MAT 2114 Intro to Linear Algebra

Joe Wells
Virginia Tech

Spring 2024

Last updated: April 23, 2024

Contents

Preface

There are many different approaches to linear algebra, and everyone has their preference. This document is largely compiled from the course I taught in the Spring of 2020 at Virginia Tech, where the book (*Linear Algebra: A Modern Introduction* 4th Ed. by David Poole) was out of my control. Over the years, I've modified the precise topic ordering, which mostly follows the book with a few exceptions.

Although not formally stated anywhere, this class is 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 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 – it seems perverse to leave the phrase "vector space" until the 6^{th} chapter! Nevertheless, I did my best to gently introduce concepts as needed in order to more smoothly segue the topics. As well, many of the homework exercises are designed to bridge certain theoretical gaps in the material and introduce concepts much earlier than the text (notably, linear transformations).

I would like to thank the many students who inadvertently served as my copy editors over the years; now I understand why authors seem to produce new versions every few years (somehow there are typos that persist for generations).

Chapter 1

Vectors

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 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 1.1.2). 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.

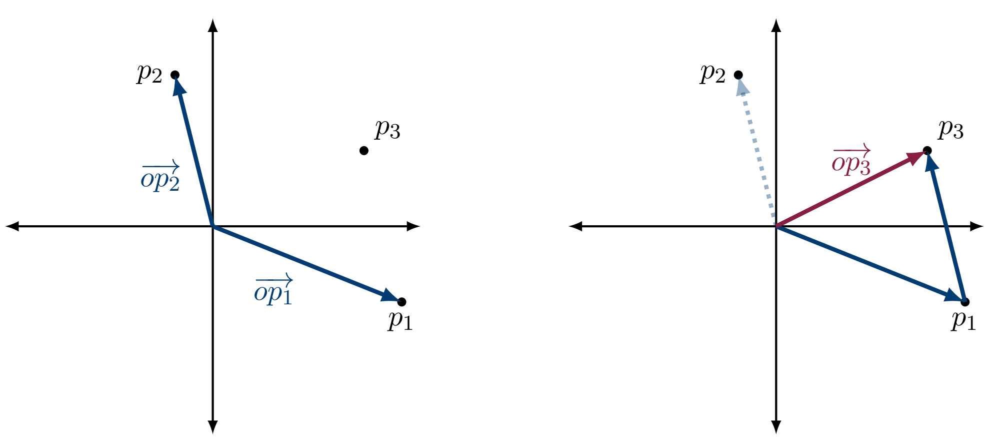


Figure 1.1: The original vectors (left) and "head-to-tail" vector addition (right).

With p_1 as before, consider some real number r. By the scalar multiplication operation described in Example 1.1.2, 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 1.2)

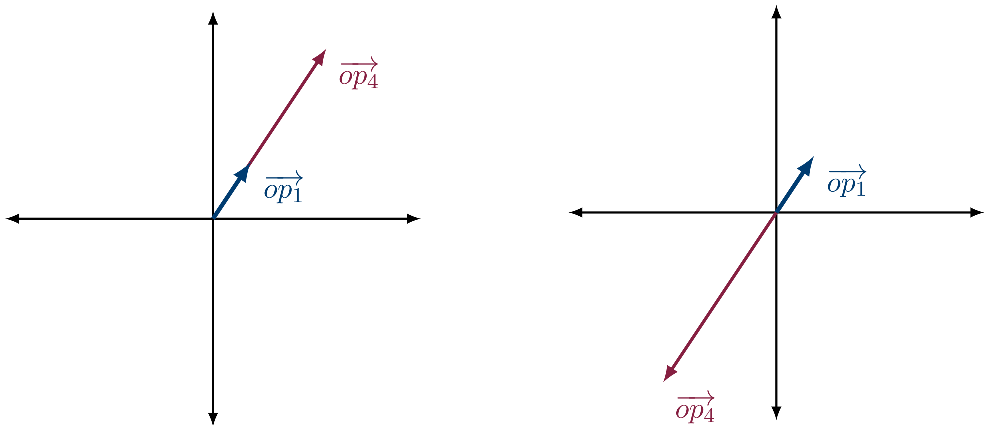


Figure 1.2: The original vector scaled by $r > 0$ (left) and $r < 0$ (right).

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.

1.1.2 Definitions and Examples

Notation: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\in$

We introduce the following notation:

- $\mathbb{Z}$ – this denotes the integers.
- $\mathbb{Q}$ – this denotes the rational numbers (i.e. fractions).
- $\mathbb{R}$ – this denotes the real numbers.
- $\in$ – this means "is contained in the set".
  Example usage would be: $x \in \mathbb{R}$, which translates to "x is a real number."
  or $x \in \{1, 2, 3\}$ means "x is one of the numbers in the list $\{1, 2, 3\}$."

Definition: (real) vector space

A **(real) vector space** is a set V of objects called **vectors** endowed with two operations:

- **(vector) addition**, denoted $+$
- **scalar multiplication**, denoted (no symbol)

satisfying the following properties: For all vector $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and for all real numbers k, ℓ called **scalars** $k, \ell \in \mathbb{R}$:

1. [closure of addition] $\mathbf{u} + \mathbf{v} \in V$
2. [commutativity of addition] $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. [associativity of addition] $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. [existence of zero] There is some vector $\mathbf{0}$, called the **zero vector** so that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
5. [existence of additive inverses] For each $\mathbf{v}$ in V, there is some vector $-\mathbf{v}$ for which $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
6. [closure of scalar multiplication] $k\mathbf{v} \in V$
7. [associativity of scalar multiplication] $(k\ell)\mathbf{v} = k(\ell\mathbf{v})$
8. [distributivity] $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
9. [distributivity] $(k + \ell)\mathbf{u} = k\mathbf{u} + \ell\mathbf{u}$
10. [existence of a multiplicative identity] If $1_{\mathbb{K}}$ is the multiplicative identity in $\mathbb{R}$, then $1_{\mathbb{K}}\mathbf{v} = \mathbf{v}$.

