{v})$
5. *For every* $\mathbf{u}$ *we have that* $\mathbf{u} \cdot \mathbf{u} \geq 0$*, with equality if and only if* $\mathbf{u} = \mathbf{0}$*.*

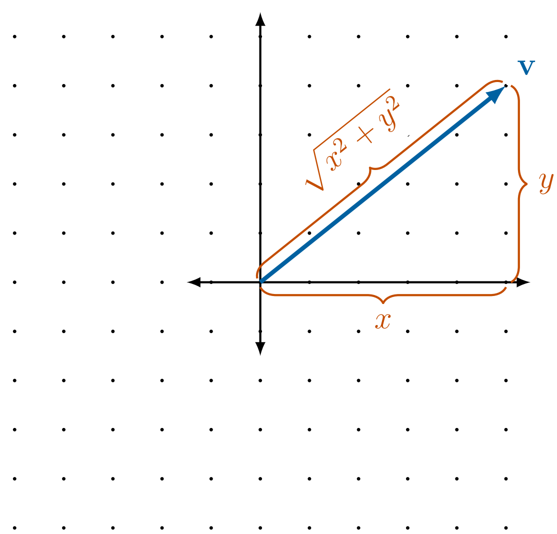
Proof. The proof is entirely straightforward and left as an exercise to the reader. □

1.2.1 Length

Notice that for a vector $\mathbf{v} = [x, y] \in \mathbb{R}^2$,

$$\mathbf{v} \cdot \mathbf{v} = x^2 + y^2,$$

which, from the Pythagorean theorem, is precisely the square of the length of $\mathbf{v}$.



Definition. The ***length*** (or ***norm***) of a vector $\mathbf{v} \in \mathbb{R}^n$ is the scalar defined by

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

The following are immediate consequences of the properties of the dot product in Theorem 1.2.1

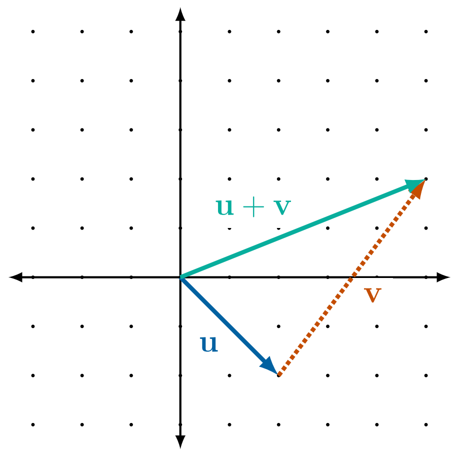
Theorem 1.2.2 (*Poole* Theorem 1.3). *For* $\mathbf{v} \in \mathbb{R}^n$ *and a scalar* k*,*

1. $\|\mathbf{v}\| = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$.
2. $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.

The following follows from the classical geometry result of the same name.

Theorem 1.2.3 (Triangle Inequality)**.** *For* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$



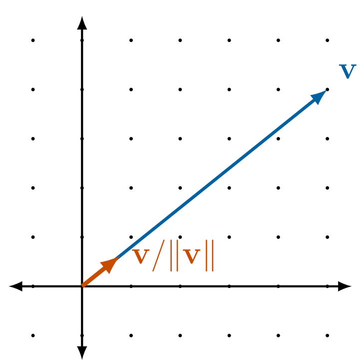
Definition. A vector $\mathbf{v}$ is called a ***unit vector*** if $\|\mathbf{v}\| = 1$.

Remark. Every unit vector in $\mathbb{R}^2$ corresponds to a point on the unit circle. Every unit vector in $\mathbb{R}^3$ corresponds to a point on the unit sphere. Generally, every unit vector in $\mathbb{R}^n$ corresponds to a point on the unit $(n-1)$-sphere.

Let $\mathbf{v}$ be any nonzero vector and let $\ell = \|\mathbf{v}\|$ be its length. Then the vector $\dfrac{\mathbf{v}}{\ell}$ is a unit vector because

$$\left\|\frac{\mathbf{v}}{\ell}\right\| = \frac{\|\mathbf{v}\|}{\ell} = \frac{\ell}{\ell} = 1$$

Definition. The process above is called ***normalization***, and it always produces a vector in the same direction as $\mathbf{v}$ but with unit length.



Remark. If $\|\mathbf{v}\| > 1$, then normalization corresponds to shrinking $\mathbf{v}$ (pictured above), but if $\|\mathbf{v}\| < 1$, then normalization stretches $\mathbf{v}$.

Remark. Despite the similarities in name, "normalization" is unrelated to the concept of a "normal vector." What you'll find is that "normal" is probably the most over-used word in mathematics. Because there aren't any around me as I type this, I'm going to go ahead and blame the physicists for the abuse of language.

1.2.2 Distances

Recall that, for two points $P(x_1, y_1)$ and $Q(x_2, y_2)$ in the plane, we have that the distance between them is given by

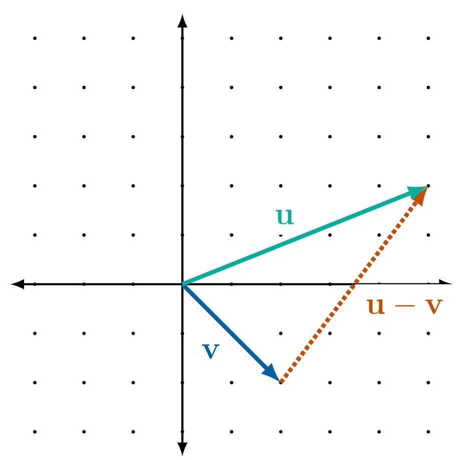
$$d(P, Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y2)^2}$$

If we identify the point $P(x_1, y_1)$ with the vector $\mathbf{u} = [x_1, y_1]$ and the point $Q(x_2, y_2)$ with the vector $\mathbf{v} = [x_2, y_2]$, then the right-hand side of the equation is just $\|\mathbf{u} - \mathbf{v}\|$. As such, we can define distances between vectors using the obvious analog.

Definition. Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the ***distance*** between $\mathbf{u}$ and $\mathbf{v}$ is

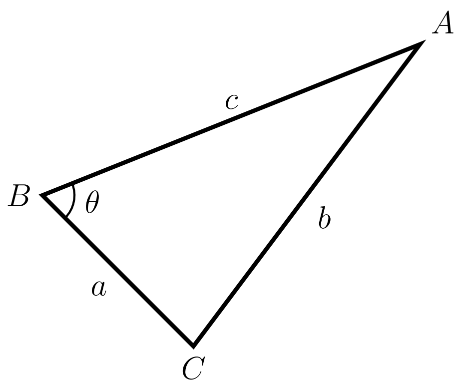
$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

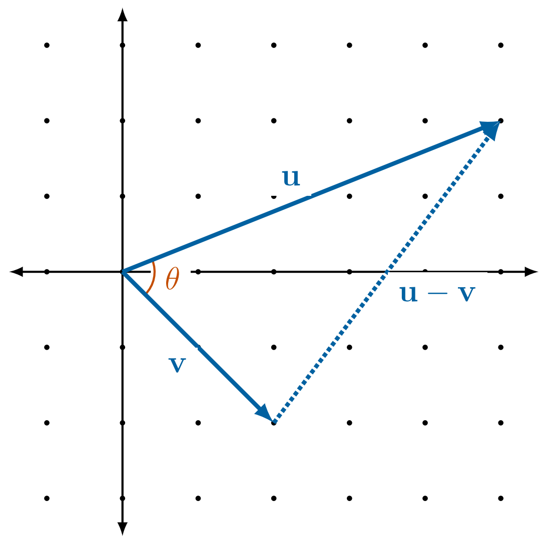
Remark. Visualizing vectors as arrows emanating from the origin, distance, as above, is actually measuring the distance between the heads of the arrows.



1.2.3 Angles

Consider a triangle $\triangle ABC$ and the angle $\theta = \measuredangle ABC$ (pictured below)





Recall that the law of cosines says

$$b^2 = a^2 + c^2 - 2ac\cos(\theta)$$

Replacing the triangle $\triangle ABC$ with the triangle formed from vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} - \mathbf{v}$ (as in the picture above on the right), we have

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

Expanding out the left-hand side of the above equation in terms of dot products, we get

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

Canceling appropriately and rearranging the equation yields

Definition. For nonzero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the ***angle*** θ between $\mathbf{u}$ and $\mathbf{v}$ satisfies

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

Example 1.2.4. Compute the angle between the vectors $\mathbf{u} = [0, 3, 3]^T$ and $\mathbf{v} = [-1, 2, 1]^T$.

From the above, we get that

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{9}{(3\sqrt{2})(\sqrt{6})} = \frac{\sqrt{3}}{2}$$

and thus

$$\theta = \arccos\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{6}.$$

The following follows immediately from the definition.

Corollary 1.2.5. $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *are perpendicular if and only if* $\mathbf{u}\cdot\mathbf{v} = 0$.

2.1 Introduction to Linear Systems

Definition. A ***linear equation*** in the variables $x_1, \ldots, x_n$ is an equation that can be written in the form

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

where $a_1, \ldots, a_n, b$ are all real numbers.

The a_i's are the ***coefficients*** and b is the ***constant term***. A ***solution*** of this equation is a vector $\mathbf{v} = [v_1, \ldots, v_n]^T$ satisfying

$$a_1v_1 + \cdots + a_nv_n = b.$$

Example 2.1.1. $4x - y = 2$ is an example of a linear equation. And notice we can rearrange it as $y = 4x - 2$, which is the equation of a line (hence why we call these "linear"). The vector $[1, 2]^T$ is a solution because

$$4(1) - (2) = 2.$$

In fact, for any real number t, the vector $[t, 4t - 2]^T$ is a solution because

$$4(t) - (4t - 2) = 2.$$

This means there are infinitely many possible solutions.

Definition. The collection of all solutions to a linear equation is called the ***solution set*** of that equation.

Noticing that

$$\begin{bmatrix} t \\ 4t-2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \begin{bmatrix} t \\ 4t \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix} + t\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

we can write the solution set to the previous example as

$$\left\{ \begin{bmatrix} 0 \\ -2 \end{bmatrix} + t\begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad \text{where} \quad t \in \mathbb{R} \right\}$$

Definition. The ***parametric form*** of the solution set is when it is written as

$$\{\mathbf{v}_0 + t_1\mathbf{v}_1 + \cdots + t_n\mathbf{v}_n \quad \text{where} \quad t_i \in \mathbb{R}\}$$

for some vectors $\mathbf{v}_i$.

Example 2.1.2. The equation

$$\sin\left(\frac{\pi}{82364423}\right) x + \sqrt{540.6464}y + z = e^{71}$$

is a linear equation with variables x, y, z because, $\sin\left(\frac{\pi}{82364423}\right)$, $\sqrt{540.6464}$, and e^{71} are just real numbers.

Example 2.1.3. The equation

$$x + xy + y + yz = 7$$

is *not* a linear equation with variables x, y, z because of the xy and yz terms.

Example 2.1.4. The equation

$$x^2 + 3^y + \log(z) = 8$$

is *not* a linear equation with variables x, y, z because of the x^2, 3^y, and $\log(z)$ terms.

Definition. A ***system of linear equations*** is a finite set of linear equations, each with the same variables (and probably different coefficients). A ***solution*** of a system of linear equations is a vector that is simultaneously a solution for each linear equation in the system. A ***solution set*** is the collection of all possible solutions to the system.

Example 2.1.5. The system

$$\begin{cases} 2x - y = 3 \\ x + 3y = 5 \end{cases}$$

has the vector $[2, 1]^T$ as a solution; in fact, this is the only solution.

Definition. A system of linear equations is called ***consistent*** if it has at least one solution, and ***inconsistent*** if it has no solutions.

Example 2.1.6. The system in Example 2.1.5 is consistent and the solution is unique.

Example 2.1.7. The system

$$\begin{cases} x - y = 0 \\ 2x - 2y = 0 \end{cases}$$

is consistent. It has the solution $[x, y]^T = [1, 1]^T$, but this is not the only solution. For any real number t, the vector $[t, t]^T$ is a solution, so there are infinitely many.

Example 2.1.8. The system

$$\begin{cases} x + y = 0 \\ x + y = 2 \end{cases}$$

has no solutions.

Definition. Two systems of linear equations are called ***equivalent*** if they have the same solution set.

Notice how easy the next system of equations is to solve by ***back-substitution***.

Example 2.1.9. Consider the system

$$\begin{cases} x + 3y + 5z = 7 \\ \quad\; 2y - 4z = 6 \\ \qquad\quad 8z = 16 \end{cases}$$

Because of this kind of "triangular structure," we quickly deduce $z = 2$, and then $2y - 4(2) = 6$ implies that $y = 7$, and then $x + 3(7) + 5(2) = 7$ implies that $x = -24$.

Since the variables themselves aren't changing, we can save time and represent any linear system by a matrix.

Definition. Given a system of linear equations

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & \cdots & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & \cdots & a_{2n}x_n & = & b_2 \\ \vdots & + & \vdots & \vdots & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & \cdots & a_{mn}x_n & = & b_m \end{cases}$$

the corresponding ***augmented matrix*** is

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

and the corresponding ***coefficient matrix*** is

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Remark. If A is the coefficient matrix for some system and $\mathbf{b} = [b_1, \ldots, b_m]^T$ is the column vector of constant terms, we may write $\left[A \mid \mathbf{b}\right]$ to represent the augmented matrix.

Remark. We will always be very explicit when we are making claims about augmented matrices specifically, and we will take care to always draw the line for an augmented matrix. When programming with matrices, however, the vertical line isn't there, so you'll have to be especially careful when considering whether the matrix you've used is representative of an augmented matrix or something else.

Example 2.1.10. The "triangular structure" of the system in Example 2.1.9 is also apparent in the corresponding augmented and coefficient matrices:

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & 7 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 8 & 16 \end{array}\right] \quad \text{and} \quad \begin{bmatrix} 1 & 3 & 5 \\ 0 & 2 & 4 \\ 0 & 0 & 8 \end{bmatrix}$$

2.1.1 Geometric Interpretation of Linear Systems

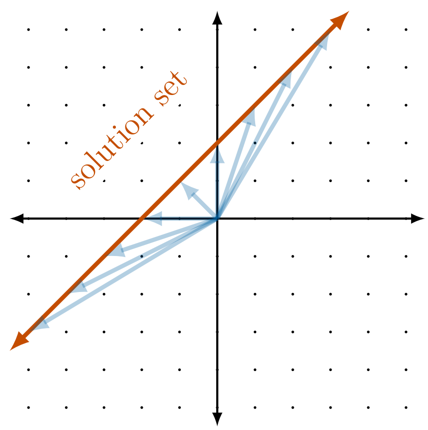
Let's consider a simple system of one linear equation in two variables

$$\{-x + y = 2.$$

The solution set to this equation is all vectors $[x, y]^T$ where $y = x + 2$. Parametrically, we would write

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} t + \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \text{where } t \in \mathbb{R} \right\}.$$

To visualize this solution set, we can make a plot in the xy-plane and try to draw all possible solution vectors.



What we find is that the head of all vectors in this solution set live along the line $y = x + 2$. In this way, we can say that the equation $-x + y = 2$ represents a line (because the solution set has one parameter) in $\mathbb{R}^2$ (because there are two variables).

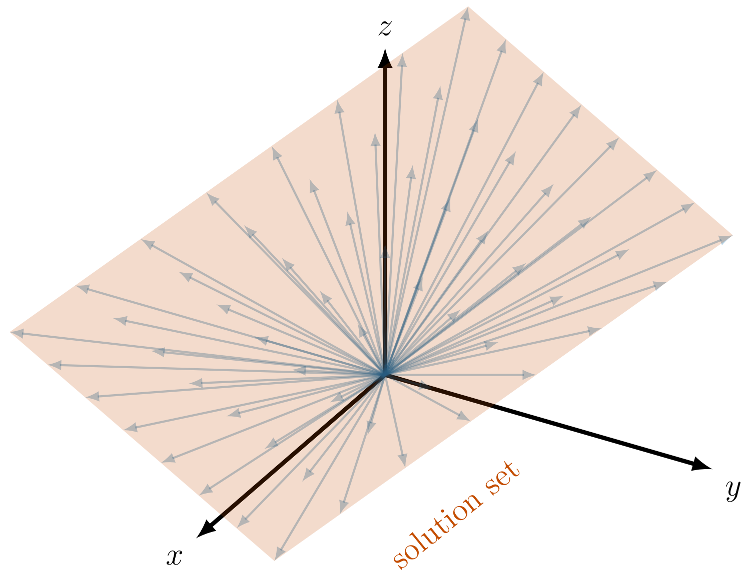
In a similar fashion, consider the linear system consisting of one equation in three variables.

$$\{-x - y + z = 2.$$

The solution set to this equation is all vectors $[x, y, z]^T$ where $z = x + y + 2$. Parametrically, we would write

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} s + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} t + \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \quad \text{where } s, t \in \mathbb{R} \right\}$$

There are three variables, so the solution set is an object in $\mathbb{R}^3$. Because the solution set has two parameters, it is a 2-dimensional object (a plane).



In general, a single equation in n variables corresponds to an $(n-1)$-dimensional object in $\mathbb{R}^n$.

What if we have two equations?

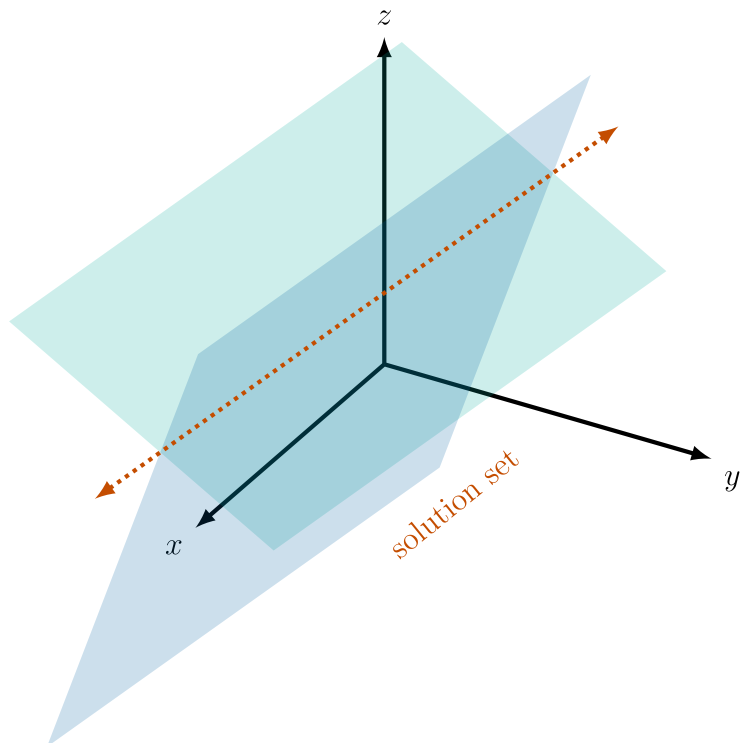
$$\begin{cases} -x - y + z = 2 \\ x - y + z = 0 \end{cases}$$

Each equation independently corresponds to a plane in $\mathbb{R}^3$, and the solution set is the set of all points *common to both* of these planes. Two arbitrary planes in $\mathbb{R}^3$ can only come in three configurations - either they are parallel (and don't have any points in common), they intersect in a line, or they are the same plane.

In this case, the linear system above has the solution set

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} t + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \text{where } t \in \mathbb{R} \right\}.$$

Since this has only one parameter, it represents a 1-dimensional object (a line) common to both planes. Geometrically, that means that the system represents two planes intersecting in a line.



By thinking of different configurations of planes, and knowing that their intersections correspond to solution sets, you can convince yourself of the following

Theorem 2.1.11. *A system of linear equations with real coefficients has exactly one of the following:*

(a) a unique solution (consistent),

(b) infinitely many solutions (consistent), or

(c) no solutions (inconsistent).

2.2 Direct Methods for Solving Linear Systems

2.2.1 Row Operations

Example 2.2.1. In Example 2.1.5 we saw that the system

$$\begin{cases} 2x - y = 3 \\ x + 3y = 5 \end{cases}$$

was consistent and had the unique solution $[x, y]^T = [2, 1]^T$. The following three systems also have the same (unique) solution $[x, y]^T = [2, 1]^T$ (this is left as an exercise for the reader), and so they are all equivalent.

$$\begin{cases} x + 3y = 5 \\ 2x - y = 3 \end{cases} \qquad \begin{cases} 2x - y = 3 \\ 100x + 300y = 500 \end{cases} \qquad \begin{cases} 2x - y = 3 \\ 3x + 2y = 8 \end{cases}$$

Looking more closely, the first systems obtained by merely swapping the equations. The second system is obtained by scaling the second equation. The third system is obtained by replacing the second equation with the sum of the first and second equations.

It turns out that this fact isn't specific to this system, but is generally true of any linear system: these three operations do not change the solution set of the system! The "elimination method" (which you may be familiar with from a previous algebra/precalculus class) uses this fact to solve systems of linear equations. If we think about what this is doing to the corresponding augmented matrices, we get what we call the *elementary row operations*.

Definition. The ***elementary row operations*** of a given matrix are the following operations:

1. Swapping Row i and Row j (denoted $R_i \leftrightarrow R_j$).
2. Multiplying Row i by a nonzero constant (denoted $kR_i \mapsto R_i$).
3. Adding (a multiple of) Row j to Row i (denoted $R_i + kR_j \mapsto R_i$).

Remark. These operations are not specific to augmented matrices, but are true of any matrices. In fact, unless explicitly stated otherwise, you should probably not ever assume that a matrix is augmented.

Given two (augmented) matrices, the above operations do not change the solution set for the corresponding linear system. So since two linear systems are equivalent if they have the same solution set, the following is a natural definition

Definition. Two matrices A and B are ***row equivalent*** if there is a sequence of elementary row operations transforming A into B.

Example 2.2.2. Using the systems in Example 2.2.1, we will show that the corresponding augmented matrices are row equivalent:

$$\left[\begin{array}{cc|c} 2 & -1 & 3 \\ 1 & 3 & 5 \end{array}\right]$$

$$\xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{cc|c} 1 & 3 & 5 \\ 2 & -1 & 3 \end{array}\right] \qquad \xrightarrow{100R_2 \mapsto R_2} \left[\begin{array}{cc|c} 2 & -1 & 3 \\ 100 & 300 & 500 \end{array}\right] \qquad \xrightarrow{R_1 + R_2 \mapsto R_2} \left[\begin{array}{cc|c} 2 & -1 & 3 \\ 3 & 2 & 8 \end{array}\right]$$

2.2.2 (Reduced) Row Echelon Form

The following systems are equivalent (it's again an exercise to the reader to verify this):

$$\begin{cases} x - y - z = 2 \\ 3x - 3y + 2z = 16 \\ 2x - y + z = 9 \end{cases} \qquad \begin{cases} x - y - z = 2 \\ y + 3z = 5 \\ 5z = 10 \end{cases} \qquad \begin{cases} x = 3 \\ y = -1 \\ z = 2 \end{cases}$$

and thus they correspond to the following row equivalent augmented matrices

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

The second and third systems are much more useful for actually *solving* the system because they have the nice triangular structure that allows us to back-substitute (or in the case of the third one, simply reading off the solution). Let's give names to this triangular structure that we like so much.

Definition. A matrix is in ***row echelon form*** (REF) if it satisfies the following properties:

(a) Any rows consisting entirely of zeros are at the bottom.

(b) In each nonzero row, the first nonzero entry (the ***leading entry***) is in a column to the left of any leading entries below it. The column containing the leading entry is sometimes called the ***pivot column***.

Example 2.2.3. The following matrices are in row echelon form.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 6 & 7 & 8 & 9 \\ 0 & 0 & 0 & 0 & 10 & 11 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Example 2.2.4. The following matrices are <u>not</u> in row echelon form. (Why?)

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

Definition. The ***reduced row echelon form*** (RREF) of a matrix is essentially the same as the row echelon form with the following additional requirements:

1. Each leading entry is 1.
2. Any entries above a *leading 1* are also 0.

Example 2.2.5. The following matrices are in reduced row echelon form.

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Example 2.2.6. The following matrices are <u>not</u> in reduced row echelon form. (Why?)

$$\begin{bmatrix} 0 & 4 & 5 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 2 & 3 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Theorem 2.2.7. *Every matrix is equivalent to a matrix in (reduced) row echelon form.*

The proof of this is actually procedural, so let's see it done in the context of an example.

Example 2.2.8.

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right]$$

1. Working left to right, find the first nonzero column in the matrix.
   The first column is nonzero
2. Among all of the rows with nonzero entries in this column, choose one and move it to Row 1.
   We'll just keep the first row where it is
3. Use elementary row operations to clear all other nonzero entries in this column (below Row 1).

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right] \xrightarrow{R_2-3R_1 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{array}\right] \tag{2.2.1}$$

$$\xrightarrow{R_3-2R_1 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{array}\right] \tag{2.2.2}$$

(2.2.3)

4. Ignoring Row 1, find the next nonzero column in this matrix.
   Ignoring Row 1, the second column is now the next nonzero column.

5. Among all of the rows below Row 1 with nonzero entries in this column, choose one and move it to Row 2.

$$\xrightarrow{R_2 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right] \tag{2.2.4}$$

6. Use elementary row operations to clear all other nonzero entries in this column (below Row 2).
   Already done.
7. Repeat this process until the matrix is in row echelon form.
   Huzzah, the matrix in Equation 2.2.4 is in row echelon form!
8. Now scale every row so that the leading term is a 1. The result will be in reduced row echelon form.

$$\xrightarrow{\frac{1}{5}R_3 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.2.5}$$

9. Working from left to right, use elementary row operations to clear all nonzero entries above each leading 1.

$$\xrightarrow{R_1 + R_2 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 2 & 7 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.2.6}$$

$$\xrightarrow{R_1 - 2R_3 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.2.7}$$

$$\xrightarrow{R_2 - 3R_3 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.2.8}$$

Remark. The row echelon form of a given matrix is not unique.

Remark. The reduced row echelon form of a matrix is unique.

Definition. The process described in the example above is called ***row reduction***.

Theorem 2.2.9 (*Poole* Theorem 2.1)**.** *Matrices A and B are row equivalent if and only if they can be row reduced to the same echelon form.*

2.2.3 Gaussian Elimination and Gauss–Jordan Elimination

Definition. Given a linear system with augmented matrix $[A|\mathbf{b}]$ in (reduced) row echelon form, the pivot columns correspond to ***leading variables*** in the system, and the other nonzero columns correspond to ***free variables*** in the system.

Definition. ***Gaussian elimination*** is the following process:

1. Write a linear system as an augmented matrix.
2. Put the matrix into row echelon form.
3. Reinterpret as a linear system and use back-substitution to solve the system for the leading variables.

Definition. ***Gauss–Jordan Elimination*** is the following process:

1. Write a linear system as an augmented matrix.
2. Put the matrix into reduced row echelon form.
3. Reinterpret as a linear system and solve the system.

Both processes take about the same amount of time by hand. But since the reduced row echelon form is unique and matrix algebra software has an RREF feature, Gauss–Jordan is usually more practical.

Example 2.2.10. Use Gaussian–Jordan elimination to find the solution set for the given system

$$\begin{cases} x_1 - x_2 + x_3 + 4x_4 = 0 \\ 2x_1 + x_2 - x_3 + 2x_4 = 9 \\ 3x_1 - 3x_2 + 3x_3 + 12x_4 = 0 \end{cases}$$

We set up the augmented matrix and row-reduce

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & 4 & 0 \\ 2 & 1 & -1 & 2 & 9 \\ 3 & -3 & 3 & 12 & 0 \end{array}\right] \xrightarrow{R_2-2R_1 \mapsto R_2} \left[\begin{array}{cccc|c} 1 & -1 & 1 & 4 & 0 \\ 0 & 3 & -3 & -6 & 9 \\ 3 & -3 & 3 & 12 & 0 \end{array}\right]$$

$$\xrightarrow{R_3-3R_1 \mapsto R_3} \left[\begin{array}{cccc|c} 1 & -1 & 1 & 4 & 0 \\ 0 & 3 & -3 & -6 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{\frac{1}{3}R_2 \mapsto R_2} \left[\begin{array}{cccc|c} 1 & -1 & 1 & 4 & 0 \\ 0 & 1 & -1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_1+R_2 \mapsto R_1} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & -1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The corresponding system is

$$\begin{cases} x_1 + 2x_4 = 3 \\ x_2 - x_3 - x_4 = 3 \end{cases}$$

Solving for the leading variables, we get

$$\begin{cases} x_1 = 3 - 2x_4 \\ x_2 = 3 + x_3 + 2x_4 \end{cases}$$

and hence any solution is of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3-2x_4 \\ 3+x_3+2x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

Replacing our free variables x_3 and x_4 with parameters s and t (respectively), our solution set is

$$\left\{ \begin{bmatrix} 3 \\ 3 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 2 \\ 0 \\ 1 \end{bmatrix} \quad \text{where } s, t \in \mathbb{R} \right\}$$

What we have seen is that both row echelon form and reduced row echelon form are useful in the same way, but both have pros and cons. Row echelon form isn't unique and, in the case of augmented matrices, it takes a little bit more work to solve the system at the end. Reduced row echelon form is unique and makes the solution at the end easier, but requires more steps initially.

2.2.4 Rank and Number of Solutions

Example 2.2.11. In Example 2.1.8, we stated that the system

$$\begin{cases} x + y = 0 \\ x + y = 2 \end{cases}$$

was inconsistent. Look at what happens when we set up the augmented matrix and row-reduce:

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 2 \end{array}\right] \xrightarrow{R_2 - R_1 \mapsto R_2} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right]$$

That last row corresponds to the linear equation $0 = 2$, which is patently false. This means there can't possibly be a solution to the system, i.e., it is inconsistent. We state this observation as a proposition.

Proposition 2.2.12. *Let $\left[A \mid \mathbf{b}\right]$ be a system of linear equations. If the i^{th} row of A is all zeroes and the i^{th} entry of $\mathbf{b}$ is nonzero, then the system is inconsistent.*

One might ask if we can say anything about a consistent system from its (reduced) row echelon form. To answer this, we first introduce the following definition.

Definition. The ***rank*** of a matrix A is the number of nonzero rows in its (reduced) row echelon form, and is denoted $\text{Rank}(A)$.

Example 2.2.13. The rank of the coefficient matrix in Example 2.1.8 is 1, and the rank of the coefficient matrix in Example **??** is 2.

Theorem 2.2.14 (*Poole* Theorem 2.2 - The Rank Theorem)**.** *If A is the coefficient matrix of a consistent system of linear equations with n variables, then*

$$n = \text{Rank}(A) + \textit{number of free variables.}$$

Remark. It turns out this theorem is actually just a special interpretation of a much more powerful theorem called the "Rank-Nullity Theorem," but that discussion will have to wait for a later section.

Definition. A system of linear equations, $[A|\mathbf{b}]$ is ***homogeneous*** if $\mathbf{b} = \mathbf{0}$. It is ***non-homogeneous*** otherwise.

Remark. Homogeneous systems are nice because they ALWAYS have at least one solution, which is the zero vector (sometimes called the ***trivial solution***).

Theorem 2.2.15. *If $[A|\mathbf{0}]$ is a homogeneous system of m linear equations and n variables, where $m < n$, then the system has infinitely many solutions.*

Proof. Since the system is homogeneous, it has at least one solution. Since $\text{Rank}(A) \leq m$, then by the Rank Theorem

$$\text{number of free variables} = n - \text{Rank}(A) \geq n - m > 0$$

and a nonzero number of free variables implies that there are infinitely-many solutions. □

Example 2.2.16. Use Gauss-Jordan elimination to find the solution set for the given system

$$\begin{cases} x_1 - x_2 + 3x_3 + 4x_4 = 0 \\ x_1 + x_2 - x_3 - 2x_4 = 0 \end{cases}$$

Creating the augmented matrix and doing the corresponding row operations, we have

$$\left[\begin{array}{cccc|c} 1 & -1 & 3 & 4 & 0 \\ 1 & 1 & -1 & -2 & 0 \end{array}\right] \xrightarrow{R_2 - R_1 \mapsto R_2} \left[\begin{array}{cccc|c} 1 & -1 & 3 & 4 & 0 \\ 0 & 2 & -4 & -6 & 0 \end{array}\right]$$

$$\xrightarrow{\frac{1}{2}R_2 \mapsto R_2} \left[\begin{array}{cccc|c} 1 & -1 & 3 & 4 & 0 \\ 0 & 1 & -2 & -3 & 0 \end{array}\right]$$

$$\xrightarrow{R_1 + R_2 \mapsto R_1} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -2 & -3 & 0 \end{array}\right]$$

From here, we can see that x_3 and x_4 are free variables, so letting $x_3 = s$ and $x_4 = t$, we get that the solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -s - t \\ 2s + 3t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

The way this last example differs from Example 2.2.10 is that we have exactly as many free variables as we have vectors in the linear combination (instead of also having the extra constant vector added on). This is more ideal because, with the usual vector operations, the collection of all of these solutions is actually a vector space! We will explore this idea a bit further in the next section.

2.3 Spanning Sets and Linear Independence

2.3.1 Span and Spanning Sets

Notice that we can rewrite the linear system

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

as an equation of vectors

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

In this way a solution to the system corresponds to a linear combination.

Theorem 2.3.1 (*Poole* Theorem 2.4)**.** *A system of linear equations* $[A \mid \mathbf{b}]$ *is consistent if and only if* $\mathbf{b}$ *is a linear combination of the columns of* A*.*

The *number of solutions to the system* also tells us *how many ways we can make such a linear combination.* If there is a unique solution, then there is exactly one way. If there are infinitely-many solutions, there are infinitely-many ways to make the linear combination, so it may be reasonable to ask more qualitative questions about the set of all possible linear combinations and study the space of linear combinations instead.

Definition. Given a set of vectors $S = \{\mathbf{v_1}, \ldots, \mathbf{v_k}\}$ in a vector space V, we define the ***span*** of $\mathbf{v_1}, \ldots, \mathbf{v_n}$ to be the set of all linear combinations of these vectors, and we write $\text{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ or $\text{Span}(S)$. If $V = \text{Span}(S)$, then we call S a ***spanning set*** for V

With this definition, we can restate Theorem 2.3.1 as follows:

Theorem 2.3.2. *A system of linear equations* $[A \mid \mathbf{b}]$ *is consistent if and only if* $\mathbf{b}$ *is in* $\text{Span}(\mathbf{a_1}, \ldots, \mathbf{a_n})$ *(where* $\mathbf{a_i}$ *is the* i^{th} *column of* A*).*

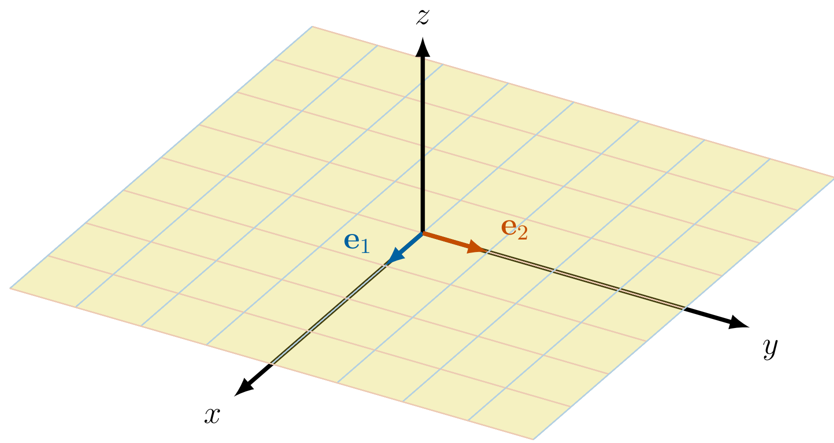
Exercise 2.3.1. If V is a (real) vector space and $\mathbf{v_1}, \ldots, \mathbf{v_k}$ is some collection of vectors in V, then the set $\text{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is also (real) vector space.

Example 2.3.3. Let $\mathbf{e_1} = [1, 0]^T, \mathbf{e_2} = [0, 1]^T$ be the standard basis vectors for $\mathbb{R}^2$. By definition, an arbitrary vector in $\text{Span}(\mathbf{e_1}, \mathbf{e_2})$ is of the form

$$x\mathbf{e_1} + y\mathbf{e_2} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

and so $\text{Span}(\mathbf{e_1}, \mathbf{e_2}) = \mathbb{R}^2$.

Example 2.3.4. Let $\mathbf{e_1} = [1, 0, 0]^T$ and $\mathbf{e_2} = [0, 1, 0]^T$ be standard basis vectors in $\mathbb{R}^3$. Span$(\mathbf{e_1}, \mathbf{e_2})$ is the collection of all vectors in $\mathbb{R}^3$ of the form $[x, y, 0]$, which is just the xy-plane. This set does not span $\mathbb{R}^3$, however, because it is missing all vectors with a nonzero 3rd coordinate (i.e. the z-direction).



Example 2.3.5. For any two vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^3$ (with $\mathbf{u}$ not a scalar multiple of $\mathbf{v}$), then Span$(\mathbf{u}, \mathbf{v})$ is a plane through the origin in $\mathbb{R}^3$.

Example 2.3.6. Consider the set $S = \left\{ \mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Then

$$\text{Span}(\mathbf{e_1}, \mathbf{e_2}, \mathbf{v}) = \{a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{v} \ : \ a, b, c \in \mathbb{R}\} = \left\{ \begin{bmatrix} a + c \\ b + c \end{bmatrix} \ : \ a, b, c \in \mathbb{R} \right\}$$

But since a, b, c can take the values of any real number, then so can $a + c$ and $b + c$. Renaming $x = a + c$ and $y = b + c$, then we exactly have that Span$(\mathbf{e_1}, \mathbf{e_2}, \mathbf{v})$ is the collection of all vectors $[x, y]^T$, which is still $\mathbb{R}^2$.

So the spanning set can, in some sense, contain redundant information, and the span may have lower "dimension" (whatever that means intuitively) than the number of vectors in the spanning set.

We can ask a similar question about this same set of vectors

Example 2.3.7. Does the set S above span $\mathbb{R}^2$?

Of course, we know the answer is "yes", but it may be good to see it explicitly. Asking whether it spans $\mathbb{R}^2$ is equivalent to asking whether every $\mathbf{w} = [x, y]^T$ in $\mathbb{R}^2$ is a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}$. By Theorem 2.3.2, this is equivalent to checking that the following system is consistent:

$$\left[\begin{array}{ccc|c} | & | & | & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{v} & \mathbf{w} \\ | & | & | & | \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 1 & x \\ 0 & 1 & 1 & y \end{array}\right]$$

This system is already in reduced row echelon form, and it is clearly consistent. Moreover, it follows from the Rank Theorem (2.2.14) that it has infinitely many solutions, and thus there are infinitely-many ways to write $\mathbf{w}$ as a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}$. In particular:

$$\begin{bmatrix} x \\ y \end{bmatrix} = (x - t)\mathbf{e}_1 + (y - t)\mathbf{e}_2 + t\mathbf{v} \qquad \text{for any real number } t.$$

Example 2.3.8. Do the vectors $\mathbf{v}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}0\\1\\1\end{bmatrix}$ span $\mathbb{R}^3$?

We are asking whether every vector $\mathbf{w} = [x, y, z]^T \in \mathbb{R}^3$ is in $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, so by Theorem 2.3.2, this is equivalent to checking that the following system is consistent:

$$\left[\begin{array}{ccc|c} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{w} \\ | & | & | & | \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right]$$

Row reducing this augmented matrix, we get

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{2}(x+y-z) \\ 0 & 1 & 0 & \frac{1}{2}(x-y+z) \\ 0 & 0 & 1 & \frac{1}{2}(-x+y+z) \end{array}\right]$$

Which is consistent and has a unique solution. Hence $\mathbb{R}^3 = \text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$.

Example 2.3.9. Do the vectors $\mathbf{v}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}2\\1\\1\end{bmatrix}$ span $\mathbb{R}^3$?

As before, we apply Theorem 2.3.2 and checking that the following system is consistent:

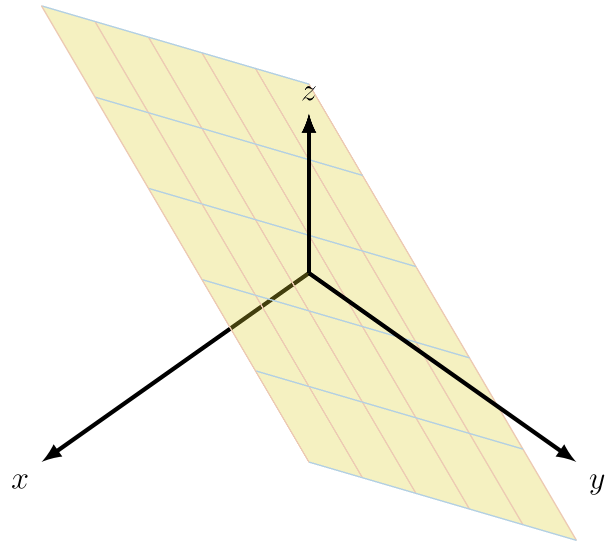
$$\left[\begin{array}{ccc|c} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{w} \\ | & | & | & | \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right]$$

Row reducing a bit, we get

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right] \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 0 & -1 & -1 & -x+y \\ 0 & 0 & 0 & z-x+y \end{array}\right]$$

and so this system is inconsistent when $z - x + y \neq 0$! In particular, this means the vector $[0, 0, 1]$ is *not* in $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$

So what is the span of these vectors? Well, it's precisely the vectors $[x, y, z]$ for which the system is consistent, i.e., the plane $z - x + y = 0$.



2.3.2 Linear (In)dependence

Suppose we can write one vector $\mathbf{w}$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$, say $\mathbf{w} = a\mathbf{u} + b\mathbf{v}$. Then $\mathbf{w}$ "depends" on $\mathbf{u}$ and $\mathbf{v}$. Clearly we can rewrite this as

$$\mathbf{w} = a\mathbf{u} + b\mathbf{v} \qquad \Longrightarrow \qquad a\mathbf{u} + b\mathbf{v} - \mathbf{w} = \mathbf{0},$$

and so we introduce the following definition.

Definition. A set of vectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$ in a vector space is ***linearly dependent*** if there are scalars $a_1, \ldots, a_k$ (not all zero) such that

$$a_1\mathbf{v_1} + \cdots + a_k\mathbf{v_k} = \mathbf{0}.$$

A set of vectors that is not linearly dependent is called ***linearly independent***.

Remark. It is always true that the equation above holds if $a_1 = \cdots = a_k = 0$, so linearly dependence says that there is some other collection of a_i's for which the equation is also true. In this way, linear independence can be thought of as saying that the only way the above equation is true is if $a_1 = \cdots = a_k = 0$.

Example 2.3.10. The set $\left\{\mathbf{v_1} = [1, 1]^T, \mathbf{v_2} = [2, 1]^T, \mathbf{v_3} = [1, 2]^T\right\}$ is linearly dependent because

$$\mathbf{v_2} + \mathbf{v_3} - 3\mathbf{v_1} = \mathbf{0}.$$

It should be clear from the way we defined linear dependence that the idea is to capture when one vector can be written as a linear combination of the others. In the above example we can easily write $\mathbf{v_3} = \frac{1}{3}\mathbf{v_2} + \frac{1}{3}\mathbf{v_2}$. In fact, this is an equivalent characterization of linear dependence.

Theorem 2.3.11 (*Poole* Theorem 2.5)**.** *The set of vectors $S = \{\mathbf{v_1}, \ldots \mathbf{v_k}\}$ in a vector space is linearly dependent if and only if at least one of the vectors can be written as a linear combination of the others.*

The proof of this fact is essentially exactly what happens in the example, so we provide it fully below. The hard part is that we have to prove two separate things (because the statement of the theorem is a "biconditional statement").

Proof. If S is linearly dependent, then we can find scalars $a_1, \ldots, a_k$, not all zero, so that

$$a_1\mathbf{v_1} + \cdots + a_k\mathbf{v_k} = \mathbf{0}.$$

Since one of the coefficients $a_i \neq 0$, then we can rearrange this as

$$\mathbf{v_i} = -\frac{a_1}{a_i}\mathbf{v_1} - \cdots \frac{a_{i-1}}{a_i}\mathbf{v_{i-1}} - \frac{a_{i+1}}{a_i}\mathbf{v_i} - \cdots - \frac{a_k}{a_i}\mathbf{v_k}.$$

Conversely, suppose $\mathbf{v_1}$ is nonzero and is a linear combination of the remaining vectors in S. Then there are constants $a_2, \ldots, a_k$, not all zero, for which

$$\mathbf{v_1} = a_2\mathbf{v_2} + \cdots + a_k\mathbf{v_k}$$

which rearranges to

$$-\mathbf{v_1} + a_2\mathbf{v_2} + \cdots + a_k\mathbf{v_k} = \mathbf{0}.$$

hence S is linearly dependent. □

Remark. The above theorem is actually a bit subtle. It doesn't say that *every* vector can be written as a linear combination of the others, just that there's at least one that can be written this way.

Exercise 2.3.2. Give an example of a dependent set of three vectors in $\mathbb{R}^n$, $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$, for which $\mathbf{u}$ is not a linear combination of $\mathbf{v}$ and $\mathbf{w}$

Why is linear independence important?

It all comes down to uniqueness. If $\mathbf{v_1}, \ldots, \mathbf{v_n}$ are linearly independent and $\mathbf{b} \in \operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_n})$, then linear independence of $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ tells us that the linear combination

$$\mathbf{b} = c_1\mathbf{v_1} + \cdots + c_n\mathbf{v_n}$$

is actually the unique one representing $\mathbf{b}$. This is great because it means that if two people can agree on a particular (ordered) linearly independent set of vectors, then we can make a vector of the coefficients $[c_1, \ldots, c_n]^T$ that unambiguously represents the vector $\mathbf{b}$. Contrast this with the following example:

Example 2.3.12. Consider the vectors $\mathbf{v_1} = [1, 0]^T$, $\mathbf{v_2} = [0, 1]^T$, $\mathbf{v_3} = [1, 1]^T$. Clearly, the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly dependent. Notably, we have

$$\mathbf{0} = \mathbf{v_1} + \mathbf{v_2} - \mathbf{v_3}.$$

So we can write $\mathbf{b} = [3, 1]^T$ in multiple ways, say,

$$\begin{aligned} \mathbf{b} &= 3\mathbf{v_1} + 1\mathbf{v_2} + 0\mathbf{v_3} \\ \mathbf{b} = \mathbf{b} + \mathbf{0} &= 4\mathbf{v_1} + 2\mathbf{v_2} - 1\mathbf{v_3} \\ \mathbf{b} = \mathbf{b} + 2\mathbf{0} &= 5\mathbf{v_1} + 3\mathbf{v_2} - 2\mathbf{v_3} \\ &\vdots \end{aligned}$$

and if we tried to refer to $\mathbf{b}$ as just a vector of its coefficients, we could have that $[3, 1, 0]^T$, $[4, 2, -1]^T$, and $[5, 3, -2]^T$ all represent the same vector.

This particular idea will become especially important when we discuss bases for a vector space, but along the way there are also several other ideas that we will see are reliant upon the notion of linear independence.

2.3.3 Using Matrices to Determine Linear (In)dependence

Given a collection of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$, determining linear (in)dependence comes down to finding whether (or not) there exist nonzero real numbers $x_1, \ldots, x_n$ such that

$$x_1\mathbf{v_1} + \cdots + x_n\mathbf{v_n} = \mathbf{0}$$

and this equivalent to checking whether or not the following system has any nontrivial solutions (i.e. solutions other than the zero vector).

$$\left[\begin{array}{ccc|c} \mathbf{v_1} & \cdots & \mathbf{v_n} & \mathbf{0} \end{array}\right]$$

Theorem 2.3.13 (*Poole* Theorem 2.6)**.** *Let $\mathbf{v_1}, \ldots \mathbf{v_m}$ be vectors in $\mathbb{R}^n$ and let A be the $n \times m$ matrix with these vectors as its columns, $A = [\mathbf{v_1}, \ldots, \mathbf{v_m}]$. Then that collection of vectors is linearly dependent if and only if the homogeneous system $[A|\mathbf{0}]$ has at least one nontrivial solution.*

The following theorem is logically equivalent to the above, but is stated to make the connection between this system and linear independence completely clear.

Theorem 2.3.14. *Let $\mathbf{v_1}, \ldots \mathbf{v_m}$ be vectors in $\mathbb{R}^n$ and let A be the $n \times m$ matrix with these vectors as its columns, $A = [\mathbf{v_1}, \ldots, \mathbf{v_m}]$. Then that collection of vectors is linearly independent if and only if the homogeneous system $[A|\mathbf{0}]$ has no nontrivial solutions.*

Example 2.3.15. Consider the column vectors $\mathbf{v_1} = [0, 1, 2]^T$, $\mathbf{v_2} = [2, 1, 3]^T$, $\mathbf{v_3} = [2, 0, 2]^T$. We can check for linear (in)dependence by row reducing $[\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}|\mathbf{0}]$ and checking the number of solutions.

$$\left[\begin{array}{ccc|c} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc|c} 0 & 2 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 3 & 2 & 0 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 2 & 3 & 2 & 0 \end{array}\right]$$

$$\xrightarrow[R_3 - 2R_1 \mapsto R_3]{\frac{1}{2}R_2 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right]$$

$$\xrightarrow[R_3 - R_2 \mapsto R_3]{R_1 - R_2 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

$$\xrightarrow[R_2 - R_3 \mapsto R_2]{R_1 + R_3 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

The system thus has no nontrivial solutions (because there are no free variables). By Theorem 2.3.14, the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly independent.

Example 2.3.16. Consider the column vectors $\mathbf{v_1} = [0,1,2]^T$, $\mathbf{v_2} = [2,1,3]^T$, $\mathbf{v_3} = [2,0,1]^T$. We can check linear dependence by using Gaussian (or Gauss-Jordan) Elimination:

$$\begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & | & \mathbf{0} \end{bmatrix} = \left[\begin{array}{ccc|c} 0 & 2 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

It's an exercise to the reader to check that the reduced row echelon form is correct. The system thus has no nontrivial solutions. By Theorem 2.3.13 the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly dependent. Of course, this isn't hard to see, as $\mathbf{v_2} = \mathbf{v_1} + \mathbf{v_3}$.

Example 2.3.17. Let $\mathbf{v_1} = [1,1]^T$, $\mathbf{v_2} = [1,-1]^T$, $\mathbf{v_3} = [x,y]^T$ be column vectors in $\mathbb{R}^2$. Using the same method from Theorem 2.3.13 we can check for linear dependence via Gaussian (or Gauss-Jordan) Elimination:

$$\begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & | & \mathbf{0} \end{bmatrix} = \left[\begin{array}{ccc|c} 1 & 1 & x & 0 \\ 1 & -1 & y & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{ccc|c} 1 & 0 & \frac{x-y}{2} & 0 \\ 0 & 1 & \frac{x+y}{2} & 0 \end{array}\right]$$

It's an exercise to the reader to check that the reduced row echelon form is correct. The system has nontrivial solutions for every vector $[x,y]^T$, so by Theorem 2.3.13 the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly dependent.

Of course, three nonzero vectors in $\mathbb{R}^2$ being linearly dependent doesn't sound too unreasonable - the corresponding linear system must have rank at most 2 and thus at least one free variable. We deduce from this the following result:

Theorem 2.3.18 (*Poole* Theorem 2.8)**.** *If $m > n$, then any set of m vectors in $\mathbb{R}^n$ is linearly dependent.*

3.1 Matrix Operations

3.1.1 Matrix Basics

Definition. A ***matrix*** is an array of numbers (called ***entries***) and has ***size*** $m \times n$ if it has m rows and n columns.

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The subscripts on the entries a_{ij} tell us that we're looking at the entry in the i^{th} row and the j^{th} column (counted top-to-bottom, left-to-right).

Fact. Two matrices are equal if and only if both (1) their sizes are equal and (2) their corresponding entries are all equal.

Remark. It's common to write $\mathbb{R}^{m\times n}$ represents the collection of $m \times n$ matrices with real number entries.

Example 3.1.1. Consider the matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, \qquad \text{and} \qquad C = \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix}.$$

$A \neq B$ because their sizes are different, and $A \neq C$ because their corresponding entries are not equal.

Definition. A matrix $A = [a_{ij}]$ is ***square*** if it has size $n \times n$. The ***diagonal of A*** is all of the entries where $i = j$:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

A square matrix is called ***diagonal*** if the only nonzero entries are along the diagonal. You may see this written as $A = \text{diag}(a_{11}, \ldots, a_{nn})$.

A bit of a subtly – the definition of diagonal just says that nonzero entries must occur along the diagonal, but the diagonal entries *do not necessarily need to be nonzero.*

Example 3.1.2. The matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are both diagonal because every entry off of the diagonal is 0.

Definition. A matrix is ***scalar*** if it is a diagonal matrix and the diagonal entries are all equal, i.e., the matrix $A = \text{diag}(r, r, \ldots, r)$ for some $r \in \mathbb{R}$.

Definition. The **zero matrix**, often denoted $\mathcal{O}$ is the matrix for which all entries are 0. Its size should be clear from context, but we may write $\mathcal{O}_{m\times n}$ if we need to specify.

Definition. The **Kroenecker delta**, denoted δ_{ij}, δ_i^j, or δ^{ij}, is the following:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Definition. The **identity matrix** I_n is the diagonal $n \times n$ matrix with all 1's along the diagonal. You may sometimes see this written as $I_n = [\delta_{ij}]$.

Example 3.1.3. It may be useful to see exactly how the Kroenecker delta leads to the identity matrix.

$$I_3 = [\delta_{ij}] = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3.1.2 Matrix Operations

Definition. Given two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$, the **matrix sum** is

$$A + B = [a_{ij} + b_{ij}] = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

Example 3.1.4. For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$,

$$A + B = \begin{bmatrix} 1+7 & 2+8 & 3+9 \\ 4+10 & 5+11 & 6+12 \end{bmatrix} = \begin{bmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}.$$

Remark. When adding two matrices, they must have the same size.

Definition. For a matrix $A = [a_{ij}]$ and a scalar r, the **scalar multiple of A** is the matrix

$$rA = [ra_{ij}]$$

Example 3.1.5. Given the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and the scalar $r = 5$,

$$rA = \begin{bmatrix} 5(1) & 5(2) \\ 5(3) & 5(4) \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix}.$$

Definition. With the notion of addition and scalar multiplication, ***subtraction*** of matrices is then defined in the obvious way:

$$A - B = A + (-1)B.$$

Remark. The zero matrix satisfies the properties you want it to: $A + \mathcal{O} = A$ and $A - A = \mathcal{O}$

3.1.3 Matrix Multiplication

A one-variable linear equation looks like

$$ax = b$$

where a, b are constants, x is some indeterminate, and ax is good old-fashioned multiplication. If we have a linear system

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & \cdots & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & \cdots & a_{2n}x_n & = & b_2 \\ \vdots & + & \vdots & \ddots & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & \cdots & a_{mn}x_n & = & b_m \end{cases}$$

it would be convenient to write it in a similar form. We thus define the ***product of a matrix and a vector*** so that the above system is captured by the equation

$$A\mathbf{x} = \mathbf{b}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

In other words, $b_i = \mathrm{Row}_i(A) \cdot \mathbf{x}$. We take this idea and extend it to products of two matrices.

Definition. Given an $m \times n$ matrix A and an $n \times p$ matrix B with columns $\mathbf{b_i}$, the ***matrix product*** of A and B is the $n \times p$ matrix

$$AB = \begin{bmatrix} | & | & & | \\ A\mathbf{b_1} & A\mathbf{b_2} & \cdots & A\mathbf{b_p} \\ | & | & & | \end{bmatrix}.$$

(And the above form is called the ***matrix-column representation*** of the product AB.) More explicitly, if A has row vectors $\mathbf{A_i}$, then AB is the $m \times p$ matrix with entries

$$AB = \begin{bmatrix} \mathbf{A_1} \cdot \mathbf{b_1} & \mathbf{A_1} \cdot \mathbf{b_2} & \cdots & \mathbf{A_1} \cdot \mathbf{b_p} \\ \mathbf{A_2} \cdot \mathbf{b_1} & \mathbf{A_2} \cdot \mathbf{b_2} & \cdots & \mathbf{A_2} \cdot \mathbf{b_p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A_m} \cdot \mathbf{b_1} & \mathbf{A_m} \cdot \mathbf{b_2} & \cdots & \mathbf{A_m} \cdot \mathbf{b_p} \end{bmatrix}$$

Example 3.1.6. Let A and B be the following matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} \text{—}\mathbf{A_1}\text{—} \\ \text{—}\mathbf{A_2}\text{—} \\ \text{—}\mathbf{A_3}\text{—} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{b_1} & \mathbf{b_2} \\ | & | \end{bmatrix}$$

We thus have that the product AB is

$$AB = \begin{bmatrix} | & | \\ A\mathbf{b_1} & A\mathbf{b_2} \\ | & | \end{bmatrix} = \begin{bmatrix} \mathbf{A_1} \cdot \mathbf{b_1} & \mathbf{A_1} \cdot \mathbf{b_2} \\ \mathbf{A_2} \cdot \mathbf{b_1} & \mathbf{A_2} \cdot \mathbf{b_2} \\ \mathbf{A_3} \cdot \mathbf{b_1} & \mathbf{A_3} \cdot \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} 23 & 28 \\ 57 & 64 \\ 89 & 100 \end{bmatrix}$$

Remark. Two matrices A, B can be multiplied even if their sizes are different. As long as the number of columns of A is equal to the number of rows of B, then the product AB exists. Moreover

$$\underset{m\times n}{A}\ \underset{n\times p}{B} = \underset{m\times p}{AB}$$

Fact. If A is an $m \times n$ matrix, then we have that

$$I_m A = A \qquad \text{and} \qquad A I_n = A$$

and it is for this reason that we call I_n the identity matrix.

3.1.4 Matrix Powers

If A is a square matrix, then for positive integers k, then we can define the power of a matrix in the intuitive way,

$$A^k = \underbrace{AA\cdots A}_{k \text{ factors}}$$

and the usual rules for exponents hold:

- $A^r A^s = A^{r+s}$
- $(A^r)^s = A^{rs}$

Example 3.1.7. If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then

$$A^2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix}$$

and

$$A^3 = A^2 A = \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 37 & 54 \\ 81 & 118 \end{bmatrix}.$$

3.1.5 Transpose

Definition. If $A = [a_{ij}]$ is an $m \times n$ matrix, then its ***transpose***, A^T is the $n \times m$ matrix who $(i, j)^{\text{th}}$ entry is a_{ij}. In other words, one obtains A^T by turning A's rows into columns and vice versa.

Visually, the transpose amounts to flipping the matrix across the red line below

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \xrightarrow{\text{flip}} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = A^T$$

Definition. A matrix A is ***symmetric*** if $A = A^T$.

Example 3.1.8. Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

One sees A is symmetric and B is not.

Remark. If A has size $m \times n$, then A^T has size $n \times m$, so the only way that $A = A^T$ is if $m = n$. In other words, symmetric matrices are always square matrices.

Example 3.1.9. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$. Compute $A^T A$ and AA^T.

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} AA^T &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 11 & 17 \\ 11 & 25 & 39 \\ 17 & 39 & 61 \end{bmatrix} \end{aligned}$$

What's interesting to notice is that, while A is not symmetric (and not even square), both AA^T and $A^T A$ are symmetric (and hence also a square matrix). These particular matrices are useful when considering *inner products* and *outer products* of vectors, respectively, although we won't be covering either of those ideas in this course.

3.2 Matrix Algebra

Theorem 3.2.1 (*Poole* Theorem 3.2 - Algebraic Properties of Matrix Addition and Scalar Multiplication)**.** *Let $A, B, C \in \mathbb{R}^{m\times n}$ and let $c, d \in \mathbb{R}$. The following are true:*

(a) $A + B = B + A$

(b) $(A + B) + C = A + (B + C)$

(c) $A + O_{m\times n} = A$

(d) $A + (-A) = O_{m\times n}$

(e) $c(A + B) = cA + cB$

(f) $(c + d)A = cA + dA$

(g) $c(dA) = (cd)A$

(h) $1A = A$

Remark. In short, Theorem 3.2.1 above says that $\mathbb{R}^{m\times n}$ is a real vector space (see page **??**).

3.2.1 Properties of Matrix Multiplication

Matrix multiplication is *not commutative in general*, and it is often the case that $AB \neq BA$. This fact clear if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$ where $m \neq n$ (just compare the sizes of AB and BA), but is possibly less obvious in the case where A, B are both square matrices. It is an exercise to find an example of this in the case of 2×2 matrices.

Exercise 3.2.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. What conditions must a, b, c, d satisfy to ensure that $AB = BA$?

So what properties does matrix multiplication have?

Theorem 3.2.2 (*Poole* Theorem 3.3 - Properties of Matrix Multiplication)**.** *Let A, B, C be matrices (whose sizes are such that the following exist) and $k \in \mathbb{R}$ a scalar. Then*

(a) $A(BC) = (AB)C$

(b) $A(B + C) = AB + AC$

(c) $(A + B)C = AC + BC$

(d) $k(AB) = (kA)B = A(kB)$

(e) $I_m A = A = AI_n$ *(if A is $m \times n$)*

Remark. This theorem implies that $\mathbb{R}^{n\times n}$ is a fancy object called a *(non-commutative) algebra*. Informally, this is a vector space with an additional operation that lets us multiply two vectors together (which, if you look closely, isn't a feature of vector spaces normally). This is outside the scope of the course, but it may be interesting to you to know that such things exist and that these properties are not unique to $\mathbb{R}^{n\times n}$.

The proof of this theorem will require the properties of the dot product (recall Proposition 1.2.1).

Proof. For simplicity, we'll introduce some notation. For a matrix M

- (this one is standard notation) M_{ij} denotes the $(i,j)^{\text{th}}$ entry of M,
- $\text{row}_i(M)$ denotes the i^{th} row of M, and
- $\text{col}_j(M)$ denotes the j^{th} column of M.