Remark. For a real vector space, $\mathbb{R}$ is sometimes referred to as the **field of scalars**.

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: $\mathbb{R}$ is a vector space.

The real numbers, denoted $\mathbb{R}$, form a real vector space when endowed with the normal addition and multiplication operations.

It is an exercise to the reader to show that this satisfies all of the vector space axioms.

Example 1.1.2: The xy-plane is a vector space.

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.

It is an exercise to the reader to show that this satisfies all of the vector space axioms.

Example 1.1.3: Polynomials are a vector space.

The set of all polynomials with real coefficients and degree at most n (these are polynomials of the form $a_n x^n + \cdots + a_1 x + a_0$), denoted $\mathcal{P}_n(\mathbb{R})$, is a vector space when considered the usual addition and scalar multiplication.

- addition:

$$(a_n x^n + \cdots + a_0) + (b_n x^n + \cdots + b_0) = (a_n + b_n)x^n + \cdots + (a_0 + b_0)$$

- scalar multiplication:
$$r(a_nx^n + \cdots + a_0) = (ra_n)x^n + (ra_0)$$

The number 0 (i.e. the constant polynomial 0) is the zero vector in this space, and this space is sometimes denoted $\mathcal{P}_n(\mathbb{R})$.

It is an exercise to the reader to show that this satisfies all of the vector space axioms.

Example 1.1.4: Continuous functions are a vector space

The set of all continuous real-valued functions on $\mathbb{R}$ (these are functions $f : \mathbb{R} \to \mathbb{R}$), denoted $C(\mathbb{R})$ 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 constant function $f(x) = 0$ is the zero vector in this space, and this space is denoted $C(\mathbb{R})$.

It is an exercise to the reader to show that this satisfies all of the vector space axioms.

Notation: $\mathbb{R}^n$

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

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

In each array above, the real number x_i is called the $i^{\textbf{th}}$ **component** of the array.

Remark. Order matters. Arrays are always ordered left-to-right, top-to-bottom.

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: row/column representations of vectors

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: linear combination

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

$$\mathbf{u} = r_1\mathbf{v}_1 + \cdots + r_k\mathbf{v}_k.$$

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

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. We also note that there is no obvious relationship between k and n in the definition above. It could be that $k = n$, $k \leq n$, or that $k > n$.

Definition: standard basis vectors

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

Show that the vector $\mathbf{u} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$ is a linear combination of the standard basis vectors, $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$.

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

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

Show that 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

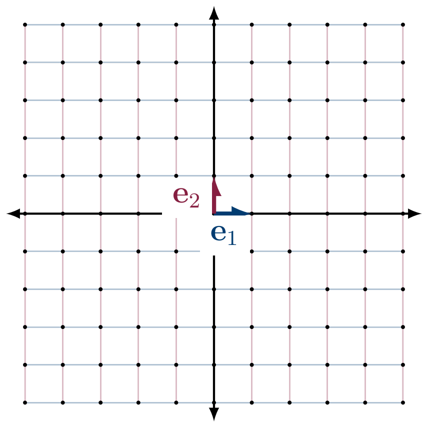
The reader is probably familiar with the Cartesian grid, which provides a useful geometric depiction of the algebra. We similarly want to construct a grid that is uniquely suited to a given set of vectors in $\mathbb{R}^n$.

Definition: coordinate grid (nonstandard terminology)

The **coordinate grid** associated to a collection of vectors $\mathbf{v_1}, \ldots, \mathbf{v_n} \in \mathbb{R}^n$ is the grid formed from all *integer* linear combinations of $\mathbf{v_1}, \ldots, \mathbf{v_n}$.

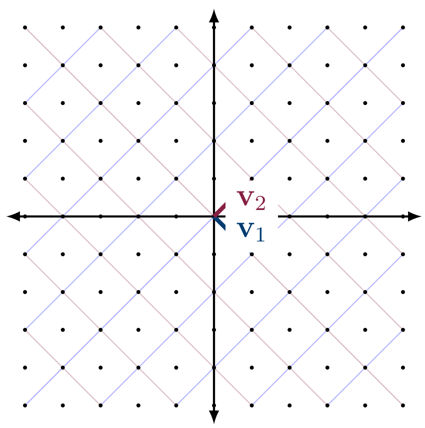
Example 1.1.8

Draw the coordinate grid in $\mathbb{R}^2$ formed from the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. See that it is the usual Cartesian grid.



Example 1.1.9

Draw the coordinate grid in $\mathbb{R}^2$ formed from the vectors $\mathbf{v}_1 = [1, -1]^T$ and $\mathbf{v}_2 = [1, 1]^T$.

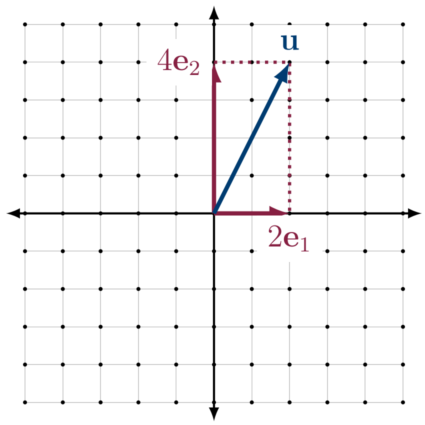


Combined with the geometric intuition about vector addition and scalar multiplication, these coordinate grids provide us with a way to visually identify the linear combination.

Example 1.1.10

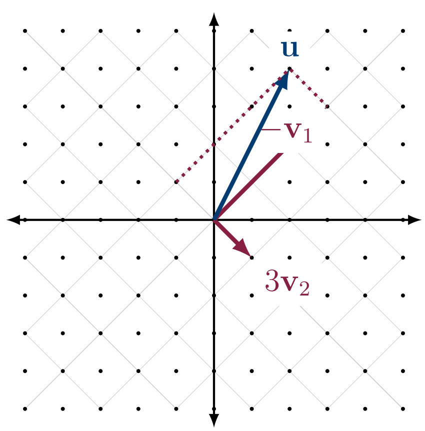
Draw a picture which shows how the vector $\mathbf{u} = [2, 4]^T$ is a linear combination of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

$$\mathbf{u} = 2\mathbf{e}_1 + 4\mathbf{e}_2$$



Example 1.1.11

Draw a picture showing the vector $\mathbf{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ as a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Determine the precise algebraic linear combination.



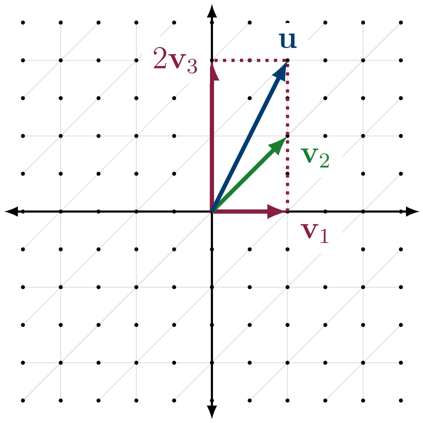
From the above figure it follows that

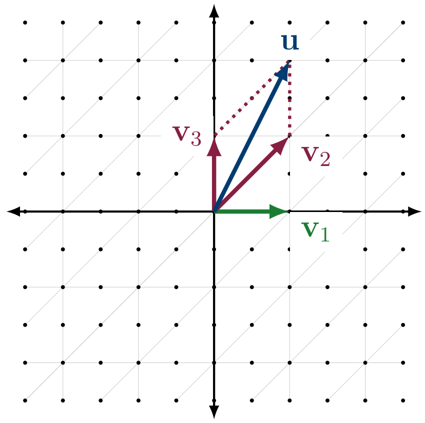
$$\mathbf{u} = 3\mathbf{v}_1 - \mathbf{v}_2.$$

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

Example 1.1.12

Show that $\mathbf{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ is 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.





From the above we see that

$$\mathbf{u} = \mathbf{v_1} + 2\mathbf{v_3} \quad \text{and} \quad \mathbf{u} = \mathbf{v}_2 + \mathbf{v}_3.$$

1.2 Length and Angle: The Dot Product

Definition: Dot Product

For vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, the **dot product** of $\mathbf{u}$ and $\mathbf{v}$ 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.*

Remark. When $\mathbf{u}, \mathbf{v}$ are written as column vectors, the product

$$\mathbf{u}^T \mathbf{v} = [u_1 v_1 + \cdots + u_n v_n]$$

is a vector in $\mathbb{R}^1$, so by identifying $\mathbb{R}^1$ with $\mathbb{R}$, we have that $\mathbf{u}^T \mathbf{v} = \mathbf{u} \cdot \mathbf{v}$ (the product $\mathbf{u}^T \mathbf{v}$ is called an *inner product*).

The dot product has the following nice properties.

Theorem 1.2.1: Properties of the Dot Product

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 $\mathbf{u} = \mathbf{0}$.

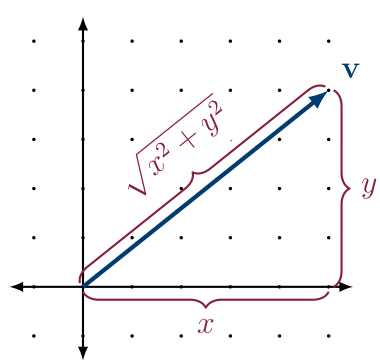
Proof. The proof is entirely straightforward and left as an exercise to the reader. □

1.2.1 Length

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

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

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



Definition: length

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: Properties of Length

For $\mathbf{v} \in \mathbb{R}^n$ and a scalar k,

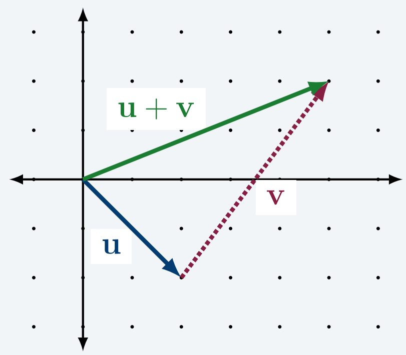
1. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
2. $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.

By thinking of vector addition in terms of triangles, we observe the following classical geometry fact, restated in terms of our dot product/length:

Theorem 1.2.3: Triangle Inequality

For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

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



Definition: unit vector

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.