(a) Note that AB has size $m \times p$ and BC has size $n \times r$, hence both $(AB)C$ and $A(BC)$ have size $m \times r$, and thus they are equal if they're corresponding coefficients are equal.

$$((AB)C)_{ij} = \sum_{k=1}^{p}(AB)_{ik}C_{kj} = \sum_{k=1}^{p}\left(\sum_{\ell=1}^{n} A_{i\ell}B_{\ell k}\right)C_{kj} = \sum_{k=1}^{p}\sum_{\ell=1}^{n} A_{i\ell}B_{\ell k}C_{kj} = \cdots$$

$$\cdots = \sum_{\ell=1}^{n}\sum_{k=1}^{p} A_{i\ell}B_{\ell k}C_{kj} = \sum_{\ell=1}^{n} A_{i\ell}\left(\sum_{k=1}^{p} B_{\ell k}C_{kj}\right) = \sum_{\ell=1}^{n} A_{i\ell}\,(BC)_{\ell j} = (A(BC))_{ij}$$

(b) Let $A \in \mathbb{R}^{m\times n}$ and $B, C \in \mathbb{R}^{n\times p}$. Notice that $A(B+C)$ and $AB + AC$ have the same size, hence they are equal if they have the same corresponding elements.

$$\begin{aligned}(A(B+C))_{ij} &= \text{row}_i(A)\cdot\text{col}_j(B+C)\\ &= \text{row}_i(A)\cdot(\text{col}_j(B)+\text{col}_j(C))\\ &= \text{row}_i(A)\cdot\text{col}_j(B)+\text{row}_i(A)\cdot\text{col}_j(C) = (AB)_{ij} + (AC)_{ij}.\end{aligned}$$

(c) Let $A, B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times p}$. Notice that $(A+B)C$ and $AC + BC$ have the same size, hence they are equal if they have the same corresponding elements.

$$\begin{aligned}((A+B)C)_{ij} &= \text{row}_i(A+B)\cdot\text{col}_j(C)\\ &= (\text{row}_i(A)+\text{row}_i(B_i))\cdot\text{col}_j(C)\\ &= \text{row}_i(A)\cdot\text{col}_j(C)+\text{row}_i(B)\cdot\text{col}_j(C) = (AC)_{ij} + (BC)_{ij}.\end{aligned}$$

(d) Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, and $k \in \mathbb{R}$. Notice that $k(AB)$, $(kA)B$ and $A(kB)$ all have the same size $m \times p$, hence they are equal if they have the same corresponding elements.

$$\begin{aligned}(k(AB))_{ij} &= k\,(\text{row}_i(A)\cdot\text{col}_j(B))\\ &= \text{row}_i(kA)\cdot\text{col}_j(B) = ((kA)B)_{ij}\\ &= \text{row}_i(A)\cdot\text{col}_j(kB) = (A(kB))_{ij}\end{aligned}$$

(e) Let $A \in \mathbb{R}^{m\times n}$. Writing the $m \times m$ identity matrix $I_m = [\delta_{ij}]$ using the Kroenecker delta (c.f. page 37), we note that $I_m A$ and A have the same size, hence they are equal if they have the same corresponding elements.

$$\begin{aligned}(I_m A)_{ij} &= \text{row}_i(I_m)\cdot\text{col}_j(A)\\ &= \delta_{i1}A_{1j} + \delta_{i2}A_{2j} + \cdots + \delta_{im}A_{mj}\\ &= \delta_{ii}A_{ij} && \text{(the only nonzero term in the sum)}\\ &= A_{ij}\end{aligned}$$

Similarly, for the $n \times n$ identity matrix I_n,

$$\begin{aligned}
(AI_n)_{ij} &= \text{row}_i(A) \cdot \text{col}_j(I_n) \\
&= A_{i1}\delta_{1j} + A_{i2}\delta_{2j} + \cdots + A_{in}\delta_{nj} \\
&= A_{ij}\delta_{jj} && \text{(the only nonzero term in the sum)} \\
&= A_{ij}
\end{aligned}$$

□

Example 3.2.3. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$. Then

$$\begin{aligned}
A(BC) &= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \left(\begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix} \right) \\
&= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \left(\begin{bmatrix} 9 & -14 \\ 1 & -2 \\ -7 & 10 \end{bmatrix} \right) \\
&= \begin{bmatrix} -10 & 12 \\ -1 & -6 \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
(AB)C &= \left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix} \right) \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix} \\
&= \left(\begin{bmatrix} 2 & -4 \\ 11 & -4 \end{bmatrix} \right) \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix} \\
&= \begin{bmatrix} -10 & 12 \\ -1 & -6 \end{bmatrix}
\end{aligned}$$

Theorem 3.2.4 (*Poole* Theorem 3.4 - Properties of the Transpose)**.** *Let A and B be matrices (whose sizes are such that the indicated operations can be performed) and let k be a scalar. Then*

(a) $(A^T)^T = A$

(b) $(kA)^T = k(A^T)$

(c) $(A + B)^T = A^T + B^T$

(d) $(AB)^T = B^T A^T$

(e) $(A^r)^T = (A^T)^r$ *for all nonnegative integers r.*

Proof. We use the same notation as in the proof of Theorem 3.2.2.

(a) If A has size $m \times n$, then A^T has size $n \times m$, and then $(A^T)^T$ has size $m \times n$. Thus these matrices are equal if they have equal corresponding entries.

$$((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij}$$

(b) If $A \in \mathbb{R}^{m\times n}$, then $kA \in \mathbb{R}^{m\times n}$ and thus $(kA)^T$ has size $n \times m$. As well, since A^T has size $m \times n$, then kA^T has size $m \times n$.

$$((kA)^T)_{ij} = (kA)_{ji} = kA_{ji} = k(A^T)_{ij}$$

(c) Let $A, B \in \mathbb{R}^{m\times n}$. It is straightforward to see that $(A+B)^T$ and A^T, B^T have size $n \times m$. Then

$$((A+B)^T)_{ij} = (A+B)_{ji} = A_{ji} + B_{ji} = (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij}$$

(d) Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$. Note that $(AB)^T$ an $B^T A^T$ both have the same size, hence they are equal if their corresponding entries are equal.

$$\begin{aligned} \left((AB)^T\right)_{ij} &= (AB)_{ji} \\ &= \text{row}_j(A) \cdot \text{col}_i(B) \\ &= \text{col}_j(A^T) \cdot \text{row}_i(B^T) \\ &= \text{row}_i(B^T) \cdot \text{col}_j(A^T) = (B^T A^T)_{ij}. \end{aligned}$$

(e) This is a corollary of item (d).

□

Example 3.2.5. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix}$. Then

$$(AB)^T = \left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix} \right)^T = \begin{bmatrix} 2 & -4 \\ 11 & -4 \end{bmatrix}^T = \begin{bmatrix} 2 & 11 \\ -4 & -4 \end{bmatrix}$$

and

$$B^T A^T = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 11 \\ -4 & -4 \end{bmatrix}$$

and

$$A^T B^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 3 & 1 & -1 \\ 2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 11 & 1 & -9 \\ 16 & 2 & -12 \\ 21 & 3 & -15 \end{bmatrix}$$

So clearly $(AB)^T = B^T A^T$, but $(AB)^T \neq A^T B^T$ in general.

3.3 The Inverse of a Matrix

Motivation: If $a, b \in \mathbb{R}$ and x is some unknown and we wanted to solve for x in the equation

$$ax = b,$$

we would do so multiplying both sides by $a^{-1} = \frac{1}{a}$ to get that $x = a^{-1}b$. We'd like to be able to do this same thing for the system of linear equations

$$A\mathbf{x} = \mathbf{b}$$

where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. But alas, we don't have a notion of division of matrices.

Let's think – if $a \in \mathbb{R}$ is some nonzero number, then a^{-1} is just some other real number for which $aa^{-1} = a^{-1}a = 1$. Since the $n \times n$ identity matrix I_n plays the role of 1, multiplicatively, then the natural way to define the matrix we desire is

Definition. For an nonzero $n \times n$ matrix A, the ***inverse of A***, denoted A^{-1}, is the $n \times n$ matrix satisfying

$$AA^{-1} = A^{-1}A = I_n.$$

If the inverse exists, we say that A is ***invertible***.

Fact. Not every nonzero matrix is invertible, and we'll devote the latter half of this section to exploring when a matrix is invertible.

Remark. We only define inverses for square matrices. You will explore what happens in your homework for non-square matrices.

Theorem 3.3.1 (*Poole* Theorem 3.6)**.** *The inverse is unique.*

Proof. Suppose $A \in \mathbb{R}^{n\times n}$ is invertible, with inverses X and Y. Then

$$X = X(I_n) = X(AY) = (XA)Y = (I_n)Y = Y.$$

□

Theorem 3.3.2 (*Poole* Theorem 3.9)**.** *If $A, B \in \mathbb{R}^{n\times n}$ are invertible and $c \in \mathbb{R}$ is some nonzero scalar, then*

a. A^{-1} *is invertible and* $(A^{-1})^{-1} = A$.

b. $(cA)^{-1} = \frac{1}{c}A^{-1}$

c. $(AB)^{-1} = B^{-1}A^{-1}$

d. A^T *is invertible and* $(A^T)^{-1} = (A^{-1})^T$

e. A^n *is invertible for all positive integers* n *and* $(A^k)^{-1} = (A^{-1})^k$

Proof. Since inverses are unique, and matrices are invertible if their inverses exist, then each of these proven by merely checking that the multiplication is correct.

a. $(A^{-1})(A) = I_n$

b. $(cA)(\frac{1}{c}A^{-1}) = \frac{c}{c}AA^{-1} = I_n$.

c. $(AB)(B^{-1}A^{-1}) = AI_nA^{-1} = AA^{-1} = I_n$.

d. $A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$.

e. $(A^k)(A^{-1})^k = \underbrace{A\cdots A}_{k}\underbrace{A^{-1}\cdots A^{-1}}_{k} = \underbrace{A\cdots A}_{k-1}I_n\underbrace{A^{-1}\cdots A^{-1}}_{k-1} = \cdots = I_n$.

□

Remark. Because of the above theorem, some will use the notation A^{-n} (for n a positive integer) and A^{-T} (the transpose) to mean the obvious things:

$$A^{-n} = (A^{-1})^n = (A^n)^{-1}$$
$$A^{-T} = (A^T)^{-1} = (A^{-1})^T$$

3.3.1 Finding A^{-1}

If $A = [a_{ij}]$ is given and $X = [x_{ij}]$ is a matrix of unknowns indeterminates that we hope to solve, then by comparing the entries, the matrix equation $AX = I_n$ yields a linear system (with n^2 equations and n^2 unknowns)

$$\begin{cases} \sum_{k=1}^n a_{1k}x_{k1} &= 1 \\ \sum_{k=1}^n a_{1k}x_{k2} &= 0 \\ &\vdots \\ \sum_{k=1}^n a_{ik}x_{kj} &= \delta_{ij} \qquad \text{(the Kroenecker delta)} \\ &\vdots \\ \sum_{k=1}^n a_{nk}x_{kn} &= 1 \end{cases}$$

and you can use standard techniques to solve this system. But this system is huge and "sparse" (that is, the coefficient matrix has many 0's). Instead, let's write X and I_n in terms of their column vectors

$$X = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \quad \text{and} \quad I_n = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ | & | & & | \end{bmatrix}$$

Now we have that

$$AX = I_n$$
$$\begin{bmatrix} | & | & & | \\ A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ | & | & & | \end{bmatrix}$$

and so rather than one massive system, we can deduce from this n smaller systems

$$A\mathbf{x}_1 = \mathbf{e}_1, \quad A\mathbf{x}_2 = \mathbf{e}_2, \quad \ldots \quad A\mathbf{x}_n = \mathbf{e}_n.$$

But of course, if we do something like Gauss–Jordan elimination to solve each of these, we're always doing the same row reduction steps on the coefficient matrix! So if we just write the augmented matrix $[A|I_n]$, then our row reduction will allow us to simultaneously solve all of these systems. In particular, because A is invertible, we have that $\text{RREF}(A) = I_n$ (this isn't obvious, so you'll have to trust me), and so row reducing $[A|I_n]$ yields $[I_n|A^{-1}]$.

Definition. Suppose $A \in \mathbb{R}^{n\times n}$ is invertible. The ***Gauss–Jordan method*** is the following procedure for finding the inverse:

1. Write the augmented matrix $[A|I_n]$.
2. Row reduce fully to put A into RREF.
3. You now have $[I_n|A^{-1}]$.

Example 3.3.3. Find A^{-1} given $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

$$\left[A \mid I_2\right] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right] \xrightarrow{R_2 - 3R_1 \mapsto R_2} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right]$$
$$\xrightarrow{\frac{-1}{2} R_2 \mapsto R_2} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right]$$
$$\xrightarrow{R_1 - 2R_2 \mapsto R_1} \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right] = \left[I_2 \mid A^{-1}\right]$$

And we can check that the result is indeed the inverse:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

One can, of course, do the above process for generic 2×2 matrices, which yields the following result.

Theorem 3.3.4 (Poole thm 3.8)**.** *If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then the inverse is given by $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, provided $ad - bc \neq 0$.*

Proof. Using the Gauss-Jordan method above,

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \xrightarrow{\frac{1}{a} R_1 \mapsto R_1} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right]$$
$$\xrightarrow{R_2 - cR_1 \mapsto R_2} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{a} & -\frac{c}{a} & 1 \end{array}\right]$$
$$\xrightarrow{\frac{a}{ad-bc} R_2 \mapsto R_2} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$
$$\xrightarrow{R_1 - \frac{b}{a} R_2 \mapsto R_1} \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

□

You could do this same system-solving process for larger matrices, but the formulas are significantly worse.

How does this help us solve our system?

If A is invertible and $A\mathbf{x} = \mathbf{b}$ then, by design, it should have the solution $\mathbf{x} = A^{-1}\mathbf{b}$. Moreover, since A^{-1} is unique, we expect this solution to be unique.

Theorem 3.3.5 (*Poole* Theorem 3.7)**.** *If A is an invertible $n \times n$ matrix, then for every $\mathbf{b} \in \mathbb{R}^n$, the linear system $A\mathbf{x} = \mathbf{b}$ is consistent and has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.*

Example 3.3.6. Solve the system $A\mathbf{x} = \mathbf{b}$ given $A = \begin{bmatrix} 1 & 4 \\ 3 & 13 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$.

We could solve this the old way, or we can try our nifty new method. We quickly deduce that A^{-1} is given by

$$A^{-1} = \begin{bmatrix} 13 & -4 \\ -3 & 1 \end{bmatrix}$$

(which can be seen either by appealing to Theorem 3.3.4 or using the Gauss-Jordan Method). Hence the solution is

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 13 & -4 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = \begin{bmatrix} -15 \\ 4 \end{bmatrix}.$$

Example 3.3.7. Solve the system

$$\begin{cases} x + y = 1 \\ -x - y = -1 \end{cases}$$

Notice that this system is equivalent to the system $\{x + y = 1\}$, which has infinitely-many solutions. Notice also that the coefficient matrix for this system is

$$A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

which isn't invertible (because otherwise, attempting to apply Theorem 3.3.4 we would be dividing by 0).

Okay, so tell me, when is A invertible?

Putting it all together, we can wrap it up into the following theorem

Theorem 3.3.8 (Poole Theorem 3.12 - The Fundamental Theorem of Invertible Matrices: Pt I)**.** *Let $A \in \mathbb{R}^{n \times n}$. The following are equivalent:*

a. A is invertible.

b. A is row equivalent to I_n (i.e. its reduced row echelon form is I_n).

c. A is the product of elementary matrices.

d. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

e. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

f. The columns of A are linearly independent.

For part (c) above, we saw that it worked in one particular example. In fact, it's true for all $\mathbf{b}$ because, if A is row equivalent to the identity, then it will never have a row of all 0's in its row-reduced form (which is precisely what happened in the cases that $A\mathbf{x} = \mathbf{b}$ had no solutions or infinitely many solutions).

3.3.2 Elementary Matrices

Definition. An ***elementary matrix*** is a matrix obtained by performing an elementary row operation on the identity matrix.

The other way to think about it is that, given a matrix $A \in \mathbb{R}^{m\times n}$, an elementary matrix $E \in \mathbb{R}^{m\times m}$ is one for which the product EA has the same effect as doing an elementary row operation on A.

One can then get the inverse as the product of all of the elementary matrices.

Example 3.3.9. Given $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Use elementary matrices to compute A^{-1}.

In the following string of equalities, we'll denote the row reduction on the left-hand side and the corresponding product by elementary matrices on the right-hand side.

$$\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} &= A \\
\xrightarrow{R_2 - 3R_1 \mapsto R_2} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} A \\
\xrightarrow{\frac{-1}{2} R_2 \mapsto R_2} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} A \\
\xrightarrow{R_1 - 2R_2 \mapsto R_1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} A \\
I_2 &= \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} A
\end{aligned}$$

so again we get that $A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$.

If we write the elementary matrices above as E_1, E_2, E_3, respectively, (so that $A^{-1} = E_3E_2E_1$) then applying them to the matrix equation $AX = I$ looks something like the following

$$\begin{aligned}
AX &= I \\
E_1AX &= E_1 \\
E_2E_1AX &= E_2E_1 \\
E_3E_2E_1AX &= E_3E_2E_1 \\
IX &= E_3E_2E_1 \\
X &= E_3E_2E_1
\end{aligned}$$

What this tells us is that computing the inverse can just be done with row operations on an augmented matrix $[A|I]$ until we get $[I|A^{-1}]$ (and hence provides proof that the *Gauss-Jordan Method* works).

3.5 Subspaces, Basis, Dimension, and Rank

We've thought about solution sets as spans of vectors and also, alternatively, as lines and planes in 3-dimensional space. Now we'll formalize these ideas so that we can talk about these things in more generality. Recall the definition of a real vector space:

Definition. A ***(real) vector space*** V is a set of objects (called ***vectors***) with two operations ***vector addition*** (denoted $+$) and ***scalar multiplication*** (no symbol) satisfying the following properties: For all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and real numbers a, b (called ***scalars***),

(a) $\mathbf{u} + \mathbf{v}$ is in V [closure of addition]

(b) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ [commutativity of addition]

(c) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ [associativity of +]

(d) There is some vector $\mathbf{0}$, called the *zero vector*, so that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all vectors $\mathbf{u}$. [additive identity]

(e) For each $\mathbf{u}$ in V, there is some vector $-\mathbf{u}$ for which $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. [additive inverse]

(f) $a\mathbf{u}$ is in V [closure of scalar mult.]

(g) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$ [distributivity]

(h) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$ [distributivity]

(i) $(ab)\mathbf{u} = a(b\mathbf{u})$ [associativity of scalar mult.]

(j) $1\mathbf{u} = \mathbf{u}$ [scalar mult. identity]

From this definition, Theorem **??** can be restated as

Theorem 3.5.1 (Restatement of *Poole* Theorem 1.1)**.** *$\mathbb{R}^n$ is a real vector space.*

Remark. As it turns out, it's entirely sufficient to just think about $\mathbb{R}^n$ when discussing (finite-dimensional) real vector spaces, so if it's more comfortable for you, any time you read "vector space," you can replace it with "$\mathbb{R}^n$" in your mind and not lose any understanding.

With this in mind, we introduce the following definition:

Definition. Let V be a vector space and let W be a subset of vectors in V. We say that W is a ***subspace*** of V if it is also a vector space (with the same vector addition/scalar multiplication operations).

In order to check that a set of vectors is a subspace, one would have to check all of the axioms of the vector space definition – eww. Instead, here is an equivalent characterization of a subspace (note: this is typically a theorem in most textbooks, but your book presents it as the definition).

Definition. Let V be a vector space and let W be a subset of vectors in V. W is a ***subspace*** of V if it has the following properties:

1. $\mathbf{0}$ is in W (where $\mathbf{0}$ is the same zero vector in V).
2. If $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$. [closure of addition]
3. If $\mathbf{u} \in W$ and $k \in \mathbb{R}$ is a scalar, then $k\mathbf{u} \in W$. [closure of scalar multiplication]

(where vector addition and scalar multiplication in W are the same operations for V).

Example 3.5.2. Every vector space V is a subspace of itself.

Example 3.5.3. For any vector space V, the set $\{\mathbf{0}\}$ is a subspace of V (sometimes called the ***trivial subspace***).

Example 3.5.4. Let W be the set of all vectors in $\mathbb{R}^3$ of the form $[x, y, 0]^T$. Then W is a subspace of $\mathbb{R}^3$ because

1. $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in W$
2. $\begin{bmatrix} x_1 \\ y_1 \\ 0 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 0 \end{bmatrix} \in W$
3. $k \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ 0 \end{bmatrix} \in W$

Example 3.5.5. Let W be the set of all vectors in $\mathbb{R}^3$ of the form $[x, 0, z]^T$. Then W is a subspace of $\mathbb{R}^3$ (and the justification for this is nearly identical to the example above).

Remark. You may notice that, in the previous two examples, the subspaces were each defined by only two real numbers and both were planes (one was the xy-plane and the other was the xz-plane). You may be inclined to call either of these "$\mathbb{R}^2$," but really they're both *copies* of $\mathbb{R}^2$ living inside of $\mathbb{R}^3$. In fact, any plane in $\mathbb{R}^3$ is just a copy of $\mathbb{R}^2$. As such, there is no canonical choice of plane, so it really doesn't make sense to call any of these infinitely-many planes "$\mathbb{R}^2$."

Example 3.5.6. Let W be the set of all vectors in $\mathbb{R}^3$ of the form $[x, y, 1^T$. Then W is *not* a subspace of $\mathbb{R}^3$ because:

1. $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \notin W$
2. $\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 2 \end{bmatrix} \notin W$
3. For any $k \neq 1$, $k \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ k \end{bmatrix} \notin W$.

Theorem 3.5.7 (*Poole* Theorem 3.19)**.** *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_\mathbf{k} \in \mathbb{R}^n$. *Then* $\mathrm{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ *is a subspace of* $\mathbb{R}^n$.

Proof. For simplicity, let $W = \text{Span}(\mathbf{v_1}, \dots, \mathbf{v_k})$. We check that W satisfies the criteria for the definition of a subspace.

1. Since $\mathbf{0} = 0\mathbf{v_1} + \cdots + 0\mathbf{v_k}$, then $\mathbf{0} \in W$.
2. Let $\mathbf{u} = c_1\mathbf{v_1} + \cdots + c_k\mathbf{v_k}$ and $\mathbf{w} = d_1\mathbf{v_1} + \cdots + d_k\mathbf{v_k}$ be vectors in W. Then

$$\mathbf{u} + \mathbf{w} = (c_1 + d_1)\mathbf{v_1} + \cdots + (c_k + d_k)\mathbf{v_k}.$$

   Since $\mathbf{u} + \mathbf{w}$ is a linear combination of the vectors $\mathbf{v_i}$, then $\mathbf{u} + \mathbf{w} \in W$.
3. Let $\mathbf{u}$ be as above and $r \in \mathbb{R}$ be some scalar. Then

$$r\mathbf{u} = (rc_1)\mathbf{v_1} + \cdots + (rc_k)\mathbf{v_k}.$$

   Since $r\mathbf{u}$ is a linear combination of the vectors $\mathbf{v_i}$, then $r\mathbf{u} \in W$.

□

3.5.1 Subspaces Associated with Matrices

Armed with the notion of a "subspace" in mind, let's try to revisit some ideas involving matrices. First, a new definition

Definition. Let A be an $m \times n$ matrix.

1. The ***column space of*** A is a subspace $\mathbb{R}^m$ spanned by the columns of A. We denote it as $\text{Col}(A)$.
2. The ***row space of*** A is a subspace of $\mathbb{R}^n$ spanned by the rows of A. We denote it as $\text{Row}(A)$

Remark. Since we will prefer to think about column vectors whenever possible, it may be more useful to define $\text{Row}(A) := \text{Col}(A^T)$.

Theorem 3.5.8 (*Poole* Theorem 3.21)**.** *Let A be an $m \times n$ matrix and let N be the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then N is a subspace of $\mathbb{R}^n$.*

Proof. As before, we will approach by showing that N satisfies the three subspace criteria.

1. Homogeneous systems always have te trivial solution, hence N contains $\mathbf{0}$.
2. Let $\mathbf{x_1}, \mathbf{x_2}$ be in N. Then

$$A(\mathbf{x_1} + \mathbf{x_2}) = A\mathbf{x_1} + A\mathbf{x_2} = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

   so N is closed under addition.
3. Let $\mathbf{x}$ be in N and k be a scalar. Then

$$A(k\mathbf{x}) = k(A\mathbf{x}) = k\mathbf{0} = \mathbf{0}$$

   so N is closed under scalar multiplication.

□

Definition. N, as above, is called the ***null space of A***, and is denoted Null(A). (In some texts, it is called the *kernel of A* and is denoted ker(A).)

Exercise 3.5.1. If W is the set of solutions to the system $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b}$ is *not* the zero vector, is W a subspace of $\mathbb{R}^n$?

Example 3.5.9. Compute Col(A), Row(A), and Null(A) for $A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \end{bmatrix}$.

Letting $\mathbf{a_i}$ denote the i^{th} column of A, we see that $\mathbf{a_3} = 2\mathbf{a_2} = 4\mathbf{a_1}$, hence

$$\text{Col}(A) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right).$$

Similarly, letting $\mathbf{A_i}$ denote the i^{th} row of A, we see that $\mathbf{A_3} = \mathbf{A_2} = \mathbf{A_1}$, hence

$$\text{Row}(A) = \text{Span}([1, 2, 4]).$$

Examining the homogeneous system $A\mathbf{x} = \mathbf{0}$,

$$[A \mid \mathbf{0}] = \left[\begin{array}{ccc|c} 1 & 2 & 4 & 0 \\ 1 & 2 & 4 & 0 \\ 1 & 2 & 4 & 0 \end{array}\right] \xrightarrow[R_2 - R_1 \mapsto R_2]{R_3 - R_1 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

we see that A has rank 1, hence there are two free variables in this system, $x_2 = s$ and $x_3 = t$. We thus get that the solution set is of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2s - 4t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}$$

hence

$$\text{Null}(A) = \text{Span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}\right).$$

Definition. Let W be a subspace of a vector space and $\mathcal{B} = \{w_1, \ldots, w_k\}$ a set of vectors in W. $\mathcal{B}$ is a ***basis*** for W if

1. $W = \text{Span}(\mathcal{B})$ and
2. $\mathcal{B}$ is a linearly independent set.

Remark. Since every vector space is a subspace of itself, this definition is valid for all vector spaces. We've merely stated it in terms of subspaces to make it clear that the basis vectors must each be contained *in that subspace.*

Example 3.5.10. The standard basis vectors $\mathbf{e_i}$ in $\mathbb{R}^n$ form a basis for $\mathbb{R}^n$. We refer to $\mathcal{E} = \{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ as the ***standard basis*** for $\mathbb{R}^n$.

Example 3.5.11. Find basis for the column space $A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{bmatrix}$.

Let $\mathbf{a_i}$ represent the i^{th} column vector for A and let $\mathbf{r}_j$ denote the j^{th} column vector for RREF(A). So we have

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_5} \\ | & | & & | \end{bmatrix}$$

$$\text{and} \qquad \text{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_5 \\ | & | & & | \end{bmatrix}$$

Notice that we clearly have

$$\mathbf{r_3} = 1\mathbf{r}_1 + 2\mathbf{r}_2 \qquad \text{and} \qquad \mathbf{r}_5 = -1\mathbf{r}_1 + 3\mathbf{r}_2 + 4\mathbf{r}_4,$$

whence

$$\mathbf{a_3} = 1\mathbf{a_1} + 2\mathbf{a_2} \qquad \text{and} \qquad \mathbf{a_5} = -1\mathbf{a_1} + 3\mathbf{a_2} + 4\mathbf{a_4}.$$

(and you can check that this is true, just to confirm). It follows that

$$\text{Col}(A) = \text{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5) = \text{Span}(\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_4}).$$

Moreover,

$$\text{RREF}(\left[\mathbf{a_1} \quad \mathbf{a_2} \quad \mathbf{a_4} \mid \mathbf{0}\right]) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and so $\{\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_4}\}$ is a linearly independent set. Hence

$$\{\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_4}\} = \left\{ \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

is a basis for Col(A).

Strategy for finding a basis for the column space of a matrix:

1. Row reduce the matrix (just row-echelon form is fine)
2. Take as a basis every column (<u>in the original matrix</u>) which contains a leading entry.

Remark. It's important that you take the basis vectors from the columns of the original matrix. For example, $\text{Col}\left(\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\right) = \text{Span!}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$, but $\text{Col}(\text{RREF}(A)) = \text{Col}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$

Example 3.5.12. Find basis for the null space of $A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{bmatrix}$.

Notice that

$$\text{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

When we go to solve the system

$$A\mathbf{x} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{0}$$

we see that x_3 and x_5 are free variables and

$$\begin{aligned} x_1 &= -x_3 + x_5, \\ x_2 &= -2x_3 - 3x_5, \\ x_4 &= -4x_5. \end{aligned}$$

By setting $x_3 = s$ and $x_5 = t$, we can parameterize the solution set as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -s+t \\ -2s-3t \\ s \\ -4t \\ t \end{bmatrix} = s \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix} = s\mathbf{v_1} + t\mathbf{v_2}.$$

Clearly $\text{Null}(A) = \text{Span}(\mathbf{v_1}, \mathbf{v_2})$ and it is straightforward to check that $\{\mathbf{v_1}, \mathbf{v_2}\}$ are linearly independent, hence

$$\{\mathbf{v_1}, \mathbf{v_2}\} = \left\{ \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\text{Null}(A)$.

Strategy for finding a basis for the null space of a matrix A:

1. Row reduce the matrix.
2. Solve the system $A\mathbf{x} = \mathbf{0}$.
3. Write the solution set in parametric form $\{\mathbf{v}_1 t_1 + \cdots + \mathbf{v}_k t_k \;:\; t_1, \ldots, t_k \in \mathbb{R}\}$. (*You should have as many parameters as free variables.*)
4. Take $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ as a basis for $\text{Null}(A)$.

3.5.2 Dimension and Rank

Question: *How many bases can a vector space have?*
Answer: *Infinitely-many.*

Exercise 3.5.2. Show that, for any nonzero real numbers m, n, the set $\left\{ \begin{bmatrix} m \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ n \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$.

Maybe the better question is

Question: *How many vectors must a basis have? Can two different bases have different numbers of vectors?*

Example 3.5.13. Suppose V is a subspace of $\mathbb{R}^4$ and it has two different bases, $\mathcal{B}_1 = \{\mathbf{u_1}, \mathbf{u_2}\}$ and $\mathcal{B}_2 = \{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$.

Since $\mathcal{B}_1$ is a basis, we can write each of the $\mathcal{B}_2$-basis vectors as a linear combination of the $\mathcal{B}_1$-basis vectors

$$\begin{aligned} \mathbf{v_1} &= a_{11}\mathbf{u_1} + a_{21}\mathbf{u_2} \\ \mathbf{v_2} &= a_{12}\mathbf{u_1} + a_{22}\mathbf{u_2} \\ \mathbf{v_3} &= a_{13}\mathbf{u_1} + a_{23}\mathbf{u_2} \end{aligned}$$

Now if we consider the vector equation

$$\mathbf{0} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3} \tag{3.5.1}$$

it *should be* that the only solution is when each of the $x_i = 0$ (since the we claim the $\mathbf{v_i}$'s are linearly independent). Notice, however, that

$$\begin{aligned} \mathbf{0} &= x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3} \\ \mathbf{0} &= x_1(a_{11}\mathbf{u_1} + a_{21}\mathbf{u_2}) + x_2(a_{12}\mathbf{u_1} + a_{22}\mathbf{u_2}) + x_3(a_{13}\mathbf{u_1} + a_{23}\mathbf{u_2}) \\ \mathbf{0} &= (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)\mathbf{u_1} + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3)\mathbf{u_2} \end{aligned}$$

Since the $\mathbf{u_i}$'s are linearly independent, we must have that

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0 \end{cases}$$

which is a homogeneous linear system in 3 variables (and only 2 equations). From the Rank Theorem, it follows that there are infinitely many solutions $[x_1, x_2, x_3]^T$ and, in particular, there is a nonzero vector solution. This means there are nonzero coefficients x_1, x_2, x_3 satisfying Equation (3.5.1), hence $\mathcal{B}_2$ is a linearly dependent set and therefore is *not* a basis.