Proposition 1.2.4: normalizing a vector

For any nonzero vector $\mathbf{v} \in \mathbb{R}^n$, the vector $\mathbf{u} = \dfrac{\mathbf{v}}{\|\mathbf{v}\|}$ is a unit vector in the direction of $\mathbf{v}$.

Proof. 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

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

□

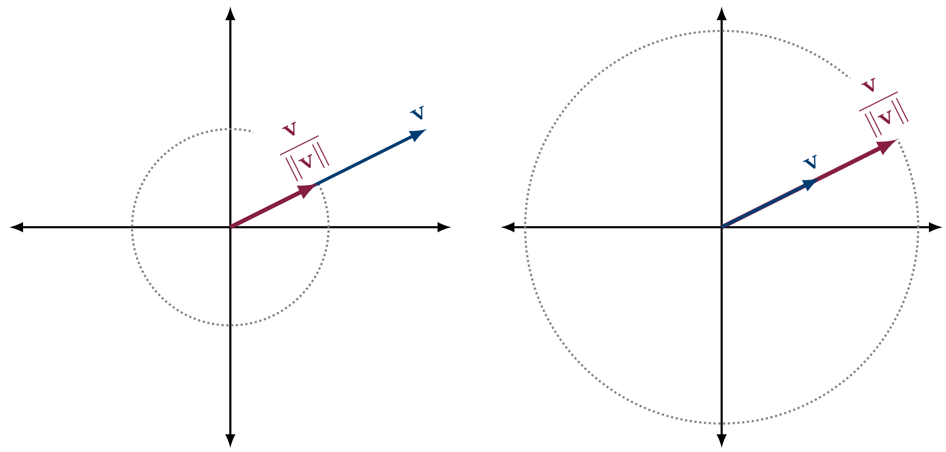


Figure 1.3: If $\|\mathbf{v}\| > 1$ (pictured on the left), normalization effectively *shrinks* the vector. If $\|\mathbf{v}\| < 1$ (pictured on the right), normalization effectively *stretches* the vector.

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

$$d(P, Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

If we identify the point $P(x_1, y_1)$ with the vector $\mathbf{u} = [x_1, y_1]^T$ and the point $Q(x_2, y_2)$ with the vector $\mathbf{v} = [x_2, y_2]^T$, 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.

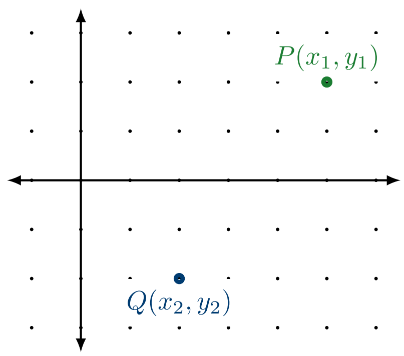
Definition: distance between vectors

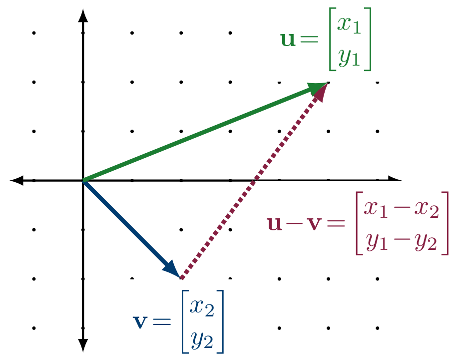
Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the **distance** between $\mathbf{u}$ and $\mathbf{v}$ is

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

Example 1.2.5: relationship to the classical distance formula

1. Let $P(x_1, y_1)$ and $Q(x_2, y_2)$ be points in the plane. Compute the distance between P and Q using the classical distance formula.
2. Let $\mathbf{v} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ be vectors in $\mathbb{R}^2$. Compute the distance between vectors $\mathbf{u}$ and $\mathbf{v}$ using the given formula.





1. INCOMPLETE
2. INCOMPLETE

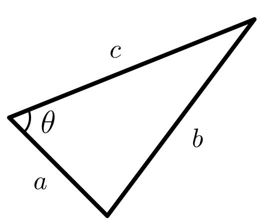
Remark. The above example shows that the distance between vectors is measuring the distance between the heads of the vectors.

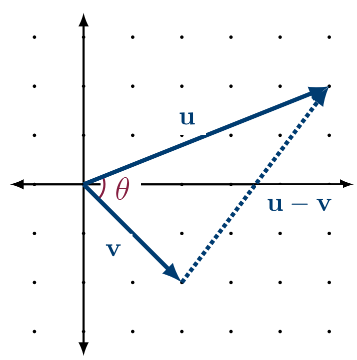
1.2.3 Angles

Definition: angle between vectors

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

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





This is a straightforward derivation using the law of cosines:

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

Replacing the triangle on the left 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 gives the desired formula.

Example 1.2.6

Compute the angle between the vectors $\mathbf{u} = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

From the definition, 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 result is an immediate consequence of the definition of the angle between vectors.

Corollary 1.2.7

Two nonzero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are perpendicular if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Chapter 2

System of Linear Equations

2.1 Introduction to Linear Systems

Definition: (system of) linear equations

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

$$a_1 x_1 + \cdots + a_n x_n = b$$

where $a_1, \dots, a_n$ are real numbers called **coefficients** and b is a real number called the **constant** term.
A **system of linear equations** is a finite set of linear equations, each with the same variables (and probably different coefficients).

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

Definition: solution (set)

A **solution** of a system of linear equations is a vector that $\mathbf{v} = [v_1, \dots, v_n]^T$ that simultaneously satisfies every equation in the system. In other words,

$$\begin{cases} a_{11}v_1 & + & a_{12}v_2 & \cdots & a_{1n}v_n & = & b_1 \\ a_{21}v_1 & + & a_{22}v_2 & \cdots & a_{2n}v_n & = & b_2 \\ \vdots & + & \vdots & \vdots & \vdots & = & \vdots \\ a_{m1}v_1 & + & a_{m2}v_2 & \cdots & a_{mn}v_n & = & b_m \end{cases}$$

The **solution set** of a linear system is the set of all solutions for a given system.

Naturally-Occuring Linear Systems

Example 2.1.1: linear combinations

Consider the vectors used in Example 1.1.7:

$$\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{v_1} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{v_2} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Write down a linear system whose solutions would yield the coefficients necessary to write $\mathbf{u}$ as a linear combination of $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$. Write down some solution vectors to this linear system.

To write $\mathbf{u}$ as a linear combination of $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$, we solve for x_1, x_2, x_3 in the following equation:

$$\mathbf{u} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3}$$

which we can expand as

$$\begin{aligned} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} &= x_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ -x_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} -x_2 \\ x_2 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_1 + x_2 + x_3 \end{bmatrix} \end{aligned}$$

and this equality of vectors holds if and only if the following system is satisfied:

$$\begin{cases} x - y &= 1 \\ -x + y &= -1 \\ x + y + z &= 0 \end{cases}$$

From Example 1.1.7, the following are each solutions to this system:

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ -3 \\ 5 \end{bmatrix}.$$

Example 2.1.2: linear combinations of polynomials

Let p_1, p_2, p_3, p_4 be the following polynomials:

$$\begin{aligned} p_1(x) &= 1 - 2x + 3x^2 - 4x^3 \\ p_2(x) &= 5x + 6x^3 \\ p_3(x) &= 2 - 9x + 6x^2 - 14x^3 \\ p_4(x) &= x^3 \end{aligned}$$

Write a linear system whose solution gives the coefficients for writing p_1 as a linear combination of p_2, p_3, p_4.

Let a, b, c be some unknown scalars. We want to write

$$\begin{aligned} p_1(x) &= ap_2(x) + bp_3(x) + cp_4(x) \\ 1 - 2x + 3x^2 - 4x^2 &= a(5x + 6x^3) + b(2 - 9x + 6x^2 - 14x^3) + cx^3 \end{aligned}$$

which rearranges to

$$1 - 2x + 3x^2 - 4x^2 = 2b + (5a - 9b)x + 6bx^2 + (6a - 14b + c)x^3$$

Two polynomials are equal if and only if their corresponding coefficients are equal, so comparing coefficients, we get

$$\begin{cases} \quad\quad 2b \quad\quad = 1 \\ 5a - 9b \quad\quad = -2 \\ \quad\quad 6b \quad\quad = 3 \\ 6a - 14b + c = -4 \end{cases}$$

Example 2.1.3: dot products

Let $\mathbf{v_1} = [9, 8, 7]^T$ and $\mathbf{v_2} = [6, 5, 4]^T$. Write down a linear system whose solutions will be vectors perpendicular to both $\mathbf{v_1}$ and $\mathbf{v_2}$. Verify that $[1, -2, 1]$ is a solution to this system.

Recall Corollary 1.2.7that two vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ are perpendicular if and only if $\mathbf{u_1} \cdot \mathbf{u_2} = 0$.

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ be a vector that simultaneously satisfies $\mathbf{v_1} \cdot \mathbf{x} = 0$ and $\mathbf{v_2} \cdot \mathbf{x} = 0$. This gives us

$$\begin{cases} \mathbf{v_1} \cdot \mathbf{x} = 0 \\ \mathbf{v_2} \cdot \mathbf{x} = 0 \end{cases}$$

which produces the system

$$\begin{cases} 9x_1 + 8x_2 + 7x_3 = 0 \\ 6x_1 + 5x_2 + 4x_3 = 0 \end{cases}$$

It is straightforward to plug in $x_1 = 1$, $x_2 = -2$, and $x_3 = 1$ into the system above and verify that each equation is true.

Example 2.1.4: functions

Let f be the function given below, whose domain is $\mathbb{R}^3$ and range is $\mathbb{R}^2$:

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_2 - x_3 \end{bmatrix}.$$

Write down a system which finds all vectors $\mathbf{x} \in \mathbb{R}^3$ for which $f(\mathbf{x}) = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$.

If we want to find all vectors $\mathbf{v}$ in $\mathbb{R}^3$ for which $f(\mathbf{v}) = [3, 7]^T$, we begin by writing this out

$$f\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$
$$\begin{bmatrix} v_1 + v_2 \\ v_2 - v_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$

and this last equation produces a linear system with two equations and three variables v_1, v_2, v_3:

$$\begin{cases} v_1 + v_2 &= 3 \\ v_2 - v_3 &= 7 \end{cases}$$

Example 2.1.5: curve fitting

Let p be a cubic polynomial

$$p(x) = c_3x^3 + c_2x^2 + c_1x^1 + c_0.$$

Write down a linear system whose solutions are the polynomials whose graphs $y = p(x)$ pass through the points $(-2, 3)$, $(-1, -6)$, $(1, 0)$, $(3, -2)$. Then verify that

$$p(x) = -x^3 + 2x^2 + 4x - 5$$

is a solution to this system.