Theorem 3.5.14 (*Poole* Theorem 3.23 - The Basis Theorem)**.** *Let V be a vector space with two different bases $\mathcal{B}_1$ and $\mathcal{B}_2$. Then $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same number of vectors.*

Because the number of vectors in the basis is invariant of the choice of basis, we can define the following term.

Definition. The ***dimension*** of a vector space V is the number of vectors in a basis for V. We denote this $\dim(V)$.

Remark. The trivial vector space $\{\mathbf{0}\}$ is defined to have dimension 0.

Example 3.5.15. $\dim(\mathbb{R}^n) = n$.

Example 3.5.16. Let $V = \operatorname{Span}\left(\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right)$ be a subspace of $\mathbb{R}^3$. What is the dimension of V?

Clearly the first vector is a linear combination of the other two (which are linearly independent), so $\dim(V) = 2$.

Theorem 3.5.17 (*Poole* Theorem 3.24)**.** *For a matrix A,* $\dim(\operatorname{Row}(A)) = \dim(\operatorname{Col}(A))$.

Definition. The ***rank*** of a matrix A is the dimension of its column space (denoted $\operatorname{Rank}(A)$). If A has size $n \times n$ and $\operatorname{Rank}(A) = n$, then sometimes we say that A *has full rank*.

Remark. This new notion of rank still agrees with our old notion, because the number of linearly independent rows in A is the same as the number of nonzero rows in $\operatorname{RREF}(A)$ and $\dim(\operatorname{Row}(A)) = \dim(\operatorname{Col}(A))$.

Theorem 3.5.18 (*Poole* Theorem 3.25)**.** *For any matrix A,* $\operatorname{Rank}(A) = \operatorname{Rank}(A^T)$.

Proof. Since $\operatorname{Row}(A) = \operatorname{Col}(A^T)$, then $\operatorname{Rank}(A) = \dim(\operatorname{Row}(A)) = \dim(\operatorname{Col}(A^T)) = \operatorname{Rank}(A^T)$. ☐

Example 3.5.19. Show that $\mathcal{B} = \left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}2\\1\\-2\end{bmatrix}, \begin{bmatrix}3\\1\\-3\end{bmatrix}\right\}$ is a basis for $\mathbb{R}^3$

We consider the matrix $A = \begin{bmatrix}1 & 2 & 3\\0 & 1 & 1\\1 & -2 & -3\end{bmatrix}$ and compute its rank.

$$\begin{bmatrix}1 & 2 & 3\\0 & 1 & 1\\1 & -2 & -3\end{bmatrix} \xrightarrow{R_3 - R_1 \mapsto R_3} \begin{bmatrix}1 & 2 & 3\\0 & 1 & 1\\0 & -4 & -6\end{bmatrix}$$
$$\xrightarrow{R_3 + 4R_2 \mapsto R_3} \begin{bmatrix}1 & 2 & 3\\0 & 1 & 1\\0 & 0 & -2\end{bmatrix}$$

Thus $\operatorname{Rank}(A) = 3$. Hence the columns of A span a 3-dimensional subspace of $\mathbb{R}^3$, i.e. $\operatorname{Col}(A) = \mathbb{R}^3$. Since the columns of A are linearly independent and span $\mathbb{R}^3$, $\mathcal{B}$ is a basis for $\mathbb{R}^3$.

Exercise 3.5.3. Show that $\mathcal{B} = \left\{\begin{bmatrix}1\\-1\\1\end{bmatrix}, \begin{bmatrix}2\\0\\2\end{bmatrix}, \begin{bmatrix}3\\-7\\3\end{bmatrix}\right\}$ is not a basis for $\mathbb{R}^3$.

Definition. The ***nullity*** of a matrix A is the dimension of its null space. We denote it by $\operatorname{nullity}(A)$.

Example 3.5.20. Find the nullity of the matrix A constructed in example 3.5.19.

Row reducing, we have

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & -2 & -3 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the unique solution, and therefore $\text{nullity}(A) = \dim(\text{Null}(A)) = \dim(\{\mathbf{0}\}) = 0$.

Theorem 3.5.21 (Rank–Nullity)**.** *If A is an $m \times n$ matrix, then*

$$\text{Rank}(A) + \text{nullity}(A) = n.$$

Theorem 3.5.22 (Fundamental Theorem of Invertible Matrices, Pt II)**.** *Suppose A is an $n \times n$ matrix. The following are equivalent:*

a. A is invertible.

b. A is row equivalent to I_n (i.e. its reduced row echelon form is I_n).

c. A is the product of elementary matrices.

d. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

e. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

f. The columns of A are linearly independent.

g. The column vectors of A span $\mathbb{R}^n$.

h. The column vectors of A form a basis for $\mathbb{R}^n$.

i. The row vectors of A are linearly independent.

j. The row vectors of A span $\mathbb{R}^n$.

k. The row vectors of A form a basis for $\mathbb{R}^n$.

l. $\text{Rank}(A) = n$

m. $\text{nullity}(A) = 0$

3.5.3 Rank/Nullity of (Special Types of) Symmetric Matrices

Recall that, for any matrix A, we had that A^TA and AA^T were always symmetric. As you might have hoped, many of the properties of A^TA and AA^T are ultimately governed by the properties of A.

Theorem 3.5.23 (*Poole* Theorem 3.28)**.** *For any matrix A,* $\text{Null}(A) = \text{Null}(A^TA)$ *and* $\text{Null}(A^T) = \text{Null}(AA^T)$.

Proof. Suppose $\mathbf{x} \in \text{Null}(A)$. Then

$$A\mathbf{x} = \mathbf{0}$$
$$A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$$

so $\mathbf{x} \in \text{Null}(A^T A)$. Conversely, suppose $\mathbf{x} \in \text{Null}(A^T A)$. Then

$$\begin{aligned} A^T A\mathbf{x} &= \mathbf{0} \\ \mathbf{x}^T A^T A\mathbf{x} &= \mathbf{x}^T \mathbf{0} = [0] \\ (A\mathbf{x})^T (A\mathbf{x}) &= [0] \end{aligned}$$

Notice that $\mathbf{v}^T \mathbf{v} = [\mathbf{v} \cdot \mathbf{v}]$, so the above line implies that $A\mathbf{x} = \mathbf{0}$, hence $\mathbf{x} \in \text{Null}(A)$.

The proof of the second equality is the same *mutatis mutandis.* □

Using the fact that $\text{Rank}(A) = \text{Rank}(A^T)$ and the Rank-Nullity Theorem, we have the following immediate consequence

Corollary 3.5.24. *For any matrix A,* $\text{Rank}(A^T A) = \text{Rank}(A) = \text{Rank}(A^T) = \text{Rank}(AA^T)$.

Example 3.5.25. Let $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$; clearly $\text{Rank}(A) = 1$. But also

$$\begin{aligned} \text{Rank}(A^T A) &= \text{Rank}\left(\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\right) = 1 \\ \text{Rank}(AA^T) &= \text{Rank}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \text{Rank}\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) = 1. \end{aligned}$$

3.5.4 Coordinates

The following is a consequence of the Fundamental Theorem of Invertible Matrices, but we'll state it to be explicit

Theorem 3.5.26. *Let V be a vector space with an ordered basis $\mathcal{B} = \{\mathbf{v_1}, \dots, \mathbf{v_k}\}$. For every vector $\mathbf{u} \in V$, there is a unique linear combination of $\mathcal{B}$-basis vectors such that*

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k.$$

Definition. The c_i in the previous theorem are called the ***coordinates of u with respect to $\mathcal{B}$*** and the column vector

$$[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

is called the ***coordinate vector of u with respect to $\mathcal{B}$***.

Example 3.5.27. Let $P(0,0)$ and $Q(3,1)$ be points in the plane and consider the vector $\mathbf{v} = \overrightarrow{PQ}$. Given the standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathbb{R}^2$, we can write

$$\mathbf{v} = 3\mathbf{e_1} + 1\mathbf{e_2}$$

hence

$$[\mathbf{v}]_{\mathcal{E}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

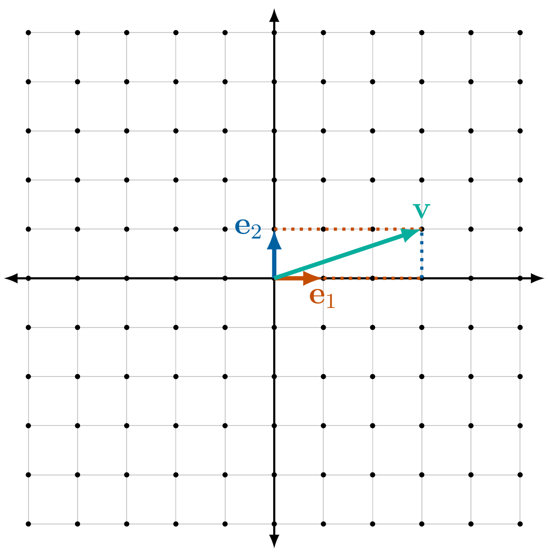
Example 3.5.28. With $P, Q, \mathbf{v}$ as above, we consider now the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ for $\mathbb{R}^2$. Since

$$\mathbf{v} = 2\mathbf{b}_1 + 1\mathbf{b}_1$$

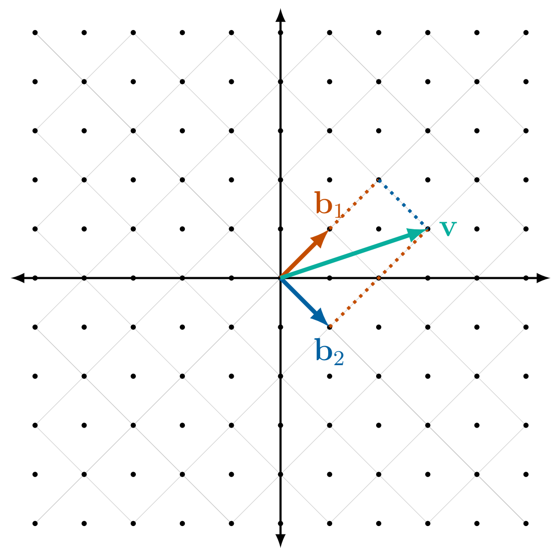
then

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Visually, the previous two examples just give a bit of formality to the hand-wavy "coordinate grid" discussion from Section 1.1



$\mathbf{v}$ in the $\mathcal{E}$-basis.



$\mathbf{v}$ in the $\mathcal{B}$-basis.

Remark. We typically don't write the subscript $\mathcal{E}$ for vectors when they are written in the standard basis.

Coordinates are actually very important in practice. On Earth, for example, exactly what does "1 unit in the x-direction" even mean? 1 meter north? 1 mile east? There's no universal agreement, so all measurements are really only relative. (In physics terms, we might say that these choices constitute a *frame of reference*.) It will be important that we figure out how to convert between these coordinate frames.

6.3 Change of Basis

Let $\mathcal{E} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be the standard basis for $\mathbb{R}^2$ and let $\mathcal{B} = \left\{ \mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ another basis for $\mathbb{R}^2$. Let $\mathbf{v}$ be a vector in $\mathbb{R}^2$ emanating from the origin to the point (x, y). By thinking in terms of coordinate grids, is straightforward to see that

$$\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 \quad \Longrightarrow \quad [\mathbf{v}]_{\mathcal{E}} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\mathbf{v} = \left(\frac{x+y}{2} \right) \mathbf{b}_1 + \left(\frac{x-y}{2} \right) \mathbf{b}_2 \quad \Longrightarrow \quad [\mathbf{v}]_{\lfloor} = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}$$

We could always figure out how to convert between $[\mathbf{v}]_{\mathcal{E}}$ or $[\mathbf{v}]_{\mathcal{B}}$ by just solving for a linear combination, but that can be cumbersome. Instead, notice that

$$\mathbf{b}_1 = 1\mathbf{e}_1 + 1\mathbf{e}_2 \quad \Longrightarrow \quad [\mathbf{b}_1]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{6.3.1}$$

$$\mathbf{b}_2 = 1\mathbf{e}_1 - 1\mathbf{e}_2 \quad \Longrightarrow \quad [\mathbf{b}_2]_{\mathcal{E}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{6.3.2}$$

We then have that

$$\begin{bmatrix} [\mathbf{b}_1]_{\mathcal{E}} & [\mathbf{b}_2]_{\mathcal{E}} \end{bmatrix} [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = [\mathbf{v}]_{\mathcal{E}}$$

And so we have that the matrix

$$\underset{\mathcal{E} \leftarrow \mathcal{B}}{P} := \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{E}} & [\mathbf{b}_2]_{\mathcal{E}} \end{bmatrix}$$

has the feature that it converts vectors from the $\mathcal{B}$-basis to the $\mathcal{E}$-basis.

Example 6.3.1. With $\mathcal{E}$ and $\mathcal{B}$ above, find the coordinate representation of $\mathbf{v} = [3, 1]^T$ in both bases, and verify that $\underset{\mathcal{B} \leftarrow \mathcal{E}}{P}$ converts these representations accordingly.

One can easily verify that

$$[\mathbf{v}]_{\mathcal{E}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \text{and} \qquad [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and from Equation **??**, one can also readily see that

$$[\mathbf{b}_1]_{\mathcal{E}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \text{and} \qquad [\mathbf{b}_2]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} [\mathbf{b}_1]_{\mathcal{E}} & [\mathbf{b}_2]_{\mathcal{E}} \end{bmatrix} [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = [\mathbf{v}]_{\mathcal{E}}.$$

Our choice to write the matrix columns in the $\mathcal{E}$-basis was just for convenience. If we had two non-standard bases, we could also do the same thing, hence we take the following definition

Definition. Let $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ and $\mathcal{C} = \{\mathbf{c_1}, \ldots, \mathbf{c_n}\}$ be two ordered bases for $\mathbb{R}^n$. The $n \times n$ matrix

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} | & | & & | \\ [\mathbf{b_1}]_{\mathcal{C}} & [\mathbf{b_2}]_{\mathcal{C}} & \cdots & [\mathbf{b_n}]_{\mathcal{C}} \\ | & | & & | \end{bmatrix}.$$

is called the ***change-of-basis matrix*** from $\mathcal{B}$ to $\mathcal{C}$. It has the effect

$$\left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{C}}.$$

Example 6.3.2. Let $\mathcal{B} = \left\{\mathbf{b_1} = \begin{bmatrix}1\\1\end{bmatrix}, \mathbf{b_2} = \begin{bmatrix}1\\-1\end{bmatrix}\right\}$ and $\mathcal{C} = \left\{\mathbf{c_1} = \begin{bmatrix}1\\-2\end{bmatrix}, \mathbf{c_2} = \begin{bmatrix}-3\\3\end{bmatrix}\right\}$ be bases for $\mathbb{R}^2$. Compute the change of basis matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$.

We need to find $[\mathbf{b_1}]_{\mathcal{C}}$ and $[\mathbf{b_2}]_{\mathcal{C}}$.

$$\mathbf{b_1} = -2\mathbf{c_2} - \mathbf{c_1} \quad \Longrightarrow \quad [\mathbf{b_1}]_{\mathcal{C}} = \begin{bmatrix}-2\\-1\end{bmatrix}$$

$$\mathbf{b_2} = 0\mathbf{c_2} - \frac{1}{3}\mathbf{c_2} \quad \Longrightarrow \quad [\mathbf{b_2}]_{\mathcal{C}} = \begin{bmatrix}0\\-\frac{1}{3}\end{bmatrix}$$

so

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\mathbf{b_1}]_{\mathcal{C}} \quad [\mathbf{b_2}]_{\mathcal{C}}] = \begin{bmatrix}-2 & 0\\-1 & -\frac{1}{3}\end{bmatrix}.$$

Proposition 6.3.3. *Given two different bases for $\mathbb{R}^n$, $\mathcal{B}$ and $\mathcal{C}$, the following are true*

- $\left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$, *and*
- $\left(\underset{\mathcal{C}\leftarrow\mathcal{E}}{P}\right)\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right) = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$,

where $\mathcal{E}$ is the standard basis.

The first feature should be obvious - since the columns of the change of basis matrix are a basis, the matrix is invertible. Hence

$$[\mathbf{v}]_{\mathcal{C}} = \left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)[\mathbf{v}]_{\mathcal{B}} \quad \Longrightarrow \quad \left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1}[\mathbf{v}]_{\mathcal{C}} = [\mathbf{v}]_{\mathcal{B}}$$

The composition seems obvious if you think about the notation as representing a function, but we'll use an example to demonstrate it.

Example 6.3.4. Using the bases from the previous example, we already computed $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. It is straightforward to see that

$$\underset{\mathcal{E}\leftarrow\mathcal{B}}{P} = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} \quad \text{and} \quad \underset{\mathcal{E}\leftarrow\mathcal{C}}{P} = \begin{bmatrix}1 & -3\\-2 & 3\end{bmatrix}.$$

Then

$$\begin{aligned}
\left(\underset{\mathcal{C}\leftarrow\mathcal{E}}{P}\right)\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right) &= \left(\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}\right)^{-1}\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right) \\
&= \begin{bmatrix} 1 & -3 \\ -2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\
&= -\frac{1}{3}\begin{bmatrix} 3 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\
&= -\frac{1}{3}\begin{bmatrix} 6 & 0 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ -1 & -\frac{1}{3} \end{bmatrix} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}.
\end{aligned}$$

Remark. It is extremely fast to find change-of-basis matrices to the standard basis, and inversion is also a fairly quick operation, so as shown in the previous example, these two facts above make it very quick to find a change-of-basis matrix between arbitrary bases.

3.6 Introduction to Linear Transformations

Definition. A ***transformation*** (aka ***function*** or ***map***) is a function T with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$, written

$$T : \mathbb{R}^n \to \mathbb{R}^m.$$

T is a ***linear transformation*** if

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and
2. $T(k\mathbf{v}) = kT(\mathbf{v})$ for all scalars $k \in \mathbb{R}$ and vectors $\mathbf{v} \in \mathbb{R}^n$.

Example 3.6.1 (*Identity transformation*). $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{v}) = \mathbf{v}$ is a linear transformation.

Example 3.6.2 (*Trivial transformation*). $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{v}) = \mathbf{0}$ is a linear transformation.

Example 3.6.3. Suppose $T : \mathbb{R}^1 \to \mathbb{R}^1$ is a linear transformation. Thinking about it as a function from $\mathbb{R}$ to $\mathbb{R}$, we can try to compute its derivative. Indeed,

$$\lim_{h\to 0} \frac{T(x+h) - T(x)}{h} = \lim_{h\to 0} \frac{T(x) + hT(1) - T(x)}{h} = T(1)$$

so T is differentiable and has a constant derivative. We thus know that $T(x) = mx + b$, for some real numbers m, b. Then

1. $T(x_1 + x_2) = mx_1 + mx_2 + b = (mx_1 + b) + (mx_2 + b) = T(x_1) + T(x_2)$ (precisely when $b = 0$), and
2. $T(kx) = mkx + b = k(mx + b) = kT(x)$ (precisely when $b = 0$)

so T is a linear transformation only when $b = 0$.

Remark. In general, the above argument shows that every component in $T(\mathbf{v}) = T([v_1, \ldots, v_n])$ looks like a linear combination of the v_i's.

Theorem 3.6.4. *If $A \in \mathbb{R}^{m\times n}$, then the transformation*

$$T_A : \mathbb{R}^n \to \mathbb{R}^m$$
$$T_A(\mathbf{x}) = A\mathbf{x}$$

is a linear transformation.

Proof. This follows quickly from the properties of matrix operations:

1. $T_A(\mathbf{x_1} + \mathbf{x_2}) = A(\mathbf{x_1} + \mathbf{x_2}) = A\mathbf{x_1} + A\mathbf{x_2} = T_A(\mathbf{x_1}) + T_A(\mathbf{x_1})$
2. $T_A(k\mathbf{x}) = A(k\mathbf{x}) = kA\mathbf{x} = kT_A(\mathbf{x})$

□

Example 3.6.5. Consider the map

$$T_A : \mathbb{R}^2 \to \mathbb{R}^3$$
$$T_A(\mathbf{x}) = A\mathbf{x}$$

where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$. Find the range and the kernel of T.

We can write

$$T_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}.$$

The range of T_A then is just

$$\begin{aligned} \text{Range}(T_A) &= \left\{ T_A\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \;:\; \text{where } x, y \in \mathbb{R} \right\} \\ &= \left\{ x \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + y \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \;:\; \text{where } x, y \in \mathbb{R} \right\} = \text{Col}(A). \end{aligned}$$

This tell us that the rank of A corresponds exactly to the dimension of Range(T). Similarly, the kernel of T is precisely the set of vectors in $\mathbb{R}^2$ for which

$$T(\mathbf{x}) = A\mathbf{x} = \mathbf{0}$$

hence $\ker(T) = \text{Null}(A)$.

Remark. The null space is a subspace of the *domain* of T.

As it turns out, we can write every linear transformation as multiplication by a matrix.

Theorem 3.6.6. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then we can write $T(\mathbf{x}) = A\mathbf{x}$ where A is the $m \times n$ matrix whose i^{th} column is the column vector $T(\mathbf{e_i})$, i.e.*

$$A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e_1}) & T(\mathbf{e_2}) & \cdots & T(\mathbf{e_m}) \\ | & | & & | \end{bmatrix}.$$

Definition. The matrix in the above theorem is called the ***standard matrix of T***. You may sometimes see $[T]$ used to denote the standard matrix of T.

Remark. This is the standard matrix because it's the transformation of the standard basis. One can always define the matrix of a linear transformation in terms of other bases as well.

Corollary 3.6.7. *Matrix multiplication corresponds to a composition of linear transformations.*

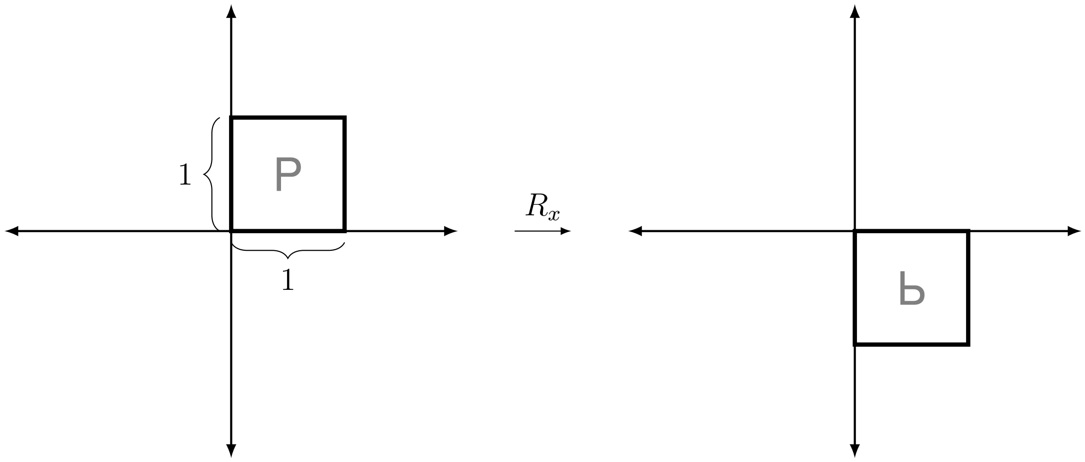
Proof. Let $T_A : \mathbb{R}^m \to \mathbb{R}^n$ and let $T_B : \mathbb{R}^n \to \mathbb{R}^p$ where A is $n \times m$ and B is $p \times n$. Then $T_B \circ T_A : \mathbb{R}^m \to \mathbb{R}^p$ and

$$T_B \circ T_A(\mathbf{x}) = T_B(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x} = T_{BA}(\mathbf{x}).$$

It follows that the standard matrix for $T_B \circ T_A$ is precisely the $p \times m$ matrix BA. □

3.6.1 Types of Linear Transformations of $\mathbb{R}^2$

Example 3.6.8. A ***reflection*** ***about the x-axis***, R_x, is a linear transformation of $\mathbb{R}^2$.



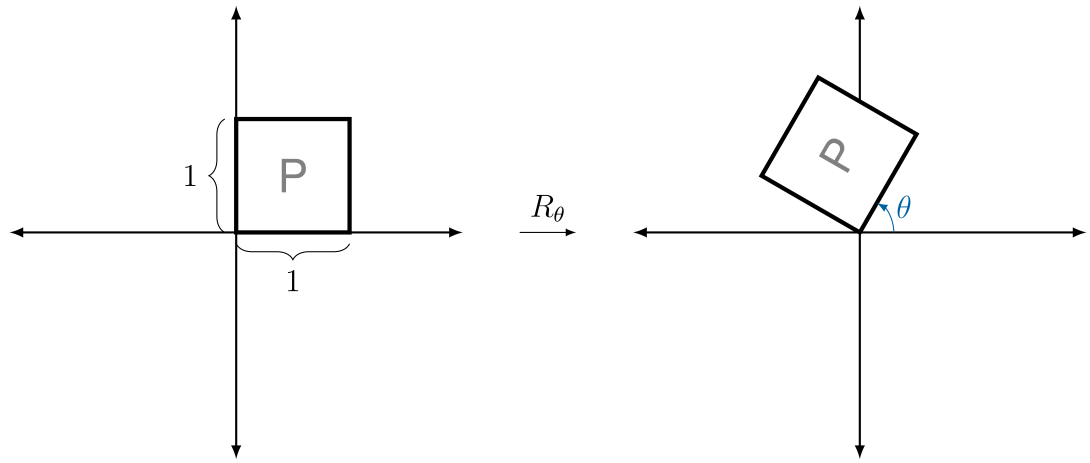
Explicitly, it sends a points (x, y) to $(x, -y)$, hence the transformation is given by

$$R_x\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ -y \end{bmatrix}$$

The standard matrix for R_x is thusly given by

$$[R_x] = \begin{bmatrix} R_x\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) & R_x\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Example 3.6.9. A *rotation* ***by an angle*** θ ***about the origin***, R_θ, is a linear transformation of $\mathbb{R}^2$.



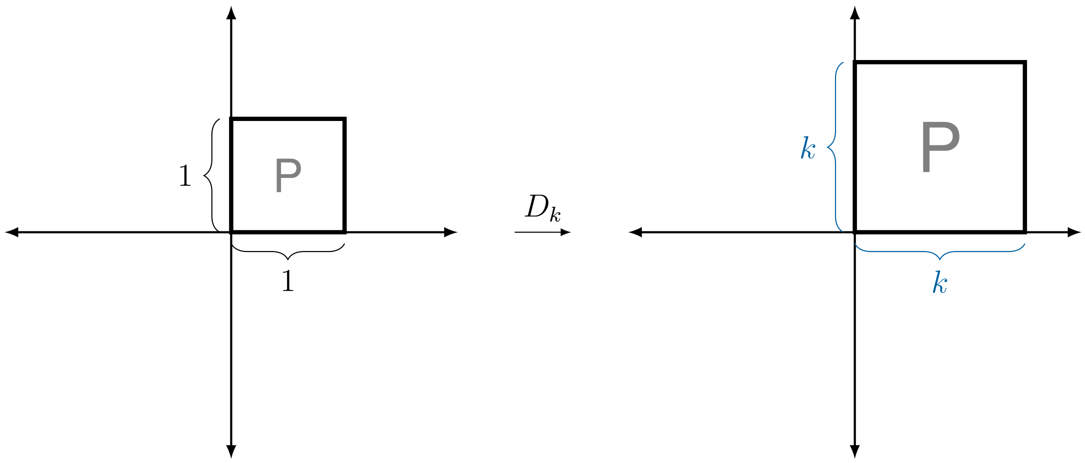
For this one, we'll first find the standard transformation matrix. Note that $\mathbf{e_1} = [\cos(0), \sin(0)]^T$ and $\mathbf{e_2} = [\cos(\frac{\pi}{2}), \sin(\frac{\pi}{2})]^T$. So rotation by an angle θ should add θ to the angle arguments of sine and cosine, i.e.

$$\begin{aligned}[R_\theta] &= \left[R_\theta\left(\begin{bmatrix} \cos(0) \\ \sin(0) \end{bmatrix} \right) \quad R_\theta\left(\begin{bmatrix} \cos(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) \end{bmatrix} \right) \right] \\ &= \begin{bmatrix} \cos(0+\theta) & \cos(\frac{\pi}{2}+\theta) \\ \sin(0+\theta) & \sin(\frac{\pi}{2}+\theta) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\end{aligned}$$

hence the linear transformation is given by

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

Example 3.6.10. A ***dilation*** (with ***dilation factor*** k) is a transformation D_k that expands out from the origin by a factor of k.



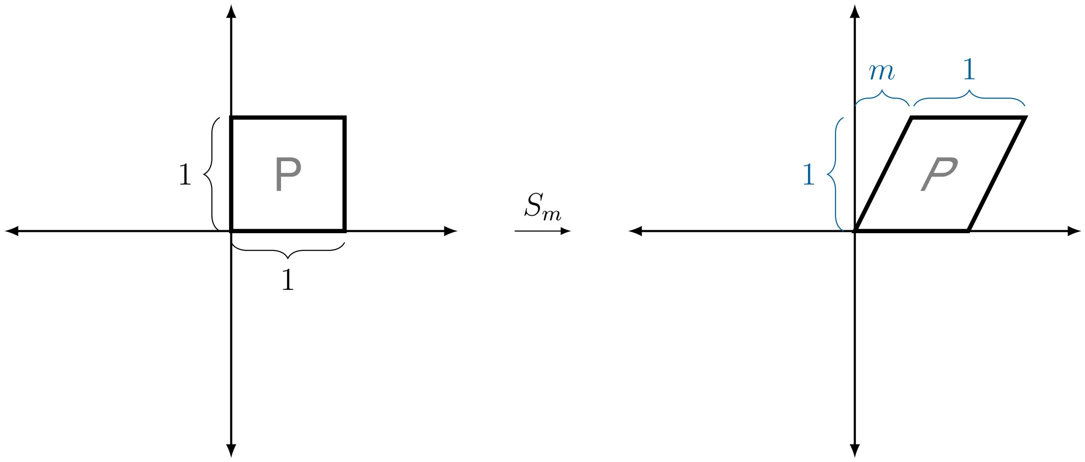
Explicitly, it sends a point (x, y) to a point (kx, y) so for vectors,

$$D_k\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} kx \\ ky \end{bmatrix}$$

The standard matrix for D_k is given by

$$[D_k] = \begin{bmatrix} D_k\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) & D_k\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}.$$

Example 3.6.11. A (***horizontal***) ***shear*** (with ***shear factor*** m), S_m, is a transformation that slides the top edge of the unit square m units to the right (making a parallelogram).