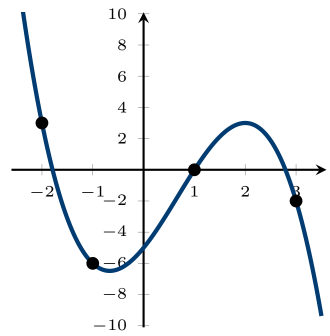
We are searching for a polynomial for which $p(-2) = 3$, $p(-1) = -6$, $p(1) = 0$, $p(3) = -2$. That is,

$$\begin{cases} p(-2) &= 3 \\ p(-1) &= -6 \\ p(1) &= 0 \\ p(3) &= -2 \end{cases}$$

which produces the system

$$\begin{cases} -8c_3 + 4c_2 - 2c_1 + c_0 &= 3 \\ -c_3 + c_2 - c_1 + c_0 &= -6 \\ c_3 + c_2 + c_1 + c_0 &= 0 \\ 27c_3 + 9c_2 + 3c_1 + c_0 &= -2 \end{cases}$$

It is straightforward to plug in $c_3 = -1$, $c_2 = 2$, $c_1 = 4$ and $c_0 = -5$ in each equation above to verify that they are correct.



Example 2.1.6

Show that, for every real number t, the vector $\mathbf{v} = \begin{bmatrix} 0 \\ 3 \\ -4 \end{bmatrix} + t \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$ is a solution to the system

$$\begin{cases} x_1 + x_2 & = & 3 \\ x_2 - x_3 & = & 7 \end{cases}$$

Recall that $\mathbf{v} = [x_1, x_2, x_3]^T$ (or at least, this is implied). Rewriting $\mathbf{v}$ slightly, we have

$$\mathbf{v} = \begin{bmatrix} 0 \\ 3 \\ -4 \end{bmatrix} + t \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} t \\ 3-t \\ -4-t \end{bmatrix}$$

whence $x_1 = t$, $x_2 = 3 - t$, $x_3 = -4 - t$. One then quickly checks that

$$\begin{cases} (t) + (3-t) & = & 3 \\ (3-t) - (-4-t) & = & 7 \end{cases}$$

and thus, no matter the value of t, $\mathbf{v}$ (as above) is a solution to the given system.

Definition: parametric form

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 $\mathbf{v}_i$. The t_i terms above are called **parameters**.

Solution Set Analysis

Definition: (in)consistency

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

Fact. A system of linear equations with real coefficients has one of three possible solution sets:

(a) a unique solution (consistent)

(b) infinitely many solutions (consistent)

(c) no solutions (inconsistent)

You can convince yourself of the above trichotemy by considering how it works for systems of equations with two variables (whose solution sets are graphically lines in the Cartesian plane). Two lines can either intersect in a single point (if they are transverse), intersect in infinitely many points (if they coincide), or no points (if they are parallel). The proof is also very straightforward, but happens to rely on some algebraic manipulations that aren't formally covered until Chapter 3. We've recorded it here for posterity.

Proof. Suppose $A\mathbf{x} = \mathbf{b}$ has two solutions, $\mathbf{x_1}$ and $\mathbf{x_2}$. That is, suppose

$$A\mathbf{x_1} = \mathbf{b} \tag{2.1}$$

$$\text{and} \qquad A\mathbf{x_2} = \mathbf{b}. \tag{2.2}$$

The claim is that, for any scalar t, $(1-t)\mathbf{x_1} + t\mathbf{x_2}$ is also a solution. Indeed,

$$\begin{aligned} A[(1-t)\mathbf{x_1} + t\mathbf{x_2}] &= A(1-t)\mathbf{x_1} + At\mathbf{x_2} && \text{(distributive property)} \\ &= (1-t)A\mathbf{x_1} + tA\mathbf{x_2} && \text{(commutativity scalar multiplication)} \\ &= (1-t)\mathbf{b} + t\mathbf{b} && \text{(Equation 2.1 and 2.2)} \\ &= [(1-t)+t]\mathbf{b} && \text{(distributive property)} \\ &= \mathbf{b}. \end{aligned}$$

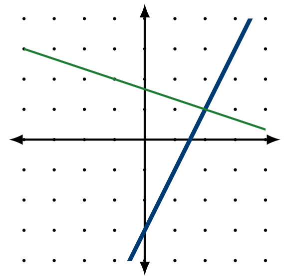
□

Example 2.1.7

Determine whether or not the following system is consistent. If it is consistent, how many solutions does it have?

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

Each equation in the linear system can be interpreted as a line in the plane. So first we draw these lines.



The lines intersect at the point $(2, 1)$, and it is straightforward to verify that $[2, 1]^T$ is a solution, hence the system is consistent.

Using ad-hoc methods, we see that the first equation is only true when $y = 2x - 3$. Any solution to the second equation must thus satisfy $x + 3(2x - 3) = 5$, which is true precisely when $x = 2$. Using this value in the equation $y = 2x - 3$ produces $y = 1$. There are no other options for solutions, hence it has a unique solution.

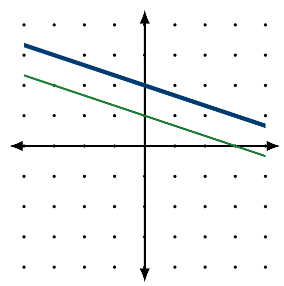
Remark. Solutions to a linear system (with two variables) correspond to points of intersections between the lines represented by the equations. Equations with 3 variables correspond to planes in 3-dimensional space, and solutions correspond again to intersections of these planes.

Example 2.1.8

Determine whether or not the following system is consistent. If it is consistent, how many solutions does it have?

$$\begin{cases} x + 3y = 6 \\ 2x + 6y = 6 \end{cases}$$

Plotting the lines as in the previous example,



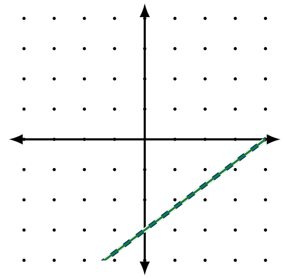
We see that the lines are parallel. It follows then that there are no solutions (because there are no points of intersection), hence the system is inconsistent.

Example 2.1.9

Determine whether or not the following system is consistent. If it is consistent, how many solutions does it have?

$$\begin{cases} 3x - 4y = 12 \\ -6x + 8y = -24 \end{cases}$$

Plotting the lines as in the previous example,



Since the two lines intersect at every point, there are infinitely-many solutions to the system.

Definition: equivalent linear systems

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

Example 2.1.10: equivalent systems

Determine which, if any, of the following systems are equivalent.

(a) $\left\{\begin{array}{rcrcr} x & + & 3y & = & 5 \\ 2x & - & y & = & 3 \end{array}\right.$

(c) $\left\{\begin{array}{rcrcr} x & - & 3y & = & 6 \\ 2x & - & y & = & 3 \end{array}\right.$

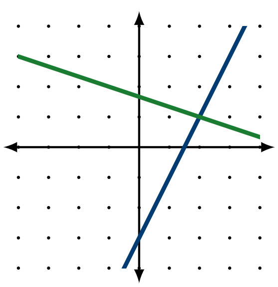
(e) $\left\{\begin{array}{rcrcr} x & + & 3y & = & 5 \\ & & -7y & = & -7 \end{array}\right.$

(b) $\left\{\begin{array}{rcrcr} 2x & - & y & = & 3 \\ x & + & 3y & = & 5 \end{array}\right.$

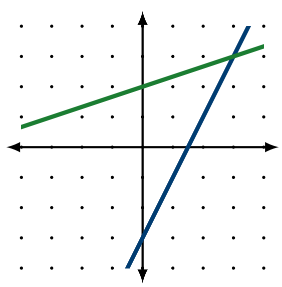
(d) $\left\{\begin{array}{rcrcr} 10x & + & 30y & = & 50 \\ 2x & - & y & = & 3 \end{array}\right.$

(f) $\left\{\begin{array}{rcrcr} 10x & + & 30y & = & 50 \\ x & + & 3y & = & 5 \end{array}\right.$

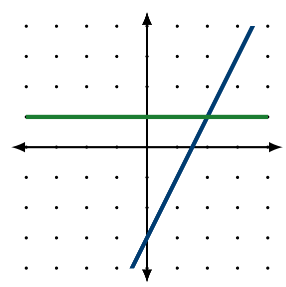
(a)



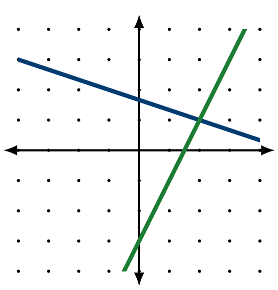
(c)



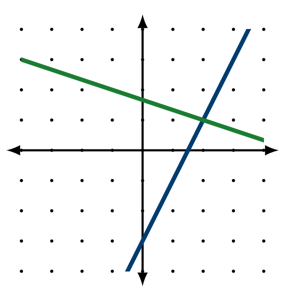
(e)



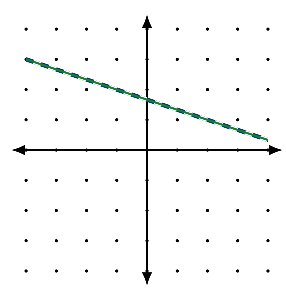
(b)



(d)



(f)



From the above, we see that (a), (b), (d), and (e) all have the solution set $\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$, hence are equivalent. (c)'s solution set is $\left\{\begin{bmatrix} 3 \\ 3 \end{bmatrix}\right\}$ and (f)'s solution set is $\left\{\begin{bmatrix} 5 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 1 \end{bmatrix} \;:\; \text{where } t \in \mathbb{R}\right\}$.

Definition: augmented/coefficient matrix

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, $\mathbf{x} = [x_1, \ldots, x_n]^T$ is the vector of indeterminates, and $\mathbf{b} = [b_1, \ldots, b_m]^T$ is the column vector of constant terms, we may write $A\mathbf{x} = \mathbf{b}$ to denote the system or $\left[A \mid \mathbf{b}\right]$ 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 2.1.11

Solve the linear system represented by the augmented matrix below.

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

THERE IS A TYPO IN ROW 2 BELOW AND THE EXAMPLE WORK IS WRONG This augmented matrix represents the system

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

The "triangular structure" of this system makes it easy to solve. We quickly deduce that

$$8z = 16 \implies z = 2,$$

then

$$2y - 4(2) = 6 \implies y = 7,$$

and finally

$$x + 3(7) + 5(2) = 7 \implies x = -24.$$

So the solution is the vector

$$\begin{bmatrix} -24 \\ 7 \\ 2 \end{bmatrix}.$$

The technique employed in solving the last system is called **back-substitution**.