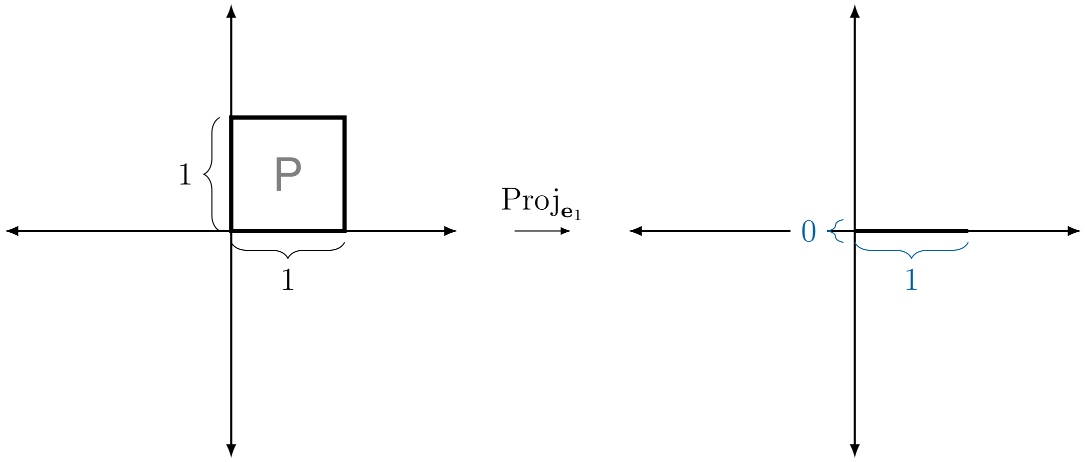
In particular, it sends (x, y) to the point $(x + my, y)$,

$$S_m\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + my \\ y \end{bmatrix}$$

The standard matrix for S_m is given by

$$[S_m] = \begin{bmatrix} S_m\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) & S_m\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$$

Example 3.6.12. A ***projection*** (onto the x-axis) is a transformation $\text{Proj}_{\mathbf{e}_1}$ that sends the vector $[x, y]$ to the vector $[x, 0]$.



Since

$$\text{Proj}_{\mathbf{e}_1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ 0 \end{bmatrix}$$

The standard matrix for $\text{Proj}_{\mathbf{e}_1}$ is given by

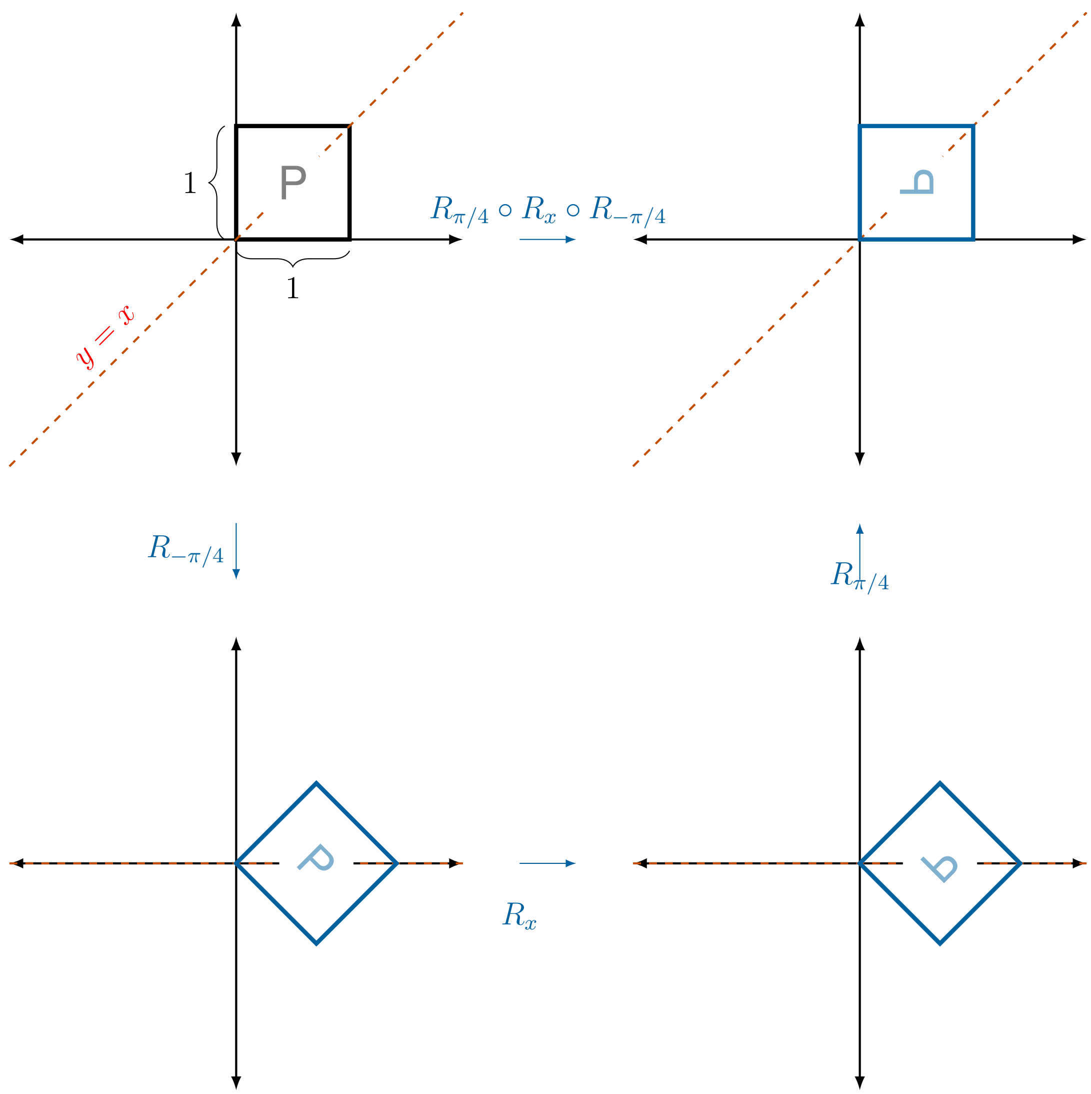
$$[\text{Proj}_{\mathbf{e}_1}] = \left[\text{Proj}_{\mathbf{e}_1}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \quad \text{Proj}_{\mathbf{e}_1}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Example 3.6.13. Find the matrix corresponding to a reflection of $\mathbb{R}^2$ across the line $y = x$.

A reflection across the line $y = x$ sends $\begin{bmatrix}1\\0\end{bmatrix}$ to $\begin{bmatrix}0\\1\end{bmatrix}$ and vice versa, so the standard matrix for this transformation is

$$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

Alternatively, notice that we can think about a reflection about this line as a composition of the following moves: first rotate the line $y = x$ to $y = 0$, reflect across this line, then rotate $y = 0$ back to $y = x$.



We have that

$$[R_{-\pi/4}] = \begin{bmatrix}\cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right)\\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right)\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\-1 & 1\end{bmatrix}$$

$$[R_x] = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$$

$$[R_{\pi/4}] = \begin{bmatrix}\cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right)\\ \sin\left(\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right)\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\1 & 1\end{bmatrix}$$

Now we have that

$$[R_{\pi/4} \circ R_x \circ R_{-\pi/4}] = [R_{\pi/4}][R_x][R_{-\pi/4}] = \frac{1}{2}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix} = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

Theorem 3.6.14. *Every linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ is a composition of the four five transformations described above, and the standard matrix can be obtained as a product of the corresponding matrices.*

Proof. Let $A = \begin{bmatrix}a & b\\ c & d\end{bmatrix}$ with $c \neq 0$. It is straightforward to verify that

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix} = \begin{bmatrix}1 & \frac{a}{c}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\frac{bc-ad}{c} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & d\\ 0 & 1\end{bmatrix}\begin{bmatrix}c & 0\\ 0 & 1\end{bmatrix}.$$

The case when $c = 0$ is left as an exercise to the reader. □

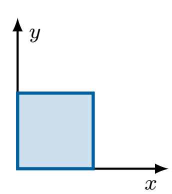
More generally,

3.6.2 One-to-One and Onto

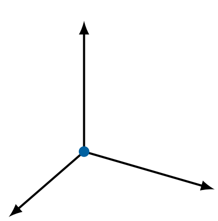
Definition. A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** (or ***injective***) if, for every vector $\mathbf{v} \in \mathbb{R}^n$, there is a unique $\mathbf{w} \in \mathbb{R}^m$ for which $T(\mathbf{v}) = \mathbf{w}$. T is ***onto*** (or ***surjective***) if, for every $\mathbf{w} \in \mathbb{R}^m$, there is at least one $\mathbf{v} \in \mathbb{R}^m$ for which $T(\mathbf{v}) = \mathbf{w}$. The ***kernel*** of T is the collection of vectors $\mathbf{x}$ in $\mathbb{R}^n$ for which $T(\mathbf{x}) = \mathbf{0}$.

For linear transformations, we have a more convenient way of thinking about these notions.

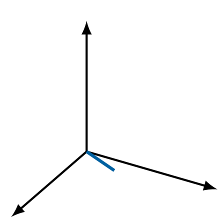
Example 3.6.15. Let's consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ with standard matrix A. Since $\mathrm{Col}(A) = \mathrm{range}(T)$, then $\mathrm{range}(T)$ is a subspace of $\mathbb{R}^3$ whose dimension is equal to $\mathrm{Rank}(A)$. Fundamentally, there are only three possible types of subspaces representing the range:



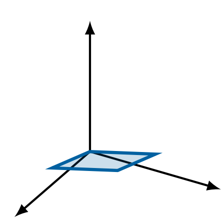
$\mathrm{domain}(T) = \mathbb{R}^2$



$\mathrm{range}(T) \cong \mathbb{R}^0$
$\mathrm{Rank}(A) = 0$



$\mathrm{range}(T) \cong \mathbb{R}^1$
$\mathrm{Rank}(A) = 1$



$\mathrm{range}(T) \cong \mathbb{R}^2$
$\mathrm{Rank}(A) = 2$

Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for $\mathbb{R}^2$, the domain of T. Note that $\{T(\mathbf{e}_1), T(\mathbf{e}_2)\}$ are precisely the columns of A.

Suppose that $\text{Rank}(A) = k$ where $k < 2$. Then $\{T(\mathbf{e}_1), T(\mathbf{e}_2)\}$ is not a linearly independent set in $\mathbb{R}^3$, so there are constants c_1, c_2 for which

$$\begin{aligned} c_1 T(\mathbf{e}_1) + c_2 T(\mathbf{e}_2) &= \mathbf{0} \\ c_2 T(\mathbf{e}_2) &= -c_1 T(\mathbf{e}_1) \\ T(c_2 \mathbf{e}_2) &= T(-c_1 \mathbf{e}_1) \end{aligned}$$

and hence we have two different vectors with the same output - T is *not* one-to-one.

This idea motivates the following

Theorem 3.6.16. *A linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if and only if* $\text{Rank}(A) = n$.

And following from Rank–Nullity,

Corollary 3.6.17. *A linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if and only if* $\text{nullity}(A) = 0$.

From example 3.6.15, it is clear that that $\{T(\mathbf{e}_1), T(\mathbf{e}_2)\}$, the columns of A, never form a basis for $\mathbb{R}^3$. Hence there is always a vector $\mathbf{v} \in \mathbb{R}^3$ for which $\mathbf{v} \notin \text{span}(T(\mathbf{e}_1), T(\mathbf{e}_2))$, and thus $\mathbf{v} \notin \text{range}(T)$. This motivates the following

Theorem 3.6.18. *A linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is onto if and only if* $\text{Rank}(A) = m$.

Example 3.6.19. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto the first coordinate:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ 0 \end{bmatrix} \quad \text{with standard matrix} \quad A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

It should be clear that any vector of the form $[0, y]^T$ is sent to $\mathbf{0}$, but this can be seen by explicitly computing

$$\text{Null}(A) = \left\{ \begin{bmatrix} 0 \\ y \end{bmatrix} \ : \ y \in \mathbb{R} \right\}.$$

By a quick rank/nullity computation, one also sees that T is neither one-to-one nor onto.

3.6.3 Inverse Transformations

Definition. Two transformations $S : \mathbb{R}^n \to \mathbb{R}^m$, and $T : \mathbb{R}^k \to \mathbb{R}^p$ are ***inverses*** if, for every $\mathbf{v} \in \mathbb{R}^n$,

$$S \circ T(\mathbf{v}) = T \circ S(\mathbf{v}) = \mathbf{v}$$

In other words, $S \circ T$ and $T \circ S$ are both the identity transformation.

Taking $\mathbf{v} \in \mathbb{R}^n$, then the only way the above equation holds is if $n = k = m = p$, and so the corresponding standard matrices $[S]$ and $[T]$ are actually $n \times n$. Since composition of linear transformations is equivalent to multiplication of the corresponding matrices, the equation in the definition can be rewritten as

$$[S][T] = [T][S] = I_n.$$

As such

Theorem 3.6.20. *If $T : \mathbb{R}^n \to \mathbb{R}^n$ has an inverse, then the standard matrix $[T]$ is invertible and $[T^{-1}] = [T]^{-1}$.*

Example 3.6.21. Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation of the plane by an angle of θ. Clearly, the inverse transformation is $R_{-\theta}$, the rotation of the plane by an angle of $-\theta$. Looking at the standard matrices

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad [R_{-\theta}] = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

and the product of these matrices is

$$[R_\theta][R_{-\theta}] = \begin{bmatrix} \cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \cos\theta\sin\theta \\ \cos\theta\sin\theta - \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Theorem 3.6.22. *A is invertible if and only if $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is both one-to-one and onto.*

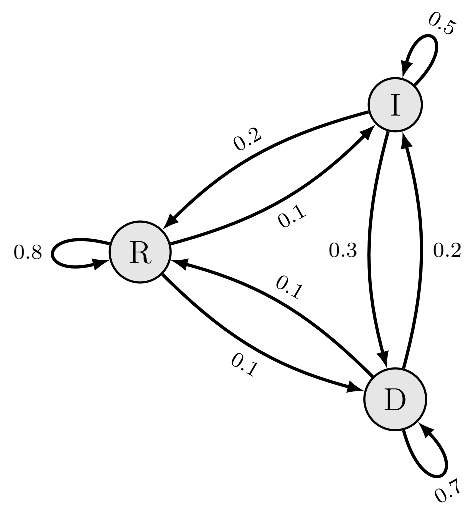
Proof. By the fundamental theorem of invertible matrices, A is invertible if and only if $\text{Rank}(A) = n$, and from Theorems 3.6.16 and 3.6.18, this is true if and only if $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is both one-to-one and onto. □

3.7 Applications

3.7.1 Markov Chains

A Markov chain is just a process by which one models probablistic scenarios

Example 3.7.1. Researchers have found that Democratic (D) voters are 70% likely to continue voting Democratic in the next election, 10% likely to vote Republican in the next election, and 20% likely to vote Independent in the next election. Similar data was compiled for Republican (R) and Independent (I) voters, and can be modeled in the following graph:



If there are D_0 Democratic voters, R_0 Republican voters, and I_0 Independent voters in this current election cycle, how many of each will there be for the next election cycle? How many will there be after k election cycles?

We can write

$$\begin{aligned} D_1 &= 0.7(D_0) + 0.1(R_0) + 0.3(I_0) \\ R_1 &= 0.1(D_0) + 0.8(R_0) + 0.2(I_0) \\ I_1 &= 0.2(D_0) + 0.1(R_0) + 0.5(I_0) \end{aligned}$$

or, as a matrix/vectors

$$\mathbf{x_1} = \begin{bmatrix} D_1 \\ R_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.1 & 0.8 & 0.2 \\ 0.2 & 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} D_0 \\ R_0 \\ I_0 \end{bmatrix} = P\mathbf{x_0}$$

It follows that, after k elections cycles, $\mathbf{x_k} = P^k\mathbf{x_0}$.

Definition. The vector $\mathbf{x_0}$ is known as the ***initial state vector***, P is known as the ***transition matrix***, and the entire process above is called a ***Markov chain*** (with 3 ***states***). If the transition matrix has columns with nonnegative entries that all sum to 1, it is called a **stochastic matrix**, the columns are called **probability vectors**, and the (i, j)-entry of P^k represents the probability that object in state i (e.g. voting preference) switches to state j after k transitions. If a stochastic matrix P has the property that some power of it P^n has all positive (not just nonnegative) entries, then P is a **regular stochastic matrix**.

Example 3.7.2. Given the process described in the previous Example 3.7.1, let $D_0 = 1000$, $R_0 = 800$ and $I_0 = 300$. Compute $\mathbf{x}_k$ for a few values of k. What happens as $k \to 0\infty$?

$$\mathbf{x_1} = \begin{bmatrix} 870 \\ 800 \\ 430 \end{bmatrix}, \qquad \mathbf{x_{20}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}, \qquad \mathbf{x_{100}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}, \quad \mathbf{x_{1000}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}$$

What we notice is that the vector $\mathbf{x_k} = P^k \mathbf{x_0}$ seems to stop changing as $k \to \infty$. Hence

$$\mathbf{x} = \lim_{k \to \infty} P^k \mathbf{x}_0 \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}$$

If you're interested in playing around with this yourself, say with different initial conditions or a different number of steps in the Markov chain process, you can use the MATLAB code below:

```
transMat = [ 0.7 0.1 0.3 ; 0.1 0.8 0.2 ; 0.2 0.1 0.5 ]; %transition matrix
x0 = [ 1000 ; 800 ; 300 ]; %initial state vector [D0,R0,I0]
maxLoop = 100; %number of iterations in Markov chain

for k=1:maxLoop
  transpose(transMat^k*x0) %outputs xk = [Dk,Rk,Ik], kth step in Markov chain process
end
```

Now, if $\lim_{k \to \infty} P^k \mathbf{x_0}$ exists and equals some vector $\mathbf{x}$, then

$$\mathbf{x} = \lim_{k \to \infty} P^k \mathbf{x_0} = \lim_{k \to \infty} P^{k+1} \mathbf{x_0} = P \left(\lim_{k \to \infty} P^k \mathbf{x_0} \right) = P\mathbf{x}.$$

Definition. In a Markov chain process with stochastic transition matrix P, a vector $\mathbf{x}$ for which $P\mathbf{x} = \mathbf{x}$ is called a **steady state vector**.

Solving for steady-state vectors is quite a bit simpler than looking at limits. Note that

$$\begin{aligned} P\mathbf{x} &= \mathbf{x} \\ P\mathbf{x} &= I\mathbf{x} \\ P\mathbf{x} - I\mathbf{x} &= \mathbf{0} \\ (P - I)\mathbf{x} &= \mathbf{0} \end{aligned}$$

which is a fairly simple linear system that can be solved in the usual way. The problem you'll run into is that this system is under-determined and there will be infinitely-many choices for $\mathbf{x}$. One natural choice is to require $\mathbf{x}$ to also be a probability vector, and this is natural because

Theorem 3.7.3. *For a regular stochastic matrix P, there is a unique steady state probability vector.*

Remark. We note that, if $\mathbf{x}_0$ is a probability vector, then $P^k\mathbf{x}_0$ is as well, so it's also natural to set up your Markov chain with the initial state vector as a probability vector.

Example 3.7.4. Find a steady-state probability vector for the Markov chain in Example 3.7.1.

We can write $(P - I)\mathbf{x} = \mathbf{0}$ as an augmented matrix and solve it in the usual way.

$$\left[P - I \mid \mathbf{0}\right] = \left[\begin{array}{ccc|c} -0.3 & 0.1 & 0.3 & 0.0 \\ 0.1 & -0.2 & 0.2 & 0.0 \\ 0.2 & 0.1 & -0.5 & 0.0 \end{array}\right] \xrightarrow{RREF} \left[\begin{array}{ccc|c} 1.0 & 0.0 & -1.6 & 0.0 \\ 0.0 & 1.0 & -1.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{array}\right]$$

so our steady state vector has the form

$$\mathbf{x} = \begin{bmatrix} 1.6t \\ 1.8t \\ t \end{bmatrix}$$

for some real number t. Requiring $\mathbf{x}$ to be a probability vector gives $1.6t + 1.8t + t = 1$, hence $t = \frac{1}{4.4} \approx 0.227$, and thus

$$\mathbf{x} \approx \begin{bmatrix} 0.364 \\ 0.409 \\ 0.227 \end{bmatrix}$$

Remark. With the above in mind, taking scaling the probability vector by 2100, we obtain the original limit. Ultimately this tells us that, no matter nonzero vector $\mathbf{x}_0$ we picked, there will *always* be a scalar $\lambda \in \mathbb{R}$ for which $\lim_{k\to\infty} P^k\mathbf{x}_0 = \lambda\mathbf{x}$.

APPENDIX C Complex Numbers

Some matrices may have complex eigenvalues (what does that mean geometrically? hmmm...), so below is an list of important properties of complex numbers.

Let $i = \sqrt{-1}$, the so-called **imaginary unity**. A **complex number** is a number $z = a + bi$ where a and b are real numbers. a is called the **real part** and b is called the **imaginary part** of z. The **conjugate of z** is the complex number $\overline{z} = a - bi$.

Addition of two complex numbers z_1 and z_2 is done via by adding the real and imaginary parts separately:

$$z_1 + z_2 = (a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + i(b_1 + b_2).$$

Multiplication of complex numbers follows the usual distributive law:

$$\begin{aligned} z_1 z_2 &= (a_1 + b_1 i)(a_2 + b_2 i) \\ &= a_1 a_2 + a_1 b_2 i + a_2 b_2 i + b_1 b_2 i^2 \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) i. \end{aligned}$$

Noting that $z\overline{z}$ is a real number for any z, *division* of complex numbers is done by multiplying by the conjugate and scaling by $1/z\overline{z}$:

$$\frac{z_1}{z_2} = \frac{z_1 \overline{z_2}}{z_2 \overline{z_2}} = \frac{1}{z_2 \overline{z_2}} (z_1 \overline{z_2}) = \frac{1}{a_2^2 + b_2^2} ((a_1 a_2 + b_1 b_2) + (-a_1 b_2 + a_2 b_1) i)$$

Exercise. To each complex number $z = a + bi$ we can associate the matrix $[z] = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Show that the product $z_1 z_2$ can be derived from multiplying the associated matrices $[z_1][z_2]$.

Matrices with complex entries

Although we won't see them in this course, matrices with complex entries appear all the time (they may even be the central objects in your instructor's dissertation) and have uses in physics and electrical engineering.

If $A = [a_{ij}]$ is a matrix with complex entries, the **conjugate** is

$$\overline{A} = \begin{bmatrix} \overline{a_{11}} & \cdots & \overline{a_{1n}} \\ \vdots & & \vdots \\ \overline{a_{m1}} & \cdots & \overline{a_{mn}} \end{bmatrix}.$$

and the **conjugate transpose** (sometimes called the **adjoint** or

Hermitian transpose) is obtained by first conjugating the matrix and then transposing it

$$\overline{A}^T = \begin{bmatrix} \overline{a_{11}} & \cdots & \overline{a_{m1}} \\ \vdots & & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{mn}} \end{bmatrix}.$$

The conjugate-transpose is also sometimes denoted $A^\dagger$ or A^*, depending on whether or not a physicist wrote the paper.

4.1 Introduction to Eigenvalues and Eigenvectors

Definition. Let A be an $n \times n$ matrix. A scalar λ is an ***eigenvalue*** of A if there is a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ so that $A\mathbf{v} = \lambda \mathbf{v}$. Such a vector is called an ***eigenvector*** of A corresponding to λ.

Remark. The prefix *eigen–* is not a name, but is derived from German and means "special" or "characteristic."

Example 4.1.1. For a Markov chain (with regular transition matrix), the steady state vector was precisely the vector $\mathbf{x}$ satisfying $P\mathbf{x} = \mathbf{x}$, so $\mathbf{x}$ was an eigenvector of P corresponding to the eigenvalue 1.

Example 4.1.2. $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvector $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Find the corresponding eigenvalue.

$$A \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

hence the corresponding eigenvalue is $\lambda = -1$.

Example 4.1.3. Show that 3 is another eigenvalue of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and find its corresponding eigenvector.

We need to find a vector $\mathbf{v}$ such that

$$A\mathbf{v} = 3\mathbf{v} \qquad \Longrightarrow \qquad A\mathbf{v} - 3\mathbf{v} = \mathbf{0} \qquad \Longrightarrow \qquad (A - 3I)\mathbf{v} = \mathbf{0}$$

so really we need to compute $\text{Null}(A - I)$.

$$\left[A - 3I \mid \mathbf{0}\right] = \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

So a vector $\mathbf{v}$ is in $\text{Null}(A - 3I)$ if it has the form $\begin{bmatrix} t \\ t \end{bmatrix}$. As such, any nonzero vector of this form is an eigenvector of A corresponding to 3.

Definition. Let A be an $n \times n$ matrix and λ an eigenvalue with corresponding eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$. The ***eigenspace*** corresponding to λ is

$$E_\lambda := \text{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k}).$$

Remark. It may at first seem surprising that any linear combination of the above $\mathbf{v_i}$'s is still an eigenvector for λ, but it is a straightforward computation to see that it is true:

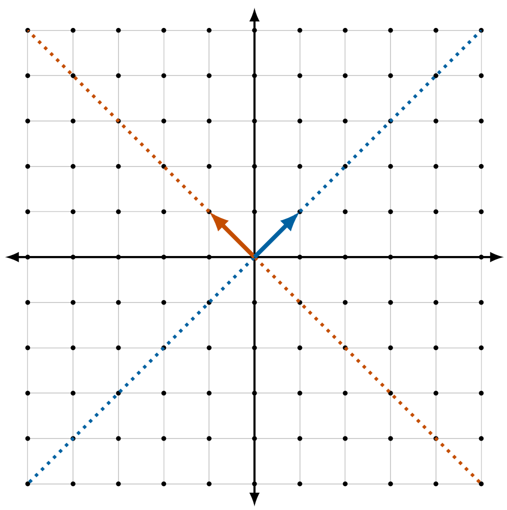
$$\begin{aligned} A(c_1\mathbf{v_1} + \cdots + c_k\mathbf{v_k}) &= c_1 A\mathbf{v_1} + \cdots + c_k A\mathbf{v_k} \\ &= c_1\lambda\mathbf{v_1} + \cdots + c_k\lambda\mathbf{v_k} \\ &= \lambda(c_1\mathbf{v_1} + \cdots + c_k\mathbf{v_k}). \end{aligned}$$

Example 4.1.4. In the previous example, $E_3 = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$.

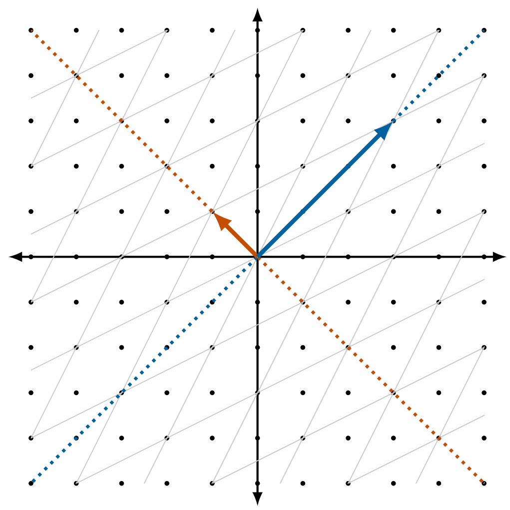
4.1.1 Geometry of Eigenvectors

Eigenvectors tell us about lines that are preserved or *stabilized* by the linear transformation (since all that happens is a vector in that line is scaled). Similarly, Eigenspaces correspond to subspaces that are stabilized by a linear transformation.

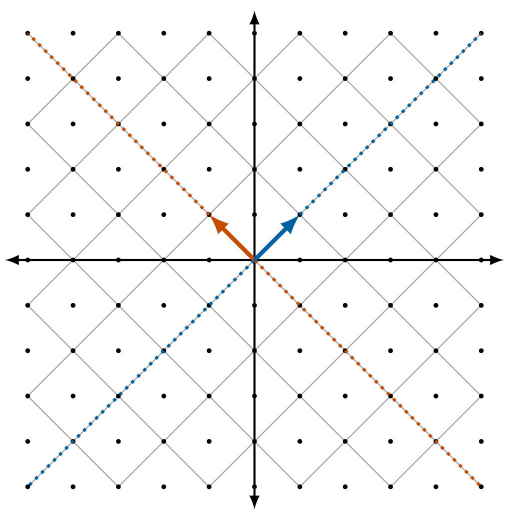
Example 4.1.5. Using $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ as before, notice that the vectors $[-1, 1]^T$ and $[1, 1]^T$ do not change direction after a transformation, but are merely scaled.



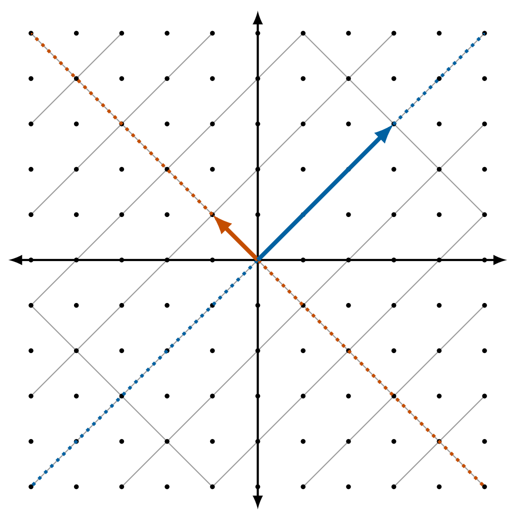
Before applying transformation A.
(Shown using the standard coordinate grid.)



After applying transformation A.
(Shown with the transformed standard coordinate grid.)



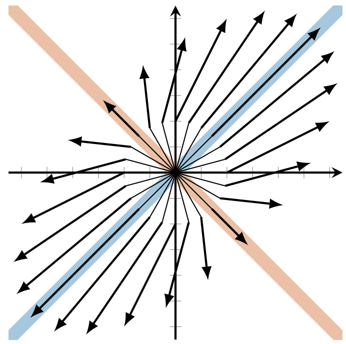
Before applying transformation A.
(Shown using the "*eigen*grid".)



After applying transformation A.
(Shown with the transformed "*eigen*grid".)

Another way to "see" eigenvectors is to plot the transformation "head-to-tail" by which you plot a vector $\mathbf{v}$ in standard position and then plot $A\mathbf{v}$ with its tail at the head of $\mathbf{v}$. Eigenvectors are then those vectors whose directions are unchanged and that are scaled by the eigenvalue.

Example 4.1.6. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ as before. Plotting a bunch of unit vectors $\mathbf{v}$ and then the corresponding vectors $A\mathbf{v}$ from the heads of $\mathbf{v}$, we see that the eigenvectors corresponding to $\lambda = 1$ are in the direction of $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and the eigenvectors corresponding to $\lambda = 3$ are in the direction of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.



4.1.2 Finding eigenvalues

Definition. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the ***determinant*** of A is

$$\det(A) = ad - bc$$

We'll define the determinants for all $n \times n$ matrices in the next section.

An eigenvector for the eigenvalue λ is a vector in $\mathrm{Null}(A - \lambda I)$. By the Fundamental Theorem of Invertible matrices tells us that $\mathrm{nullity}(A - \lambda I) \neq 0$ if and only if $A - \lambda I$ is *not* invertible. Moreover, from a previous theorem, a 2×2 matrix is *not* invertible if and only if the matrix $\det(A - \lambda I) = 0$. Since $\det(A - \lambda I)$ is a polynomial in λ, this means that eigenvalues are precisely the zeroes of this polynomial.

Example 4.1.7. Find the eigenvalues for $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

$$0 = \det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}\right) = (2-\lambda)(2-\lambda) - 1 = \lambda^2 - 4\lambda + 3.$$

Using your favorite method of solving for the zeroes of this polynomial, we exactly see that its zeroes are $\lambda = 1, 3$, which are precisely the eigenvalues we expected to get from the previous examples.

Example 4.1.8. Find the eigenvalues for $A = \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}$.

$$0 = \det(A - \lambda I) = \det\left(\begin{bmatrix} -6-\lambda & 3 \\ 4 & 5-\lambda \end{bmatrix}\right) = (-6-\lambda)(5-\lambda) - 12 = \lambda^2 + \lambda - 42$$

Using your favorite method of solving for the zeroes of this polynomial, we exactly see that its zeroes are $\lambda = -6, 7$.

Example 4.1.9. Find the eigenvalues and corresponding eigenspaces for the scalar matrix $A = \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}$.

$$0 = \det(A - \lambda I) = \det\left(\begin{bmatrix} 7-\lambda & 0 \\ 0 & 7-\lambda \end{bmatrix}\right) = (7-\lambda)(7-\lambda).$$

This matrix has a single eigenvalue $\lambda = 7$ and

$$E_7 = \text{Null}(A - 7I) = \text{Null}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

Remark. Eigenvalues can appear with multiplicity in the polynomial $\det(A - \lambda I)$, and this multiplicity may or may not be equal to the dimension of E_λ.

Exercise 4.1.1. Find the eigenvalues and corresponding eigenspaces for the horizontal shear matrix $A = \begin{bmatrix} 1 & 7 \\ 0 & 1 \end{bmatrix}$.

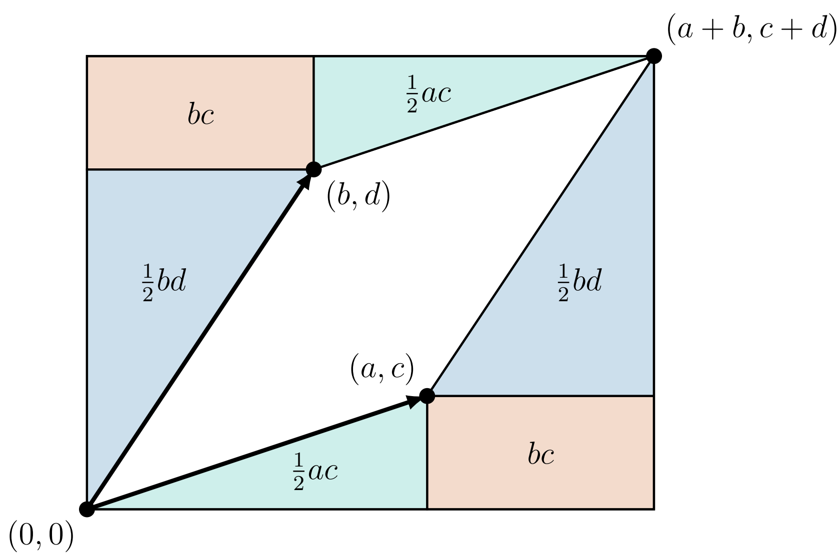
4.2 Determinants

What is the determinant of a matrix?

4.2.1 Determinant of a 2×2 Matrix

We motivate this by looking at where the determinant comes from in the 2×2 case.

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. For purposes of the picture, we will assume that $a > c > 0$ and $d > b > 0$.



The (signed) area of the parallelogram is

$$\text{area} = (a+b)(c+d) - 2bc - bd - ac = ad - bc. = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

Let's look at the properties of this determinant.

Proposition 4.2.1. $\det(I_2) = 1$

Proof. The area of the unit square is one. □

Proposition 4.2.2. $\det(A)$ *is* **multilinear** *(it behaves like a linear transformation in each column).*

Proof. Addition: This is a straightforward computation.

$$\begin{aligned} \det \begin{bmatrix} (a_1 + a_2) & b \\ (c_1 + c_2) & d \end{bmatrix} &= (a_1 + a_2)d - b(c_1 + c_2) \\ &= a_1 d - bc_1 + a_2 d - bc_2 \end{aligned}$$

$$= \det \begin{bmatrix} a_1 & b \\ c_1 & d \end{bmatrix} + \det \begin{bmatrix} a_2 & b \\ c_2 & d \end{bmatrix}$$

and $$\det \begin{bmatrix} a & (b_1 + b_2) \\ c & (d_1 + d_2) \end{bmatrix} = a(d_1 + d_2) - (b_1 + b_2)c$$
$$= ad_1 - b_1 c + ad_2 - b_2 c$$
$$= det \begin{bmatrix} a & b_1 \\ c & d_1 \end{bmatrix} + \det \begin{bmatrix} a & b_2 \\ c & d_2 \end{bmatrix}$$

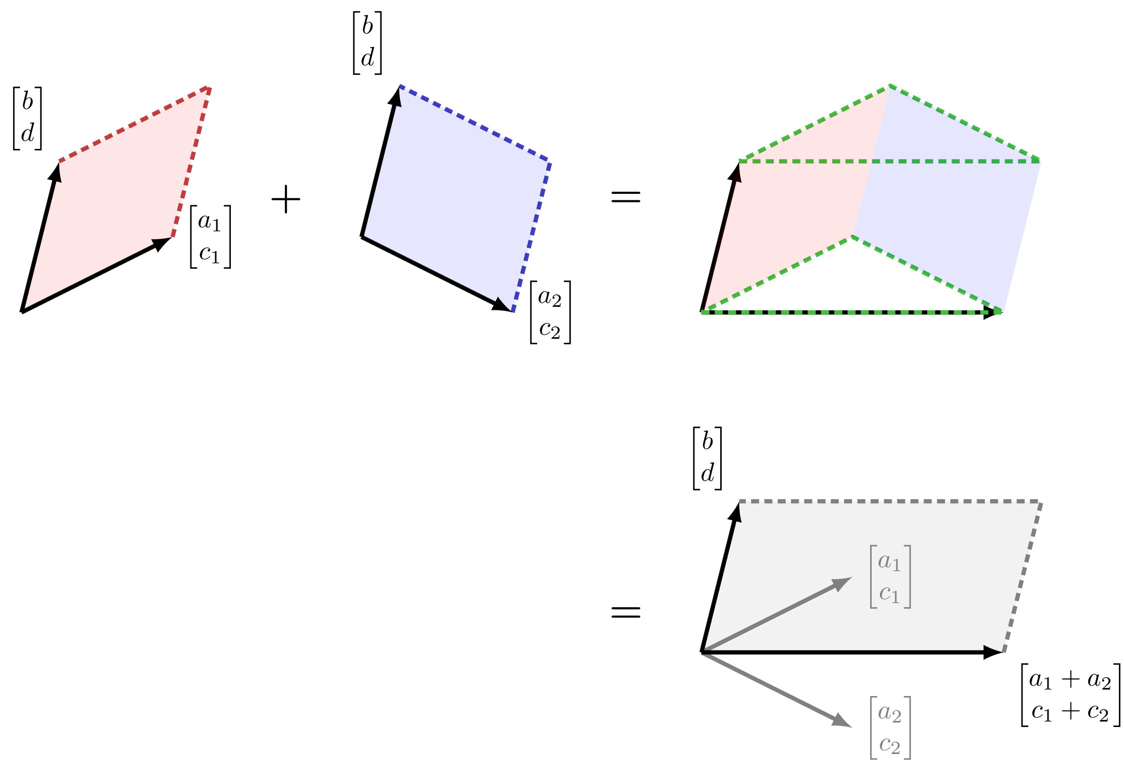
and that for any scalar k,

$$\det \begin{bmatrix} ka & b \\ kc & d \end{bmatrix} = (kad - bkc) = k(ad - bc) = k \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
and $$\det \begin{bmatrix} a & kb \\ c & kd \end{bmatrix} = (akd - kbc) = k(ad - bc) \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

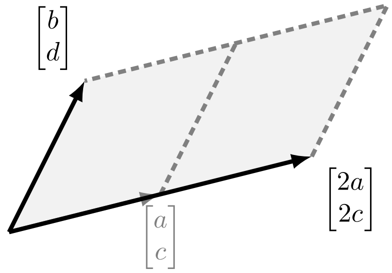
□

Here is an alternate, geometric proof of multilinearity for $\det(A)$ in the case of a 2×2 matrix.

Proof. <u>addition:</u> (drawn only in the case of the first column)



<u>scalar multiplication:</u> (drawn in the case that $k = 2$)



□

Proposition 4.2.3. *The determinant is* **alternating** *(that is, it switches sign whenever columns are swapped).*

Proof. Alternating: it is a straightforward computation to show that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = (ad - bc) = -(bc - ad) = -\det \begin{bmatrix} b & a \\ d & c \end{bmatrix}.$$

□

Remark. There is no geometric proof of the alternating condition; while it does have geometric meaning (it encodes *orientation*), this geometric meaning is inferred from the algebraic condition, so a geometric argument would be circular.

Now that we know the properties, we might define the determinant for an $n \times n$ matrix as follows, and then check to see if it still has all of those properties that the 2×2 case has.

Definition. The ***determinant*** of an $n \times n$ matrix A the signed volume of the parallelepiped formed by the columns of A.

Remark. Some write $|A|$ to mean $\det A$. We don't associate with such individuals.

Theorem 4.2.4. *Consider* det *as a function whose inputs are the columns of an $n \times n$ matrix, A, and whose output is* $\det(A)$*. This function has the following properties:*

1. $\det(I_n) = 1$.
2. det *is a multilinear function (it is a linear transformation in each input).*
3. det *is an alternating function (its sign changes whenever two columns are swapped).*

This is good news! This means that our geometric definition still has all of the same properties as in the 2×2 case. Better still, there's actually no other way we could have defined it:

Theorem 4.2.5. *The determinant is the unique function with the properties given in Theorem 4.2.4.*

4.2.2 Computing the determinant of a $n \times n$ Matrix

So how does one actually compute the determinant of an $n \times n$ matrix where $n > 2$? Computing the parallelepiped volume sounds like a bit of a chore, but thankfully we can do it iteratively.

Definition. For an $n \times n$ matrix A, the **(i,j)-minor** of A, denoted $M_{i,j}$, is the determinant of the ***submatrix*** formed by removing row i and column j from A.

Example 4.2.6. For $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, the $(2,1)$-minor is

$$M_{2,1} = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \det \begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} = a_{12}a_{33} - a_{32}a_{13}.$$

Remark. Your book uses the notation $\det A_{ij}$ to denote the (i,j)-minor, but I think it just makes things more confusing, especially since A_{ij} is common notation to represent the (i,j) entry of A.

Theorem 4.2.7 (Laplace's Theorem - cofactor expansion)**.** *The determinant of an $n \times n$ matrix A can be computed* along the i^{th} row *as the sum*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{i,j}.$$

or along the j^{th} column *as the sum*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{i,j}.$$

The quantity $(-1)^{i+j} M_{i,j}$ is sometimes called the **(i,j)-cofactor** *and the above sums are called* **cofactor expansions**.

Remark. The very rough geometric idea of cofactor expansion in dimension 3 is that the volume of the parallelepiped can be computed from the areas of three of the side parallelograms, and each of these areas can ultimately be viewed as the determinant of a 2×2 matrix. There are obviously a number of technical details to fill in here, but this construction extends up to the n-dimensional case.

Example 4.2.8. For $A = \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix}$, compute $\det A$ by expanding along the first row.

$$\begin{aligned} \det A &= a_{11}M_{1,1} - a_{12}M_{1,2} + a_{13}M_{1,3} \\ &= 3\det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} - 1\det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} + 4\det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} \\ &= 3\det \begin{bmatrix} 5 & -9 \\ 6 & 5 \end{bmatrix} - 1\det \begin{bmatrix} -1 & -9 \\ 2 & 5 \end{bmatrix} + 4\det \begin{bmatrix} -1 & 5 \\ 2 & 6 \end{bmatrix} \\ &= 3(79) - 1(13) + 4(-16) = 160. \end{aligned}$$

Example 4.2.9. With the same matrix as before, $A = \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix}$, compute $\det A$ by expanding along the second column.

$$\begin{aligned} \det A &= -a_{12}M_{1,2} + a_{22}M_{2,2} - a_{32}M_{3,2} \\ &= -1 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} + 5 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} - 6 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} \\ &= -1 \det \begin{bmatrix} -1 & -9 \\ 2 & 5 \end{bmatrix} + 5 \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} - 6 \det \begin{bmatrix} 3 & 4 \\ -1 & -9 \end{bmatrix} \\ &= -1(13) + 5(7) - 6(-23) = 160. \end{aligned}$$

Having a zero determinant tells us that the corresponding parallelepiped has no volume, which means all three column vectors must live in the same plane (or even same line). So a determinant of zero is enough to tell us that the columns are linearly dependent!

Proposition 4.2.10. *For an $n \times n$ matrix A, $\det(A) = 0$ if and only if the columns of A are linearly dependent.*

This means we can add onto the fundamental theorem of invertible matrices:

Theorem 4.2.11. *A is invertible if and only if $\det(A) \neq 0$.*

4.2.3 Properties of Determinants

Definition. A square matrix $A = [a_{ij}]$ is ***upper triangular*** if $a_{ij} = 0$ whenever $i > j$, and is ***lower triangular*** if $a_{ij} = 0$ whenever $i < j$.

Example 4.2.12. Consider the following matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

A is upper triangular, B is lower triangular, and C (a diagonal matrix) is both upper and lower triangular.

Theorem 4.2.13 (*Poole* Theorem 4.2)**.** *If A is an $n \times n$ triangular matrix, then $\det(A) = a_{11}a_{22} \cdots a_{nn}$, the product of the numbers along the diagonal.*

Proof. Perform the cofactor expansion along the first column (in the upper triangular case) or along the first row (in the lower triangular case). ☐

Example 4.2.14. Let's look at the matrices in Example 4.2.12. By computing the determinant of A along the first column, we have

$$\begin{aligned}\det A &= 1M_{1,1} - 0M_{2,1} + 0M_{3,1} \\ &= 1\det\begin{bmatrix}4 & 5\\0 & 6\end{bmatrix} = 1(4)(6).\end{aligned}$$

Now computing the determinant of B along the first row, we have

$$\begin{aligned}\det B &= 1M_{1,1} - 0M_{1,2} + 0M_{1,3} \\ &= 1\det\begin{bmatrix}3 & 0\\5 & 6\end{bmatrix} = 1(3)(6).\end{aligned}$$

Finally, computing the determinant of C along the first row (although the first column is perfectly fine as well),

$$\begin{aligned}\det C &= 1M_{1,1} - 0M_{1,2} + 0M_{1,3} \\ &= 1\det\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix} = 1(2)(3).\end{aligned}$$

Theorem 4.2.15 (*Poole* Theorems 4.7 - 4.10)**.** *If A and B are $n \times n$ matrices and k is a scalar, then*

1. $\det(AB) = (\det A)(\det B)$,
2. $\det(kA) = k^n(\det A)$,
3. $\det(A^T) = \det A$,
4. *and if A is invertible,* $\det A^{-1} = \dfrac{1}{\det A}$.

Example 4.2.16. Verify each part of the theorem above with the following:

$$A = \begin{bmatrix}3 & -1\\8 & -2\end{bmatrix}, \qquad B = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}, \qquad \text{and} \qquad k = 5.$$

1. $\det A = -6 + 8 = 2$, $\det B = 4 - 1 = 3$, and
$$\det(AB) = \det\begin{bmatrix}5 & 1\\14 & 4\end{bmatrix} = 6 = (2)(3) = (\det A)(\det B).$$
2. $\det kA = \det\begin{bmatrix}15 & -5\\40 & -10\end{bmatrix} = 50 = 25(2) = k^2 \det A.$
3. $\det A^T = \det\begin{bmatrix}3 & 8\\-1 & -2\end{bmatrix} = -6 + 8 = 2 = \det A$
4. $\det A^{-1} = \det\begin{bmatrix}-1 & \frac{1}{2}\\-4 & \frac{3}{2}\end{bmatrix} = -\dfrac{3}{2} + 2 = \dfrac{1}{2} = \dfrac{1}{\det A}$

4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices

We've already seen eigenvalues and eigenvectors for 2×2 matrices, but now that we have defined determinants for $n \times n$ matrices, we'll extend these definitions accordingly.

Definition. If A is a square matrix, then $\det(A - \lambda I)$ is a polynomial with indeterminate λ and is called the ***characteristic polynomial*** of A. The ***eigenvalues*** of A are precisely the roots of the characteristic polynomial. For each eigenvalue λ, the corresponding ***eigenspace*** is $E_\lambda = \text{Null}(A - \lambda I)$ and the nonzero vectors in E_λ are ***eigenvectors***.

Remark. Non-square matrices do not have eigenvalues/eigenvectors. If A is an $m \times n$ matrix and $\mathbf{v} \in \mathbb{R}^n$, then $A\mathbf{v} \in \mathbb{R}^m$. However, the equation $A\mathbf{v} = \lambda\mathbf{v}$ implies that $A\mathbf{v} \in \mathbb{R}^n$, so it must be that $m = n$. Non-square matrices have something called *singular values* which, in some sense, play the role of eigenvalues, but this is outside of the scope of this course.

Example 4.3.1. Find the eigenvalues and eigenvectors for $A = \begin{bmatrix} 2 & 12 & 10 \\ 0 & -4 & -4 \\ 1 & 2 & 1 \end{bmatrix}$.

We first compute $\det(A - \lambda I)$ via cofactor expansion along the first column.

$$\begin{aligned} p(\lambda) = \det(A - \lambda I) &= \det \begin{bmatrix} 2-\lambda & 12 & 10 \\ 0 & -4-\lambda & -4 \\ 1 & 2 & 1-\lambda \end{bmatrix} \\ &= (2-\lambda)\det\begin{bmatrix} -4-\lambda & -4 \\ 2 & 1-\lambda \end{bmatrix} + 1\det\begin{bmatrix} 12 & 10 \\ -4-\lambda & -4 \end{bmatrix} \\ &= (2-\lambda)\left((-4-\lambda)(1-\lambda)+8\right) + 1\left(-48 - 10(-4-\lambda)\right) \\ &= -\left(\lambda^3 + \lambda^2 - 12\lambda\right) \\ &= -\lambda(\lambda+4)(\lambda-3) \end{aligned}$$

The characteristic polynomial factors nicely and the eigenvalues are $-4, 0, 3$. The corresponding eigenspaces are

$$\begin{aligned} E_{-4} = \text{Null}(A + 4I) &= \text{Null}\left(\begin{bmatrix} 6 & 12 & 10 \\ 0 & 0 & -4 \\ 1 & 2 & 5 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}\right), \\ E_0 = \text{Null}(A - 0I) &= \text{Null}\left(\begin{bmatrix} 2 & 12 & 10 \\ 0 & -4 & -4 \\ 1 & 2 & 1 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right), \\ E_3 = \text{Null}(A - 3I) &= \text{Null}\left(\begin{bmatrix} -1 & 12 & 10 \\ 0 & -7 & -4 \\ 1 & 2 & -2 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 22 \\ -4 \\ 7 \end{bmatrix}\right). \end{aligned}$$

Example 4.3.2. Consider the matrix $A = \begin{bmatrix} 3 & -2 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. The characteristic polynomial for A is

$$p(\lambda) = (3-\lambda)(1-\lambda)^2$$

hence the eigenvalues are $1, 3$. The corresponding eigenspaces are

$$E_3 = \text{Null}(A - 3I) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right) \qquad \text{and} \qquad E_1 = \text{Null}(A - 1I) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right).$$

Notice that the above 3×3 matrix had only two distinct eigenvalues, but if we counted multiplicity (that is $(1 - \lambda)$ appears with multiplicity 2 in the characteristic polynomial), then in fact we have exactly 3 eigenvalues. It may also be interesting to notice that $\dim(E_1) = \dim(E_3) = 1$.

Example 4.3.3. Consider the matrix $A = \begin{bmatrix} 3 & -2 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. The characteristic polynomial for A is

$$p(\lambda) = (3-\lambda)(1-\lambda)^2$$

hence the eigenvalues are $1, 3$. The corresponding eigenspaces are

$$E_3 = \text{Null}(A - 3I) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right) \qquad \text{and} \qquad E_1 = \text{Null}(A - 1I) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right).$$

Just as before, the above 3×3 matrix had an eigenvalue 1 that appeared with multiplicity 2 in the characteristic polynomial, but this time $\dim(E_1) = 2$. It's not clear that there is a relationship between the multiplicity of the polynomial roots and the dimension, so let's give the following definitions:

Definition. The ***algebraic multiplicity*** of an eigenvalue λ is the multiplicity as a root of the characteristic polynomial, and the ***geometric multiplicity*** is the dimension of the eigenspace E_λ.

These two different notions of multiplicity will be important in the next section. We'll note that if A is an $n \times n$ matrix, then the sum of all of the algebraic multiplicities will always be n. This is a consequence of the fundamental theorem of algebra that every polynomial of degree n factors into n linear factors. From Example 4.3.2, it's clear that the geometric multiplicities do not need to sum to n, however.

Theorem 4.3.4 (*Poole* - Theorem 4.15)**.** *If A is a triangular (or diagonal) matrix, then the eigenvalues are precisely the entries appearing along the diagonal.*

Proof. If A is triangular, then by Theorem 4.2.13

$$p(\lambda) = \det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda).$$

□

Theorem 4.3.5. *If A is an $n \times n$ square matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily all distinct), then $\det A = \lambda_1\lambda_2\cdots\lambda_n$, the product of all of the eigenvalues.*

Proof. The above theorem is easy to see in the case that A is triangular - the product of the roots of a polynomial are precisely the constant term, and the constant term of the characteristic polynomial is exactly $a_{11}a_{22}\cdots a_{nn}$. More generally, the characteristic polynomial factors as

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$$

where each λ_i is an eigenvalue. Taking $\lambda = 0$, one gets

$$\det(A) = (\lambda_1)(\lambda_2)\cdots(\lambda_n)$$

□

From this it follows that we have yet another test for invertibility:

Theorem 4.3.6. *A square matrix A is invertible if and only if 0 is <u>not</u> an eigenvalue of A.*

4.3.1 Relationship to Matrix Operations

It is natural to ask about the interplay between eigenvalues/eigenvectors and matrix operations like inversion and exponentiation.

Example 4.3.7. Consider $A = \begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$. The characteristic polynomial is

$$p_A(\lambda) = (2 - \lambda)(3 - \lambda)$$

hence the eigenvalues are $2, 3$ and the corresponding eigenspaces are

$$E_2 = \text{Span}\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right) \qquad \text{and} \qquad E_3 = \text{Span}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right).$$

With A as above, we have that $A^{-1} = \frac{1}{6}\begin{bmatrix} 4 & 2 \\ -1 & 1 \end{bmatrix}$ and the eigenvalues are $\frac{1}{2}, \frac{1}{3}$ – reciprocals of A's eigenvalues. What's more, notice that

$$A^{-1}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -2 \\ 1 \end{bmatrix} \qquad \text{and} \qquad A^{-1}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

so the reciprocal eigenvalues of A^{-1} have the same eigenvectors as the eigenvalues of A! With A as before, we have that $A^2 = \begin{bmatrix} -1 & -10 \\ 5 & 14 \end{bmatrix}$ and the eigenvalues are $4, 9$ – squares of A's eigenvalues. What's more, notice that

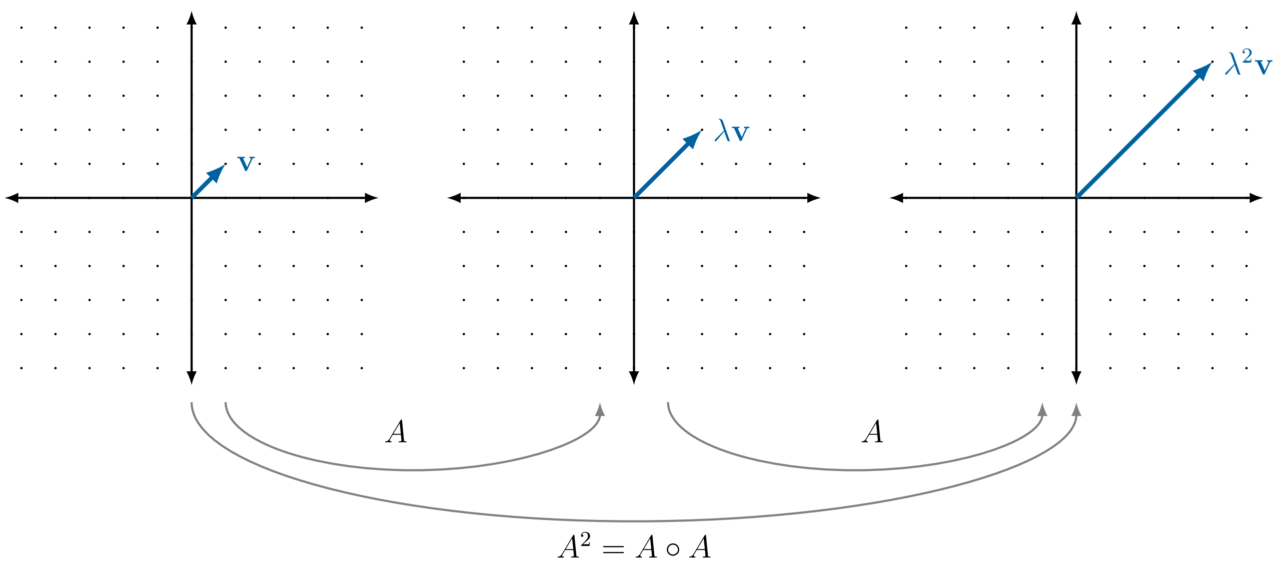
$$A^2\begin{bmatrix} -2 \\ 1 \end{bmatrix} = 2^2\begin{bmatrix} -2 \\ 1 \end{bmatrix} \qquad \text{and} \qquad A^2\begin{bmatrix} -1 \\ 1 \end{bmatrix} = 3^2\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

so the squared eigenvalues of A^2 have the same eigenvectors as the eigenvalues of A!

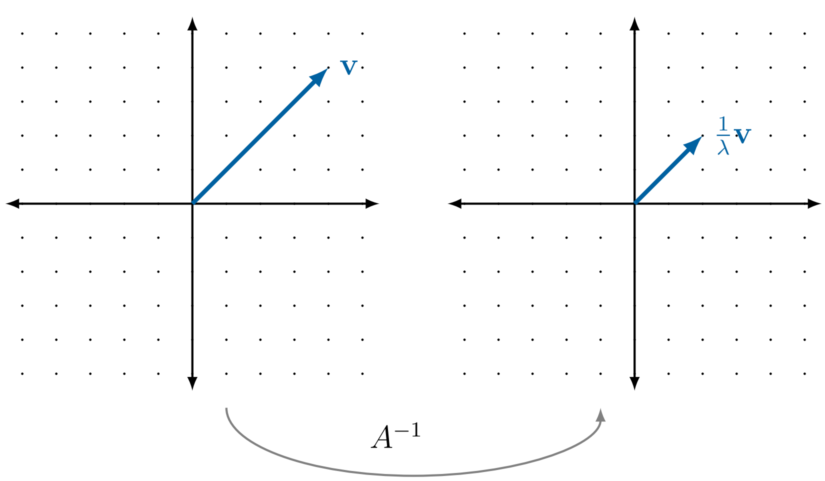
Theorem 4.3.8. *Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\mathbf{v}$.*

1. *For any positive integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{v}$.*
2. *If A is invertible, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} with corresponding eigenvector $\mathbf{v}$.*

The above theorem also makes sense geometrically. Each application of the transformation A stretches its eigenvector $\mathbf{v}$ by a factor of λ:



Similarly, each application of A^{-1} "undoes" the stretching of its eigenvector $\mathbf{v}$ by a factor of λ (i.e., stretches instead by a factor of $\frac{1}{\lambda}$: The above theorem also makes sense geometrically. Each application of A stretches its eigenvector by a factor of λ



4.4 Similarity and Diagonalization

4.4.1 Similarity

As we saw in the previous section, triangular and diagonal matrices were very nice from a computational standpoint, so it would be nice to convert a matrix into triangular form in a meaningful way. We already know that we can do this with row reduction, but this process does not preserve eigenvalues (any invertible matrix row reduces to the identity, for example), so in this section we will look at another process that does retain the useful *eigen*information.

Definition. Two $n \times n$ matrices A and B are called ***similar*** if there is an invertible $n \times n$ matrix P for which $P^{-1}AP = B$. We sometimes write "$A \sim B$" to mean "A is similar to B." We also sometimes refer to the product $P^{-1}AP$ as "conjugation of A by P."

Remark. Such a P is not unique. For example, $P^{-1}IP = I$ is true for every invertible matrix P.

Example 4.4.1. $A = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ are similar.

We aim to find $P = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ so that $P^{-1}AP = B$. So

$$\begin{aligned} P^{-1}AP &= B \\ \frac{1}{wx - yz}\begin{bmatrix} w & -y \\ -z & x \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} x & y \\ z & w \end{bmatrix} &= \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \\ \frac{1}{wx - yz}\begin{bmatrix} wx - 5yz & -4wy \\ 4xz & 5wx - yz \end{bmatrix} &= \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \end{aligned}$$

This matrix equation gives us four nonlinear equations in the entries, and playing with it a bit, one sees that $P = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$ is one such solution.

Example 4.4.2. The matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are not similar.

If they were, we could find a matrix $P = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ for which $B = P^{-1}AP$. In this case, we would have

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} &= \frac{1}{xw - yz}\begin{bmatrix} w & -y \\ -z & x \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x & y \\ z & w \end{bmatrix} \\ &= \frac{1}{xw - yz}\begin{bmatrix} wx + yz & 2wy \\ -2xz & -wx - yz \end{bmatrix} \end{aligned}$$

and this equality implies that the diagonal entries are both equal and nonzero. But $wx + yz = -(wx + yz)$ implies that $wx + yz = 0$, so this is impossible.

4.4.2 Properties of similarity and similar matrices

Theorem 4.4.3 (*Poole* Theorem 4.21)**.** *Let A, B, C be $n \times n$ matrices.*

a. reflexive*: $A \sim A$.*

b. symmetric*: If $A \sim B$ then $B \sim A$.*

c. transitive*: If $A \sim B$ and $B \sim C$ then $A \sim C$.*

Each of the following properties are easily verified, say with the matrices from Example 4.4.1.

Theorem 4.4.4 (*Poole* Theorem 4.22)**.** *Let A and B be similar $n \times n$ matrices. Then*

a. $\det A = \det B$.

b. A is invertible if and only if B is invertible.

c. A and B have the same rank.

d. A and B have the same characteristic polynomial.

e. A and B have the same eigenvalues.

f. $A^m \sim B^m$ for any positive integer m.

Proof (Sketch). Let P be a matrix for which $B = P^{-1}AP$.

a. $\det B = \det(P^{-1}AP) = (\det P^{-1})\,(\det A)\,(\det P) = \left(\frac{1}{\det P}\right)(\det A)\,(\det P) = \det A$

b. $B^{-1} = (P^{-1}AP)^{-1} = P^{-1}A^{-1}(P^{-1})^{-1} = P^{-1}A^{-1}P$

c. This follows from the fact that for any invertible matrix P,
$\text{Rank}(A) = \text{Rank}(PA) = \text{Rank}(AP)$.

d. $\det(B - \lambda I) = \det(P^{-1}AP - \lambda P^{-1}IP) = \det(P^{-1}(A - \lambda I)P) = \det(A - \lambda I)$

e. A and B have the same characteristic polynomials

f. $B^m = \underbrace{(P^{-1}AP)\cdots(P^{-1}AP)}_{m} = P^{-1}\underbrace{A\cdots A}_{m}P = P^{-1}A^mP$

□

To check whether two given matrices A and B are similar requires finding the matrix P satisfying $P^{-1}AP = B$, which as we saw from Examples 4.4.1 and 4.4.2, could be quite laborious. The above theorem is actually most useful for showing that two matrices are <u>not</u> similar (in fact, no single part of the theorem is enough to deduce that two matrices are similar).

Example 4.4.5. The matrices A and B in Example 4.4.2 are not similar because $\det A = -1$ and $\det B = 1$.

Example 4.4.6. Although matrices $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}$ have the same rank ($\text{Rank}\,A = \text{Rank}\,B = 2$) and determinant ($\det A = \det B = 6$), they are not similar because their eigenvalues and characteristic polynomials are different.

Example 4.4.7. Although matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ have the same rank (Rank A = Rank $B = 2$), determinant ($\det A = \det B = 6$), characteristic polynomial ($p(\lambda) = (1-\lambda)^2$), they are not similar because no non-identity matrix is similar to the identity matrix (for any P, $P^{-1}IP = I$).

4.4.3 Diagonalization

Definition. A matrix A is ***diagonalizable*** if it is similar to a diagonal matrix D, i.e. if there is some invertible matrix P so that $P^{-1}AP = D$.

Example 4.4.8. From Example 4.4.1, $B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ is diagonalizable since it is similar to the diagonal matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$.

Notice that the characteristic polynomial of A above is

$$\det(A - \lambda I) = (2-\lambda)(4-\lambda) - 3 = \lambda^2 - 6\lambda + 5 = (\lambda - 1)(\lambda - 5)$$

and thus B contains A's eigenvalues along the diagonal. This gives us a clue as to how one can go about finding the matrix P used to conjugate A into a diagonal matrix (if possible).

Theorem 4.4.9 (*Poole* Theorem 4.23)**.** *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.*

More precisely, $D = P^{-1}AP$ if and only if the columns of P are the eigenvectors of A and if the (i,i) entry of D is the eigenvalue corresponding to the i^{th} column of P.

I won't sketch the proof, but the core observation is the following:

If $P^{-1}AP = D$, then this rearranges to $AP = PD$. So if $\mathbf{p_i}$ is the i^{th} column of P and λ_i is the (i,i) entry in D, then

$$AP = PD$$

$$A \begin{bmatrix} \mathbf{p_1} & \cdots & \mathbf{p_n} \end{bmatrix} = \begin{bmatrix} \mathbf{p_1} & \cdots & \mathbf{p_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\begin{bmatrix} A\mathbf{p_1} & & A\mathbf{p_n} \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{p_1} & & \lambda_n\mathbf{p_n} \end{bmatrix}$$

and so $A\mathbf{p_i} = \lambda_i\mathbf{p_i}$, hence the λ_i are eigenvalues for A with corresponding eigenvectors $\mathbf{p_i}$.

Example 4.4.10. Let $A = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$. Is A diagonalizable?

The characteristic polynomial for A is

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 3 \\ 0 & 2-\lambda \end{bmatrix} = (2-\lambda)^2$$

so A has a single eigenvalue of 2 with algebraic multiplicity 2. The corresponding eigenspace is

$$E_2 = \text{Null}(A - 2I) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right).$$

and so the eigenvalue 2 has geometric multiplicity 1. This means that there are not enough linearly independent eigenvectors to form our invertible matrix P (the one for which $P^{-1}AP$ is a diagonal matrix), hence A is not diagonalizable.

Example 4.4.11. Let $A = \begin{bmatrix} 4 & -3 & -3 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix}$. is A diagonalizable?

The characteristic polynomial is

$$\det(A - \lambda I) = \det \begin{bmatrix} 4-\lambda & -3 & -3 \\ 3 & -2-\lambda & -3 \\ -1 & 1 & 2-\lambda \end{bmatrix} = -(\lambda - 1)^2(\lambda - 2)$$

and the eigenvalues are 1 and 2 (with algebraic multiplicities 2 and 1, respectively). The corresponding eigenspaces are

$$E_1 = \text{Null}(A - I) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) \qquad E_2 = \text{Null}(A - 2I) = \text{Span}\left(\begin{bmatrix} -3 \\ -3 \\ 1 \end{bmatrix} \right)$$

and so the eigenvalues 1 and 2 have geometric multiplicities 2 and 1 (respectively). It is readily seen that the vectors we used to define E_1 are linearly independent, so the following matrix is invertible:

$$P = \begin{bmatrix} 1 & 1 & -3 \\ 1 & 0 & -3 \\ 0 & 1 & 1 \end{bmatrix}.$$

We then diagonalize A:

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

4.4.4 When is a matrix diagonalizable?

What these examples have highlighted is that a matrix may only fail to be diagonalizable if it has repeated eigenvalues. Before stating the next theorem, we quickly state a definition

Definition. If $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_j\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ are sets of vectors in $\mathbb{R}^n$, their ***union*** is the set $\mathcal{B} \cup \mathcal{C} = \{\mathbf{b}_1, \ldots, \mathbf{b}_j, \mathbf{c}_1, \ldots, \mathbf{c}_k\}$.

Theorem 4.4.12 (*Poole* Theorem 4.24)**.** *Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Let $\mathcal{B}_i$ be the basis for E_{λ_i}. The union of the $\mathcal{B}_i$'s (i.e. the collection of all basis vectors in the $\mathcal{B}_i$'s) is a linearly independent set.*

Corollary 4.4.13 (*Poole* Theorem 4.25)**.** *If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.*

What is it about the repeated eigenvalues that causes the failure of diagonalizability of a matrix $A \in \mathbb{R}^{n \times n}$? Well, we need there to be n linearly independent eigenvectors, so we need the geometric multiplicity for each eigenvalue to be as large as possible.

Lemma 4.4.14 (*Poole* Lemma 4.26)**.** *The geometric multiplicity of an eigenvalue λ is less than or equal to its algebraic multiplicity.*

All of this culminates in the following result:

Theorem 4.4.15 (Diagonalization Theorem)**.** *Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. The following are equivalent:*

a. A is diagonalizable.

b. The union of the basis vectors from each E_{λ_i} is a set of n vectors. In other words, $n = \sum_{i=1}^{k} \dim(E_{\lambda_i})$.

c. For each i, the algebraic multiplicity of λ_i is equal to the geometric multiplicity of λ_i.

Example 4.4.16. The matrix $A = \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix}$ has characteristic polynomial

$$\det(A - \lambda I) = (2 - \lambda)^2(4 - \lambda).$$

The eigenvalue 2 has geometric multiplicity 2 and the eigenvalue 4 has geometric multiplicity 1, By the Diagonalization Theorem, A is diagonalizable – you can verify that an appropriate conjugating matrix is

$$P = \begin{bmatrix} -1 & -2 & 1 \\ -3 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}.$$

Example 4.4.17. The matrix $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 4 \\ 0 & 0 & 3 \end{bmatrix}$ has characteristic polynomial

$$\det(A - \lambda I) = (1 - \lambda)(3 - \lambda)^2.$$

Both eigenvalues 1 and 3 have geometric multiplicity 1, so by the Diagonalization Theorem, A is not diagonalizable.

4.4.5 Computational power of diagonal matrices

Notice that for a diagonal matrix $D = \text{diag}(d_1, \dots, d_n)$ and any positive integer k,

$$D^k = \begin{bmatrix} d_1^k & & \\ & \ddots & \\ & & d_n^k \end{bmatrix}.$$

Moreover, if D is invertible, then

$$D^{-k} = \begin{bmatrix} d_1^{-k} & & \\ & \ddots & \\ & & d_n^{-k} \end{bmatrix} = \begin{bmatrix} \frac{1}{d_1^k} & & \\ & \ddots & \\ & & \frac{1}{d_n^k} \end{bmatrix}.$$

This is instantaneous. For a general $n \times n$ matrix A, computing A^k in the usual way is *extremely* computationally expensive. However, if A is diagonalizable, we can write $P^{-1}AP = D$, hence

$$D^k = (P^{-1}AP)^k = P^{-1}A^kP \qquad \Longrightarrow \qquad A^k = PD^kP^{-1}.$$

In this way, computing the k^{th} power of A is only as computationally difficult as diagonalizing A.

Example 4.4.18. Let $A = \begin{bmatrix} 11 & -6 \\ 15 & -8 \end{bmatrix}$. One can readily check that A has eigenvalues 1, 2, hence is diagonalizable (and since the eigenvalues are all nonzero, A is invertible). Through the usual methods, we can obtain

$$A = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix}^{-1}$$

whence, for any integer k,

$$A^k = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 1^k & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 5 & -3 \end{bmatrix} = \begin{bmatrix} -9 + 10(2^k) & 6 - 6(2^k) \\ -15 + 15(2^k) & 10 - 9(2^k) \end{bmatrix}.$$

Similarly,

$$A^{-2} = \begin{bmatrix} -9 + 10(2^{-2}) & 6 - 6(2^{-2}) \\ -15 + 15(2^{-2}) & 10 - 9(2^{-2}) \end{bmatrix} = \begin{bmatrix} -13/2 & 9/2 \\ -45/4 & 31/4 \end{bmatrix}$$

4.4.6 Jordan Canonical Form

The next best thing to being a diagonal matrix is to be a block-diagonal (in fact, a diagonal matrix is just a block-diagonal matrix where all blocks are 1×1). While not everysquare matrix can be diagonalized, every square matrix can be block-diagonalized, with the blocks taking a special form.

Definition. A ***$k \times k$ Jordan block*** is the upper-triangular matrix

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

A matrix is in **Jordan canonical form** (or **Jordan normal form**) if it is a block-diagonal matrix of the form

$$\begin{bmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_n}(\lambda_n) \end{bmatrix}$$

Example 4.4.19. The following matrices are in Jordan canonical form (all blank entries are zero, zeroes were just added to highlight the shapes of the blocks.

$$\begin{bmatrix} 1 & & & & & \\ & 2 & 1 & & & \\ & 0 & 2 & & & \\ & & & 3 & 1 & 0 \\ & & & 0 & 3 & 1 \\ & & & 0 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} 5 & 1 & & & & \\ 0 & 5 & & & & \\ & & 5 & 1 & & \\ & & 0 & 5 & & \\ & & & & 5 & 1 \\ & & & & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 7 & 1 & 0 & & & \\ 0 & 7 & 1 & & & \\ 0 & 0 & 7 & & & \\ & & & 4 & 1 & 0 \\ & & & 0 & 4 & 1 \\ & & & 0 & 0 & 4 \end{bmatrix}.$$

The failure of a matrix to be diagonalizable comes down to a lack of linearly independent eigenvectors. So if a block-diagonal matrix is a generalization of a diagonal matrix, then maybe we can find vectors like eigenvectors that allow us a sufficiently large linearly independent set.

Definition. Given an $n \times n$ matrix A, a ***generalized eigenvector of rank*** m corresponding to the eigenvalue λ is a vector $\mathbf{v} \in \mathbb{R}^n$ for which $(A - \lambda I)^m \mathbf{v} = \mathbf{0}$ but $(A - \lambda I)^{m-1} \mathbf{v} \neq \mathbf{0}$.

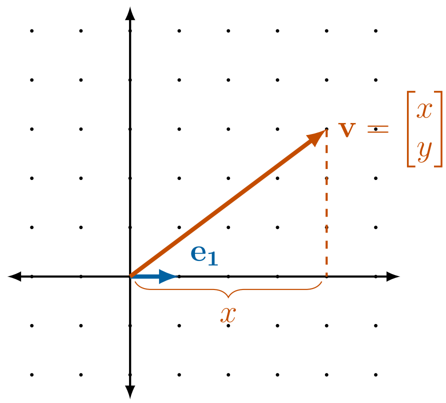
Remark. An eigenvector is a generalized eigenvector of rank 1: it is a vector $\mathbf{v}$ that satisfies $(A - \lambda I)^1 \mathbf{v} = \mathbf{0}$ but $(A - \lambda I)^0 \mathbf{v} = \mathbf{v} \neq 0$

If an eigenvalue has algebraic multiplicity a and geometric multiplicity g, then in general you'll need to find generalized eigenvectors for λ of all ranks up to $1 + a - g$.

More here

5.1 Orthogonality in $\mathbb{R}^n$

At the end of Section 1.2, we saw that orthogonality could be detected with the dot product, but otherwise we haven't really given any geometric description of what the dot product actually measures, so let's consider $\mathbf{e_1} = [1, 0]^T$ and $\mathbf{v} = [x, y]^T$. A straightforward computation shows that $\mathbf{e_1} \cdot \mathbf{v} = x$. Looking at the picture below



We see that x is precisely the length of the "shadow" cast by $\mathbf{v}$ on the line spanned by $\mathbf{e_1}$. More generally, for two vectors $\mathbf{u}$ and $\mathbf{v}$, the dot product $\mathbf{u} \cdot \mathbf{v}$ is the length of "shadow" of $\mathbf{v}$ cast onto $\mathbf{u}$, but then scaled by the length of $\mathbf{v}$ (by symmetry of the dot product, this description also works *mutatis mutandis* for the "shadow" cast by $\mathbf{u}$ onto $\mathbf{v}$).

Let's formalize this notion of a "shadow" cast by $\mathbf{v}$ onto $\mathbf{u}$ keeping in mind two things:

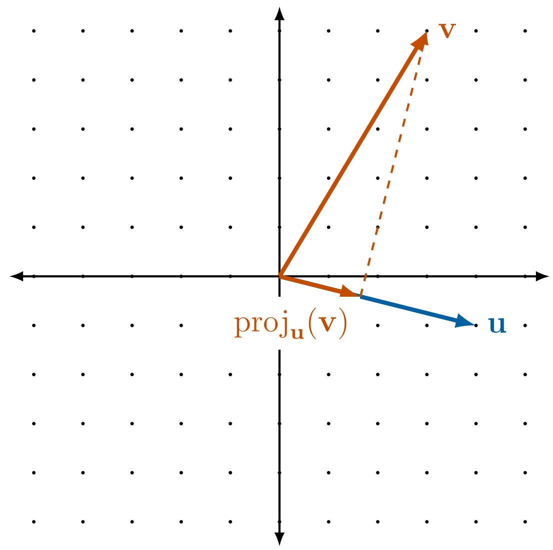
1. $\mathbf{u}$ should be normalized to be a unit vector so as to preserve the length of $\mathbf{v}$'s shadow, and
2. the shadow of a vector should also be a vector in the direction of $\mathbf{u}$ (i.e. a scalar multiple of $\mathbf{u}$).

Definition. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, the ***projection*** (or ***orthogonal projection***) ***of v onto u*** is the vector $\text{proj}_{\mathbf{u}}(\mathbf{v})$ given by

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}}{\|\mathbf{u}\|} \cdot \mathbf{v}\right) \frac{\mathbf{u}}{\|\mathbf{u}\|} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u}.$$

Example 5.1.1. Let $\mathbf{u} = [4, -1]^T$ and $\mathbf{v} = [3, 5]^T$. Using the formula from the definition we have

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u} = \left(\frac{7}{17}\right) \mathbf{u} = \left[\frac{28}{17}, \frac{-7}{17}\right]^T.$$



Notice that the vector $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ (represented by a dashed line) is perpendicular/orthogonal to $\mathbf{u}$:

$$\begin{aligned}(\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})) \cdot \mathbf{u} &= \mathbf{v} \cdot \mathbf{u} - \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u} \cdot \mathbf{u} \\ &= \mathbf{v} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v} \\ &= \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} \\ &= 0.\end{aligned}$$

Definition. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an ***orthogonal set*** of vectors in $\mathbb{R}^n$ if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$ for $i, j = 1, \ldots, k$.

Theorem 5.1.2. *If $\{\mathbf{v_1}, \ldots, \mathbf{v_k}\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then it is linearly independent.*

Proof. Let a_i be scalars for which

$$\mathbf{0} = a_1\mathbf{v_1} + \cdots + a_k\mathbf{v_k}$$

Then for each $i = 1, \ldots, k$,

$$\begin{aligned}0 &= \mathbf{0} \cdot \mathbf{v_1} \\ &= (a_1\mathbf{v_1} + \cdots + a_k\mathbf{v_k}) \cdot \mathbf{v_i} \\ &= a_1(\mathbf{v_1} \cdot \mathbf{v_i}) + \cdots + a_k(\mathbf{v_k} \cdot \mathbf{v_i}) \\ &= a_i(\mathbf{v_i} \cdot \mathbf{v_i})\end{aligned}$$

and since $\mathbf{v_i} \cdot \mathbf{v_i} \neq 0$, then it must be that $a_i = 0$. □

Definition. A basis $\mathcal{B}$ for $\mathbb{R}^n$ is an ***orthogonal basis*** if it is also an orthogonal set.

Example 5.1.3. The standard basis $\mathcal{E}$ for $\mathbb{R}^n$ is orthogonal.

Example 5.1.4. $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^2$ (this is straightforward to check).

Notice that for a vector $\mathbf{v} = [x, y] \in \mathbb{R}^2$, we have that $\mathbf{v} = x\mathbf{e_1} + y\mathbf{e_2}$

Theorem 5.1.5 (*Poole* Theorem 5.2)**.** *Suppose $\mathcal{B} = \{\mathbf{b_1}, \dots, \mathbf{b_k}\}$ is an orthogonal basis for the subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the coefficients c_i of the linear combination*

$$\mathbf{w} = c_1\mathbf{b_1} + \cdots + c_k\mathbf{b_k}$$

are obtained by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{b_i}}{\mathbf{b_i} \cdot \mathbf{b_i}}.$$

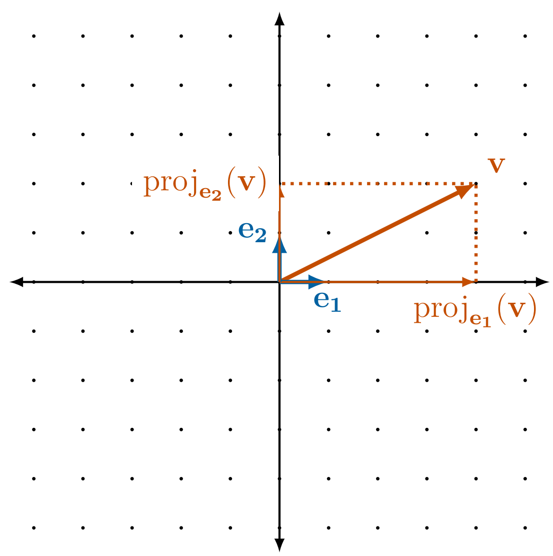
In other words,

$$\mathbf{w} = \text{proj}_{\mathbf{b_1}}(\mathbf{w}) + \cdots + \text{proj}_{\mathbf{b_n}}(\mathbf{w}).$$

Example 5.1.6. Let $\mathbf{v} = [4, 2]^T \in \mathbb{R}^2$ and consider the standard basis $\mathcal{E}$. A straightforward computation shows that

$$\text{proj}_{\mathbf{e_1}}(\mathbf{v}) = 4\mathbf{e_1} \qquad \text{and} \qquad \text{proj}_{\mathbf{e_2}}(\mathbf{v}) = 2\mathbf{e_2}$$

and clearly $\mathbf{v} = 4\mathbf{e_1} + 2\mathbf{e_2}$.

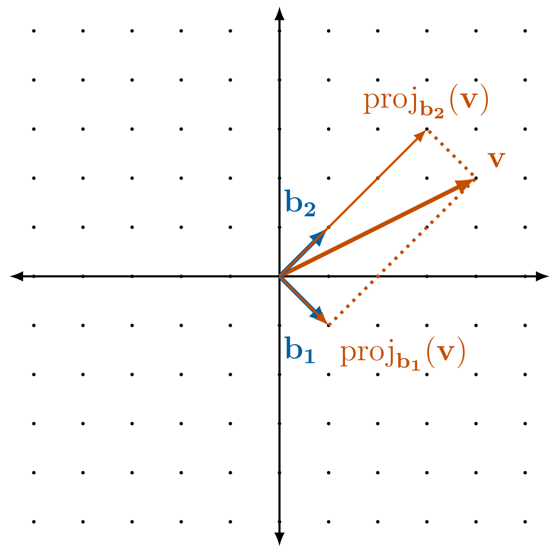


Example 5.1.7. Let $\mathbf{v}$ be as in the previous example and consider the orthogonal basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Computing the orthogonal projection of $\mathbf{v}$ onto the basis vectors we have that

$$\text{proj}_{\mathbf{b_1}}(\mathbf{v}) = \frac{4-2}{2}\mathbf{b_1} = \mathbf{b_1} \qquad \text{and} \qquad \text{proj}_{\mathbf{b_2}}(\mathbf{v}) = \frac{4+2}{2}\mathbf{b_2} = 3\mathbf{b_2}$$

and certainly

$$\mathbf{v} = \mathbf{b_1} + 3\mathbf{b_2}. \tag{5.1.1}$$



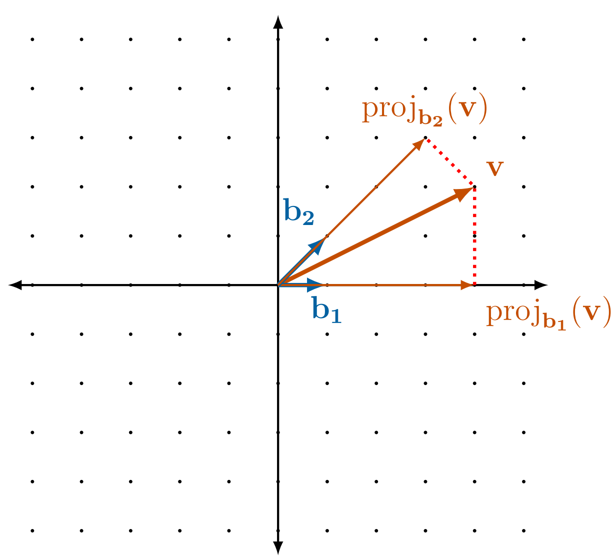
Example 5.1.8. To see why orthogonality of the basis is important, let $\mathbf{v}$ be as in the previous two examples and consider the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$, which is not orthogonal. Computing the orthogonal projection of $\mathbf{v}$ onto the basis vectors, we have that

$$\text{proj}_{\mathbf{b_1}}(\mathbf{v}) = 4\mathbf{b_1} \qquad \text{and} \qquad \text{proj}_{\mathbf{b_2}}(\mathbf{v}) = \frac{4+2}{2}\mathbf{b_2} = 3\mathbf{b_2},$$

but

$$\mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 7 \\ 3 \end{bmatrix} = 4\mathbf{b_1} + 3\mathbf{b_2}.$$

(In fact, by inspection we see that $\mathbf{v} = 2\mathbf{b_1} + 2\mathbf{b_2}$).



5.1.1 Orthonormality

Definition. An orthogonal set of vectors is ***orthonormal*** if each vector is also a unit vector. An ***orthogonal basis*** for $\mathbb{R}^n$ is an orthonormal set that is also a basis for $\mathbb{R}^n$.

Proposition 5.1.9. *Suppose* $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ *is a basis for* $\mathbb{R}^n$*. Then* $\mathcal{B}$ *is orthonormal if and only if* $\mathbf{b_i} \cdot \mathbf{b_j} = \delta_{ij}$ *for all* $i, j = 1, \ldots, n$ *(where* δ_{ij} *is the Kroenecker delta).*

Example 5.1.10. The standard basis $\mathcal{E}$ for $\mathbb{R}^n$ is orthonormal.

Example 5.1.11. The basis $\mathcal{B}$ from Example 5.1.4 is not orthonormal because

$$\left\| \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\| = \sqrt{2} \qquad \text{and} \qquad \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \sqrt{2}.$$

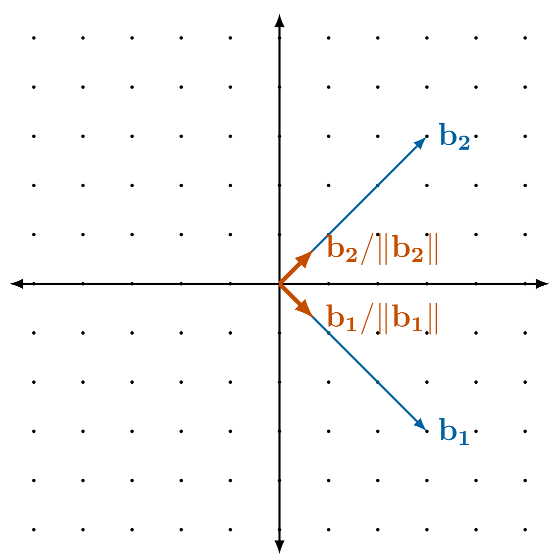
Obviously we like the standard basis and its orthonormality, so it would be nice if we could supe up an orthogonal basis to be a normal basis. Indeed, this is easily achieved by merely normalizing each vector.

Proposition 5.1.12. *If* $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ *is an orthogonal basis for* $\mathbb{R}^n$*, then* $\mathcal{B}' = \left\{ \frac{\mathbf{b_1}}{\|\mathbf{b_1}\|}, \ldots, \frac{\mathbf{b_n}}{\|\mathbf{b_n}\|} \right\}$ *is an orthonormal basis for* $\mathbb{R}^n$

Example 5.1.13. Normalizing the basis vectors in Example 5.1.4, we get that

$$\mathcal{B}' = \left\{ \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

is an orthonormal basis for $\mathbb{R}^2$.



What you may notice is that the orthonormal basis $\mathcal{B}'$ in the previous example is just a $-\frac{\pi}{4}$ rotation of the standard basis. Amazingly, this same feature is very nearly true of all orthornormal bases.

Theorem 5.1.14. *Every orthonormal basis of* $\mathbb{R}^n$ *can be obtained by a rotation of the standard basis (possibly followed by a permutation of the basis vectors).*

It follows that, given two orthonormal bases $\mathcal{B}$ and $\mathcal{C}$ of $\mathbb{R}^2$, the change of basis matrix is either a rotation matrix

$$P_{\mathcal{C} \leftarrow \mathcal{B}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

or a rotation matrix times a matrix that swaps the basis vectors

$$P_{\mathcal{C}\leftarrow\mathcal{B}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{bmatrix}$$

(although, since the particular angle θ is freely chosen, we would often make the substitution $\theta = \varphi + \frac{\pi}{2}$ to get

$$\begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}$$

as this provides some nice symmetries in the forms of the matrix).

5.2 Orthogonal Complements and Orthogonal Projections

Our goal will ultimately be to come up with a procedure for finding an orthogonal basis for a subspace. In doing this, we first need to introduce the following notion.

5.2.1 Orthogonal Complement

Definition. Let W be a subspace of $\mathbb{R}^n$. A vector $\mathbf{v} \in \mathbb{R}^n$ is ***orthogonal to*** W if it is orthogonal to every vector $\mathbf{w} \in W$. The collection of all such vectors is called the ***orthogonal complement to*** W and is denoted $W^\perp$.

Fact. If W is a subspace of $\mathbb{R}^n$, then $W^\perp$ is a subspace of $\mathbb{R}^n$.

Proof. Suppose $\mathbf{w} \in W$ and that $\mathbf{u}, \mathbf{v}$ are orthogonal to W. It is then straightforward to check that

- $\mathbf{0} \cdot \mathbf{w} = 0$
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = 0$
- $k\mathbf{v} \cdot \mathbf{w} = 0$ for any scalar k

and thus it follows that the collection of all vectors orthogonal to W is indeed a subspace. □

Example 5.2.1. Suppose W is the xy-plane in $\mathbb{R}^3$ (i.e. the set of vectors $[x, y, 0]^T$). Then $W^\perp$ is the z-axis (i.e. the set of vectors $[0, 0, z]^T$). Geometrically, $W^\perp$ is the line through the origin that is perpendicular to W

Theorem 5.2.2 (*Poole* Theorem 5.9)**.** *Let W be a subspace of $\mathbb{R}^n$.*

1. *$W^\perp$ is also a subspace of $\mathbb{R}^n$.*
2. *$(W^\perp)^\perp = W$*
3. *The only vector common to both W and $W^\perp$ is $\mathbf{0}$ (we say that W and $W^\perp$ have* trivial intersection*).*
4. *If $W = \text{Span}(\mathbf{w_1}, \ldots, \mathbf{w_k})$, then $W^\perp$ is the set of vectors perpendicular to each $\mathbf{w_i}$.*

How do we find orthogonal complements in practice?

Theorem 5.2.3 (*Poole* Theorem 5.10)**.** *Let A be an $m \times n$ matrix. Then $\text{Null}(A) = (\text{row } A)^\perp$ and $\text{Null}(A^T) = (\text{col } A)^\perp$.*

Proof. If A is an $m \times n$ matrix with rows $\mathbf{A_1}, \ldots, \mathbf{A_m}$ and $\mathbf{x} \in R^n$, then

$$A\mathbf{x} = \begin{bmatrix} \text{—}\mathbf{A_1}\text{—} \\ \vdots \\ \text{—}\mathbf{A_m}\text{—} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{A_1} \cdot \mathbf{x} \\ \vdots \\ \mathbf{A_m} \cdot \mathbf{x} \end{bmatrix}$$

so $A\mathbf{x} = \mathbf{0}$ precisely when $\mathbf{A_i} \cdot \mathbf{x} = 0$ for each $i = 1, \ldots, m$. □

As such, solving for the orthogonal complement can be done by explicitly solving for the null space of the appropriate matrix of vectors.

Example 5.2.4. Let $W = \operatorname{Span}\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\-1\\0\end{bmatrix}\right)$ be a plane in $\mathbb{R}^3$. By viewing the above vectors as row vectors of a matrix and applying the above theorem, we have that

$$W^\perp = \operatorname{Null}\left(\begin{bmatrix}1 & 2 & 3\\2 & -1 & 0\end{bmatrix}\right).$$

Since

$$\operatorname{RREF}\left(\begin{bmatrix}1 & 2 & 3\\2 & -1 & 0\end{bmatrix}\right) = \begin{bmatrix}1 & 0 & 3/5\\0 & 1 & 6/5\end{bmatrix}$$

we deduce that $W^\perp = \operatorname{Span}\left(\begin{bmatrix}3/5\\6/5\\-1\end{bmatrix}\right)$.

Remark. Your book uses the notation $\operatorname{perp}_{\mathbf{u}}(\mathbf{v})$ to mean $\mathbf{v} - \operatorname{proj}_{\mathbf{u}}(\mathbf{v})$. This is reasonable, but I can't say its particularly common.

The following definition extends the idea of orthogonal projection onto an entire subspace.

Definition. Let W be a subspace of $\mathbb{R}^n$ and $\{\mathbf{w_1}, \ldots, \mathbf{w_k}\}$ an orthogonal basis for W. For a vector $\mathbf{v} \in \mathbb{R}^n$, the ***orthogonal projection of*** $\mathbf{v}$ ***onto*** W is

$$\operatorname{proj}_W(\mathbf{v}) = \operatorname{proj}_{\mathbf{w_1}}(\mathbf{v}) + \cdots + \operatorname{proj}_{\mathbf{w_k}}(\mathbf{v})$$

and the ***component of*** $\mathbf{v}$ ***orthogonal to*** W is

$$\operatorname{perp}_W(\mathbf{v}) = \mathbf{v} - \operatorname{proj}_W(\mathbf{v}).$$

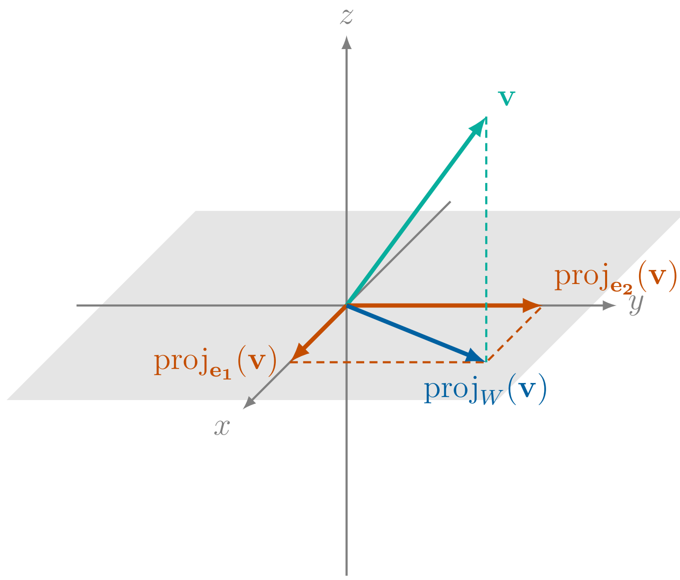
Remark. Once again, for any subspace W of $\mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$, we have the *orthogonal decomposition* of $\mathbf{v}$:

$$\mathbf{v} = \underbrace{\operatorname{proj}_{\mathbf{u}}(\mathbf{v})}_{\text{in } W} + \underbrace{(\mathbf{v} - \operatorname{proj}_{\mathbf{u}}(\mathbf{v}))}_{\text{in } W^\perp}$$

Example 5.2.5. Suppose W is the xy-plane in $\mathbb{R}^3$ and let $\mathbf{v} = [3, 4, 5]^T$. Then

$$\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{e_1}}(\mathbf{v}) + \text{proj}_{\mathbf{e_2}}(\mathbf{v}) = 3\mathbf{e_1} + 4\mathbf{e_2} = [3, 4, 0]^T.$$

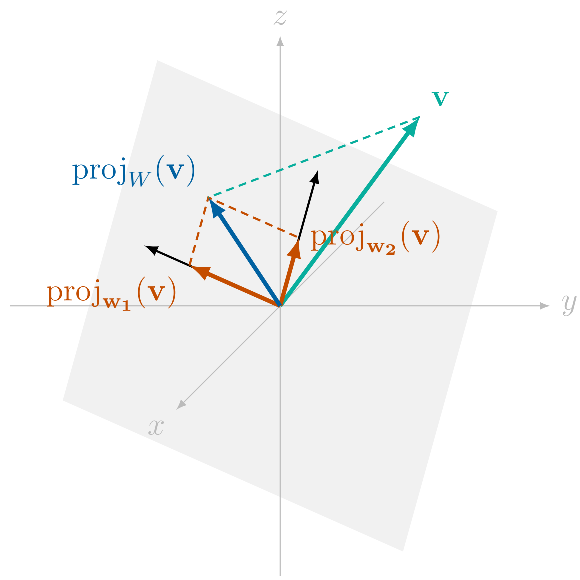
Visually,



Example 5.2.6. Suppose $\mathbf{w_1} = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$, $\mathbf{w_2} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ are vectors in $\mathbb{R}^3$ and $W = \text{Span}(\mathbf{w_1}, \mathbf{w_2})$. We first check that $\mathbf{w_1} \cdot \mathbf{w_2} = 0$, whence $\{\mathbf{w_1}, \mathbf{w_2}\}$ is an orthogonal basis for W. To compute the projection of $\mathbf{v}$ onto W

$$\begin{aligned}
\text{proj}_W(\mathbf{v}) &= \text{proj}_{\mathbf{w_1}}(\mathbf{v}) + \text{proj}_{\mathbf{w_2}}(\mathbf{v}) \\
&= \left(\frac{\mathbf{v} \cdot \mathbf{w_1}}{\mathbf{w_1} \cdot \mathbf{w_1}}\right)\mathbf{w_1} + \left(\frac{\mathbf{v} \cdot \mathbf{w_2}}{\mathbf{w_2} \cdot \mathbf{w_2}}\right)\mathbf{w_2} \\
&= \left(\frac{6 - 8 + 10}{12}\right)\mathbf{w_1} + \left(\frac{-6 + 10}{8}\right)\mathbf{w_2} \\
&= \frac{2}{3}\mathbf{w_1} + \frac{3}{4}\mathbf{w_2} \\
&= \frac{1}{6}\begin{bmatrix} -1 \\ -8 \\ 17 \end{bmatrix}.
\end{aligned}$$

Visually,



What's amazing is that the orthogonal decomposition of a vector is actually unique.

Theorem 5.2.7 (Orthogonal Decomposition Theorem)**.** *Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$. Then there are unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ for which $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$.*

In particular, $\mathbf{w} = \mathrm{proj}_W(\mathbf{v})$ and $\mathbf{w}^\perp = v - \mathrm{proj}_W(\mathbf{v})$.

Theorem 5.2.8. *If W is a subspace of $\mathbb{R}^n$, then $\dim W + \dim W^\perp = n$*

The above isn't particularly surprising; it's exactly what we saw happen in the first example of this section. More generally, given a basis of column vectors $\{\mathbf{w_1}, \ldots, \mathbf{w_k}\}$ for W, we find $W^\perp$ by using Theorem 5.2.3:

$$W^\perp = \mathrm{Null}\left(\begin{bmatrix} — \mathbf{w_1}^T — \\ \vdots \\ — \mathbf{w_k}^T — \end{bmatrix}\right).$$

The matrix above has size $k \times n$ and rank k, and so the result follows by Rank-Nullity.

Of course, the proof given in the book doesn't rely on Rank-Nullity at all. In fact, Rank-Nullity actually a corollary of the above theorem (and the proof is given by the same observation as above).

5.3 The Gram–Schmidt Process and the QR Factorization

5.3.1 Gram–Schmidt

It sure would be nice to be able to find an orthogonal basis for every subspace, huh?

Example 5.3.1. Let $\mathbf{b_1} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{b_2} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ be vectors in $\mathbb{R}^3$ and let $W = \text{Span}(\mathbf{b_1}, \mathbf{b_2})$. Recall that

$$\mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2}) = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 21/25 \\ 28/25 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/25 \\ -3/25 \\ 2 \end{bmatrix}$$

is perpendicular to $\mathbf{b_1}$ and is still contained within W (because $\text{proj}_{\mathbf{b_1}}(\mathbf{b}_2)$ is a scalar multiple of $\mathbf{b}_1$). This means that $\{\mathbf{b_1}, \mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2})\}$ is an orthogonal basis for W!

As it turns out, the above example can be extended into any dimension, and this iterative process is known as the ***Gram-Schmidt orthogonalization***.

Theorem 5.3.2 (Gram–Schmidt)**.** *Let $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_k}\}$ be a basis for W, a subspace of $\mathbb{R}^n$.*

1. *Let $\mathbf{x_1} = \mathbf{b_1}$, and let $W_1 = \text{Span}(\mathbf{x_1})$.*
2. *For each $i = 2, ..., k$, let $\mathbf{x_i} = \mathbf{b_i} - \text{proj}_{W_{i-1}}(\mathbf{b_i})$ and set $W_i = \text{Span}(\mathbf{x_1}, \ldots, x_i)$*

For each i, $\{x_1, \ldots, x_i\}$ is an orthogonal basis for W_i and $W_k = W$.

Remark. One can always scale the basis elements to have norm 1, further producing an orthonormal basis.

Example 5.3.3. Find an ortho<u>normal</u> basis for $W = \text{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \\ 2 \end{bmatrix}\right)$.

We choose

$$\mathbf{x_1} = [1, 2, 2, 0]^T$$

and we set $W_1 = \text{Span}(\mathbf{x_1})$. Then

$$\begin{aligned} \mathbf{x_2} &= [0, 1, 2, 2]^T - \text{proj}_{W_1}([0, 1, 2, 2]^T) \\ &= [0, 1, 2, 2]^T - \text{proj}_{\mathbf{x_1}}([0, 1, 2, 2]^T) \\ &= [0, 1, 2, 2]^T - \left[\frac{2}{3}, \frac{4}{3}, \frac{4}{3}, 0\right]^T \\ &= \left[-\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}, 2\right]^T \end{aligned}$$

and we set $W_2 = \text{Span}(\mathbf{x_1}, \mathbf{x_2})$. Then

$$\mathbf{x_3} = [2, 0, 1, 2]^T - \text{proj}_{W_2}([2, 0, 1, 2]^T)$$

$$
\begin{aligned}
&= [2,0,1,2]^T - \text{proj}_{\mathbf{x_1}}([2,0,1,2]^T) - \text{proj}_{\mathbf{x_2}}([2,0,1,2]^T) \\
&= [2,0,1,2]^T - \left[\frac{4}{9}, \frac{8}{9}, \frac{8}{9}, 0\right]^T - \left[-\frac{4}{9}, -\frac{2}{9}, \frac{4}{9}, \frac{4}{3}\right]^T \\
&= \left[2, -\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right]^T.
\end{aligned}
$$

and $\{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}\}$ is an orthogonal basis for W. To form an orthonormal basis, we normalize each of these vectors, hence an orthonormal basis for W is

$$
\left\{ \begin{bmatrix} 1/3 \\ 1/3 \\ 2/3 \\ 0 \end{bmatrix}, \begin{bmatrix} -2/3\sqrt{5}) \\ -1/3\sqrt{5} \\ 2/3\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \begin{bmatrix} 2/\sqrt{5} \\ -2/3\sqrt{5} \\ -1/3\sqrt{5} \\ 2/3\sqrt{5} \end{bmatrix} \right\}.
$$

5.3.2 QR Factorization

If A is an $m \times n$ matrix with linearly independent columns (implying $m \geq n$), then applying the Gram–Schmidt process to the columns yields a useful factorization of A, although its usefulness will not be immediately obvious.

Theorem 5.3.4 (QR Factorization)**.** *Let A be an $m \times n$ matrix with linearly independent columns. Then there exists a matrix Q with orthonormal columns and an invertible upper triangular matrix R for which $A = QR$.*

The proof is constructive. Let $\{\mathbf{A_1}, \ldots, \mathbf{A_n}\}$ be the columns of A and let $\{\mathbf{Q_1}, \ldots, \mathbf{Q_n}\}$ be the *orthonormal* basis produced from applying Gram-Schmidt to the $\mathbf{A_i}$'s. Notice that in the Gram-Schmidt process, we have

$$
\begin{aligned}
\mathbf{Q_1} &= c_1 \mathbf{A_1} \\
\mathbf{Q_2} &= c_2 \left(\mathbf{A_2} - \left(\frac{\mathbf{Q_1} \cdot \mathbf{A_2}}{\mathbf{Q_1} \cdot \mathbf{Q_1}} \right) \mathbf{Q_1} \right) \\
\mathbf{Q_3} &= c_3 \left(\mathbf{A_3} - \left(\frac{\mathbf{Q_1} \cdot \mathbf{A_3}}{\mathbf{Q_1} \cdot \mathbf{Q_1}} \right) \mathbf{Q_1} - \left(\frac{\mathbf{Q_2} \cdot \mathbf{A_3}}{\mathbf{Q_2} \cdot \mathbf{Q_2}} \right) \mathbf{Q_2} \right) \\
&\vdots
\end{aligned}
$$

where the c_i's are all the scalars normalizing the vectors.

Since all of the dot products are just scalars, we can write

$$
r_{ij} = \begin{cases} 1/c_j & \text{if } i = j \\ \left(\dfrac{\mathbf{Q_i} \cdot \mathbf{A_j}}{\mathbf{Q_i} \cdot \mathbf{Q_i}} \right) & \text{if } i \neq j \end{cases}
$$

and rearrange the above equations to be

$$
\mathbf{A_1} = r_{11} \mathbf{Q_1}
$$

$$\begin{aligned}\mathbf{A_2} &= r_{12}\mathbf{Q_1} + r_{22}\mathbf{Q_2}\\ \mathbf{A_3} &= r_{13}\mathbf{Q_1} + r_{23}\mathbf{Q_2} + r_{33}\mathbf{Q_3}\\ &\vdots\end{aligned}$$

The above system can be represented as the following matrix product:

$$A = \begin{bmatrix} | & & | \\ \mathbf{A_1} & \cdots & \mathbf{A_n} \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{Q_1} & \cdots & \mathbf{Q_n} \\ | & & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix} = QR$$

Remark. We can always take the diagonal entries r_{ii} to be positive: if $r_{ii} < 0$, then simply replace $\mathbf{Q_i}$ with $-\mathbf{Q_i}$.

Remark. Since Q is $m \times n$ with orthonormal columns, then $Q^T Q = I_n$, so in fact $R = Q^T A$, saving us some time in computing R.

Example 5.3.5. Compute the QR factorization of $A = \begin{bmatrix} 12 & -51 & -4 \\ 6 & 167 & 68 \\ -4 & 24 & 41 \end{bmatrix}$

We first apply the Gram-Schmidt process to the columns. Let $\mathbf{A_i}$ denote the i^{th} column of A. We take $\mathbf{x_1} = \mathbf{A_1}$. Letting $W_1 = \text{Span}(\mathbf{x_1})$,

$$\mathbf{x_2} = \mathbf{A_2} - \text{proj}_{W_1}(\mathbf{A_2}) = [-69, 158, 30]^T.$$

Letting $W_2 = \text{Span}(\mathbf{x_1}, \mathbf{x_2})$,

$$\mathbf{x_3} = \mathbf{A_3} - \text{proj}_{W_2}(\mathbf{A_3}) = \left[\frac{58}{5}, -\frac{6}{5}, 33\right]^T.$$

Now $\{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}\}$ is an orthogonal basis for $\mathbb{R}^3$. Letting $\mathbf{Q_i} = \dfrac{\mathbf{x_i}}{\|\mathbf{x_i}\|}$, we form the orthogonal matrix

$$Q = \begin{bmatrix} | & | & | \\ \mathbf{Q_1} & \mathbf{Q_2} & \mathbf{Q_3} \\ | & | & | \end{bmatrix} = \begin{bmatrix} \frac{6}{7} & -\frac{69}{175} & \frac{58}{175} \\ \frac{3}{7} & \frac{158}{175} & -\frac{6}{175} \\ -\frac{2}{7} & \frac{6}{35} & \frac{33}{35} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} 14 & 21 & 14 \\ 0 & 175 & 70 \\ 0 & 0 & 35 \end{bmatrix}.$$

7.3 Least Squares

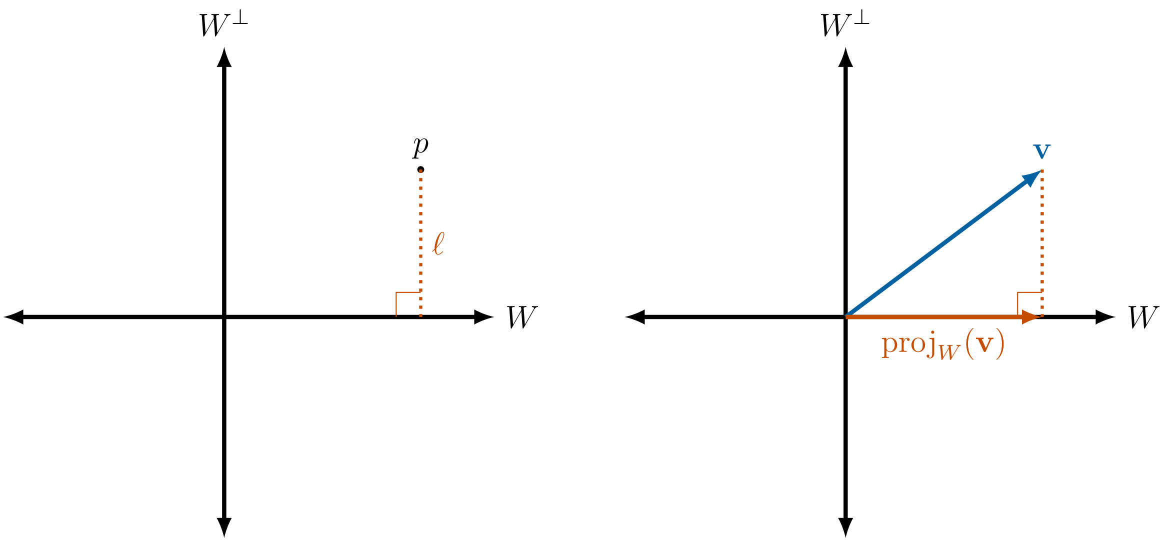
By now the power of linear algebra should be apparent, so we'd like to try to use this tool in many real-world applications, and such applications often times require us to make approximations.

Definition. Given a vector subspace W of $\mathbb{R}^n$ and a vector $\mathbf{v} \in \mathbb{R}^n$, the ***best approximation of*** $\mathbf{v} \in W$ is the vector $\overline{\mathbf{v}} \in W$ that is closest to $\mathbf{v}$, i.e., that satisfies

$$\|\mathbf{v} - \overline{\mathbf{v}}\| < \|\mathbf{v} - \mathbf{w}\|$$

for all $\mathbf{w} \in W$ with $\mathbf{w} \neq \overline{\mathbf{v}}$.

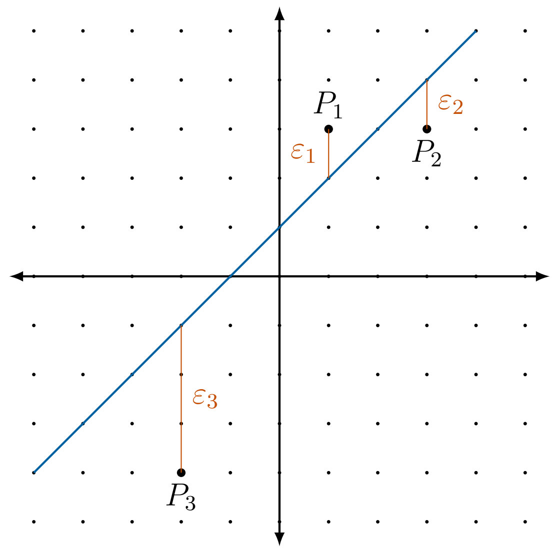
Certainly we could use some calculus techniques to find this, but we could also appeal to a fact of Euclidean geometry – the distance between a point p and a subspace W of $\mathbb{R}^n$ is the length of the line segment ℓ which is perpendicular to W and has endpoint p.



Proposition 7.3.1 (*Poole Theorem 7.8*, The Best Approximation Theorem)**.** *For any* $\mathbf{v} \in \mathbb{R}^n$ *and any subspace* W *of* $\mathbb{R}^n$, $\text{proj}_W(\mathbf{v})$ *is the best approximation of* $\mathbf{v} \in \mathbb{R}^n$.

7.3.1 Least Squares Approximation

Suppose that we have points P_1, P_2, P_3 in the plane and we approximate these three points with the line $y = mx + b$. Let ε_1, ε_2, ε_3 be the vertical distance between these points and the line. The ***error vector*** is $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \varepsilon_3]$ and the number $\|\boldsymbol{\varepsilon}\| = \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2}$ is called the ***least squares error*** of the approximation.



Example 7.3.2. The image above shows the case where $P_1(1,3)$, $P_2(3,3)$, $P_3(-2,-4)$, $y = x+1$, and $\varepsilon_1 = 1$, $\varepsilon_2 = 1$ and $\varepsilon_3 = 3$. Thus the least squares error is

$$\|\boldsymbol{\varepsilon}\| = \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2} = \sqrt{11}$$

Of course, in practice, we're most interested with finding the line that minimizes the least squares error, and such a line is called the ***least squares error approximating line*** (or the ***best fit line***).

Suppose we have a bunch of points (x_i, y_i) in the plane and we approximate them with the line $y = mx + b$. Then we have that

$$\varepsilon_i = y_i - y = y_i - mx_i - b.$$

Since the above is a linear equation, we can rewrite it slightly in terms of matrices and vectors. Letting

$$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} b \\ m \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

then we have that

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} mx_1 + b \\ \vdots \\ mx_n + b \end{bmatrix} = \mathbf{b} - A\mathbf{x}.$$

and hence that

$$\|\boldsymbol{\varepsilon}\| = \|\mathbf{b} - A\mathbf{x}\|.$$

The entries of A and $\mathbf{b}$ are fixed, and the approximating line is entirely encoded by the vector $\mathbf{x}$. So if $\overline{\mathbf{x}}$ represents the line that best approximates the points, then we should have, for any other line $\mathbf{x}$,

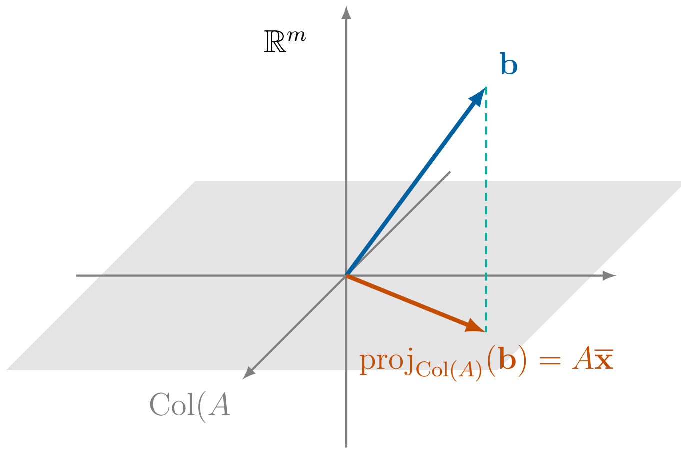
$$\|\mathbf{b} - A\overline{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|.$$

Definition. If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, then a ***least squares solution*** to the equation $A\mathbf{x} = \mathbf{b}$ is a vector $\overline{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\overline{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Consider the following picture:



Since $A\mathbf{x}$ is in $\text{Col}(A)$, by the Best Approximation Theorem, the vector $\overline{\mathbf{x}}$ for which $A\overline{\mathbf{x}}$ is closest to $\mathbf{b}$ satisfies $A\overline{\mathbf{b}} = \text{proj}_{\text{Col}(A)}(\mathbf{b})$!

Rather than computing this outright, notice that

$$\mathbf{b} - A\overline{\mathbf{x}} = \mathbf{b} - \text{proj}_{\text{Col}(A)}(\mathbf{b})$$

is perpendicular to $\text{Col}(A)$. So if $A = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$, then

$$A^T(\mathbf{b} - A\overline{\mathbf{x}}) = \begin{bmatrix} — \mathbf{A}_1^T — \\ \vdots \\ — \mathbf{A}_n^T — \end{bmatrix} (\mathbf{b} - A\overline{\mathbf{x}}) = \begin{bmatrix} \mathbf{A}_1 \cdot (\mathbf{b} - A\overline{\mathbf{x}}) \\ \vdots \\ \mathbf{A}_n \cdot (\mathbf{b} - A\overline{\mathbf{x}}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}.$$

And therefore $\overline{\mathbf{x}}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$ if and only if

$$A^T A\overline{\mathbf{x}} = A^T\mathbf{b} \tag{7.3.1}$$

Definition. Equation 7.3.1 represents the ***normal equations*** for $\overline{\mathbf{x}}$.

Theorem 7.3.3 (*Poole 7.9*, The Least Squares Theorem)**.** *Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. Then $A\mathbf{x} = \mathbf{b}$ always has at least one least squares solution $\overline{x}$. Moreover,*

1. *$\overline{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\overline{\mathbf{x}}$ is a solution of the normal equations $A^T A\overline{\mathbf{x}} = A^T\mathbf{b}$.*

2. *A has linearly independent columns if and only if A^TA is invertible, in which case the least squares solution is unique:*

$$\overline{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

Corollary 7.3.4. *If there is a unique least squares solution to $A\mathbf{x} = \mathbf{b}$ and $A = QR$ (with Q orthogonal and R upper triangular), then it is given by $\overline{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$*

Proof. The proof is straightforward, just notice that

$$A^TA = (QR)^T(QR) = R^TQ^TQR = R^TIR = R^TR.$$

□

Remark. If you have a QR-factorization of A, then finding R^{-1} is extremely fast, and so computing a least squares solution is considerably faster than it would be for $(A^TA)^{-1}A^T\mathbf{b}$. Practically speaking, the difference is indistinguishable for matrices of small size, but when working with matrix algebra software and huge matrices (think $20,000 \times 10,000$), computation time is a very important consideration.

Example 7.3.5. We want to find a least squares approximating line $y = mx + b$ for the points in Example 7.3.2. Using the above discussion, this is equivalent to finding a least squares solution to $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & -2 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} b \\ m \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}.$$

We thus look for the solution set to

$$A^TA\mathbf{x} = A^T\mathbf{b}$$
$$\begin{bmatrix} 3 & 2 \\ 2 & 14 \end{bmatrix}\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 2 \\ 20 \end{bmatrix}.$$

By row reducing the augmented system $\left[A^TA \mid A^T\mathbf{b}\right]$, we get

$$\left[A^TA \mid A^T\mathbf{b}\right] = \left[\begin{array}{cc|c} 3 & 2 & 2 \\ 2 & 14 & 20 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & -6/19 \\ 0 & 1 & 28/29 \end{array}\right]$$

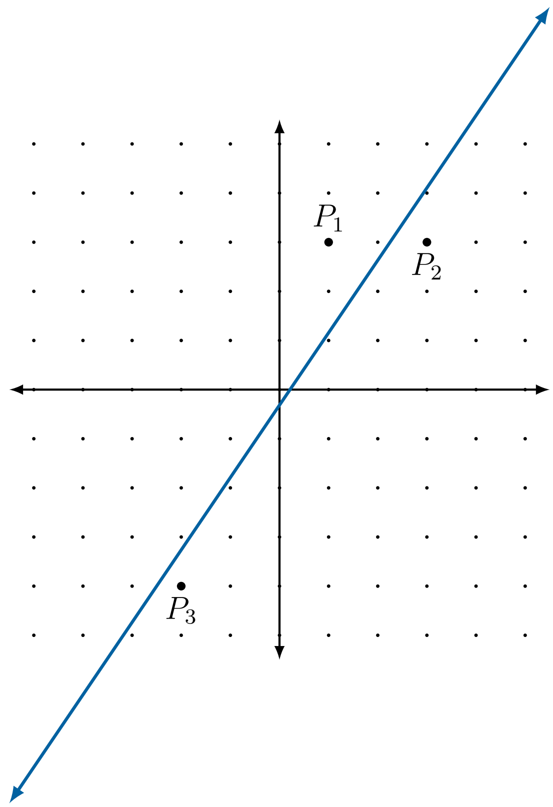
whence the least squares approximating line is

$$y = \frac{28}{19}x - \frac{6}{19}.$$

The least squares error vector for this line is

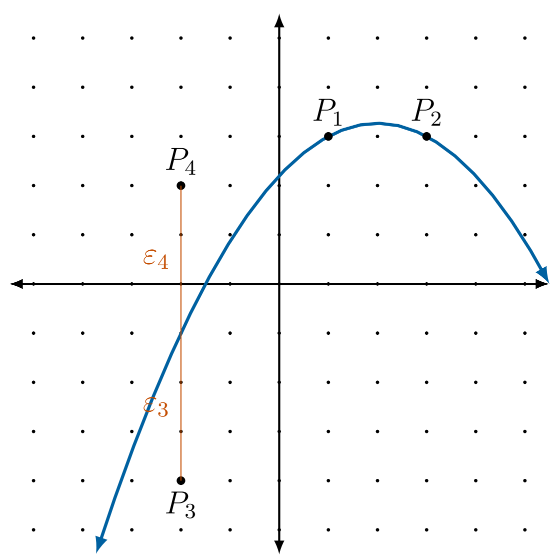
$$\boldsymbol{\varepsilon} = \mathbf{b} - A\mathbf{x} = \begin{bmatrix} 35/19 \\ -21/19 \\ -14/19 \end{bmatrix}$$

and the least squares error is $\|\boldsymbol{\varepsilon}\| \approx 2.2711$. This is better than the previous line we had, as that had an error of $\sqrt{11} \approx 3.3166$.



The above strategy can be employed for approximating planes, etc., making the obvious changes to A, $\mathbf{x}$, and $\mathbf{b}$. It can also be used for polynomials, although in general you'll need at least $n+2$ points to bother with finding a best approximation polynomial of degree n. (This is due to the following fact: for any $n+1$ points in the plane, there is a unique polynomial of degree n that passes through them. As such, there's no point in approximating when an exact solution exists.)

Example 7.3.6. Find the best least squares approximating quadratic $y = ax^2 + bx + c$ for the points $P_1(1, 3)$, $P_2(3, 3)$, $P_3(-2, -4)$, $P_4(-2, 2)$.



If the points above were on the quadratic, we would have

$$a(1)^2 + b(1) + c = 3$$

$$\begin{aligned} a(3)^2 + b(3) + c &= 3 \\ a(-2)^2 + b(-2) + c &= -4 \\ a(-2)^2 + b(-2) + c &= 2 \end{aligned}$$

and so this system can be reinterpreted as $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 9 & 3 & 1 \\ 4 & -2 & 1 \\ 4 & -2 & 1 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \\ 2 \end{bmatrix}$$

It is an exercise to the reader to show from here that the least squares approximating parabola is $y = -\frac{4}{15}x^2 + \frac{16}{15}x + \frac{11}{5}$ and the least squares error is $\|\boldsymbol{\varepsilon}\| = \sqrt{0^2 + 0^2 + 3^2 + 3^2} = 3\sqrt{2} \approx 4.243$.

Note to self: Rewrite following example to use normal equations

Example 7.3.7. Find the least squares solution set for $A\mathbf{x} = \mathbf{b}$ where $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

The size of the matrix shows that $\text{nullity}(A) > 0$ (and in fact, a quick computation shows that $\text{nullity}(A) = 2$), so A does not have linearly independent columns and thus are infinitely many least squares solutions.

To get the least squares solutions, we first compute $\text{proj}_{\text{Col}(A)}(\mathbf{b})$. We first note that

$$\text{Col}(A) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$$

and the basis for $\text{Col}(A)$ is orthogonal. Hence

$$\text{proj}_{\text{Col}(A)}(\mathbf{b}) = \text{proj}_{[1,0,1]^T}(\mathbf{b}) + \text{proj}_{[0,1,0]}(\mathbf{b}) = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

The least squares solution set is now the solution set to the equation

$$A\mathbf{x} = \text{proj}_{\text{Col}(A)}(\mathbf{b}) \qquad \Rightarrow \qquad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

hence

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow{RREF} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

hence $x_1 = 2 - x_4$ and $x_2 = 2 - x_3$, so the least squares solution set is

$$\left\{ \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \ : \ \text{where } s, t \in \mathbb{R} \right\}$$