As we'll see in the next section, our strategy will be to manipulate our systems to obtain this "triangular structure."

2.2 Direct Methods for Solving Linear Systems

2.2.1 Row Operations

In Example 2.1.10 we saw that the following systems were all equivalent, and were derived from the original system.

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

$R_1 \leftrightarrow R_2$:
$$\left[\begin{array}{cc|c} 2 & -1 & 3 \\ 1 & 3 & 5 \end{array}\right]$$

$10R_1 \mapsto R_1$:
$$\left[\begin{array}{cc|c} 10 & 30 & 50 \\ 2 & -1 & 3 \end{array}\right]$$

$-2R_1 + R_2 \mapsto R_2$:
$$\left[\begin{array}{cc|c} 1 & 3 & 5 \\ 0 & -7 & -7 \end{array}\right]$$

Proposition 2.2.1

Let k be any nonzero real number. The following systems all have the same solution sets:

(a) $\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \end{cases}$

(b) $\begin{cases} a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \end{cases}$

(c) $\begin{cases} ka_{11}x_1 + \cdots + ka_{1n}x_n = kb_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \end{cases}$

(d) $\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ (ka_{11} + a_{21})x_1 + \cdots + (ka_{1n} + a_{2n})x_n = kb_1 + b_2 \end{cases}$

Definition: Elementary Row Operations

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

1. **Row swapping**
   Swapping Row i and Row j (denoted $R_i \leftrightarrow R_j$).
2. **Row scaling**
   Multiplying Row i by a nonzero constant (denoted $kR_i \mapsto R_i$).
3. **Row addition**
   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.

Remark. The operations above do not necessarily commute; the order in which they are applied is important.

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: row equivalence

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

Show that the following matrices are row equivalent:

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

$$\begin{aligned} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right] &\xrightarrow{-3R_1+R_2 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{array}\right] \\ &\xrightarrow{-2R_1+R_3 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{array}\right] \\ &\xrightarrow{R_2 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right] \end{aligned}$$

2.2.2 (Reduced) Row Echelon Form

That "triangular structure" which made life so easy in Example 2.1.11 is given a special name.

Definition: row echelon form

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

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

(b) In each nonzero row, the leftmost nonzero entry (the **leading entry** or **pivot**) is in a column to the left of any leading entries below it.

Example 2.2.3

Which of the following matrices are in row echelon form?

(a) $\begin{bmatrix} 1 & 0 & 3 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 5 & 0 & 1 \end{bmatrix}$

(b) $\left[\begin{array}{ccc|c} 2 & 0 & 4 & 0 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$

(c) $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right]$

(d) $\begin{bmatrix} 0 & 3 & 0 & 4 \\ 0 & 0 & 5 & 6 \\ 0 & 0 & 0 & 7 \end{bmatrix}$

(e) $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$

(f) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$

Definition: reduced echelon form

A matrix is in **reduced** row echelon form (RREF) if it satisfies the following properties:

1. The matrix is already in row echelon form.
2. Each leading entry is 1.
3. Any entries *above a leading* 1 are 0.

Example 2.2.4

Which of the following matrices are in reduced row echelon form?

(a) $\begin{bmatrix} 1 & 0 & 3 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 5 & 0 & 1 \end{bmatrix}$

(b) $\left[\begin{array}{ccc|c} 2 & 0 & 4 & 0 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$

(c) $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right]$

(d) $\begin{bmatrix} 0 & 3 & 0 & 4 \\ 0 & 0 & 5 & 6 \\ 0 & 0 & 0 & 7 \end{bmatrix}$

(e) $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$

(f) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$

Theorem 2.2.5

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

Proof. The proof is given by the algorithm below performed on an $m \times n$ matrix. □

Algorithm 2.2.6: Row Reduction

```
  /*PUTTING MATRIX INTO REF*/
i ← 1;   /*Row Number*/
j ← 1;   /*Column Number*/
while i ≤ m and j ≤ n do
    if Column j contains nonzero entries then
        Use row swap to move nonzero entry to Row i;
        Use row addition to make entries below Row i all zero;
        i ← i + 1;
        j ← j + 1;
    else
        j ← j + 1;
    end if
end while
  /*FURTHER PUTTING MATRIX INTO RREF*/
i ← m;   /*Row Number*/
j ← n;   /*Column Number*/
while i ≥ 1 and j ≥ 1 do
    if Column j contains a leading entry then
        Use row scaling to make leading entry 1;
        Use row addition to make entries above Row i all zero;
        i ← i − 1;
        j ← j − 1;
```

```
    else
        j ← j − 1;
    end if
end while
```

Definition: row reduction

The process of putting a matrix into reduced row echelon form is called **row reduction**.

Remark. The row echelon form of a matrix is not unique, but the reduced row echelon form is unique.

Here is a visual of the Row Reduction Algorithm.

Step 1. Look at Rows $1 \ldots m$ in Column 1. If there are any nonzero entries, find it and move it to Row 1. Then use Row Addition to clear everything below it.

$$\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 3 & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{bmatrix}$$

Step 2. Look at Rows $2 \ldots m$ in Column 2. There are no nonzero entries here, so we move onto the next column. We won't change the range of rows.

$$\begin{bmatrix} 3 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{bmatrix}$$

Steps 3 … n. Repeat the above steps for all remaining columns. Now the matrix is in row echelon form.

$$\begin{bmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 19 \end{bmatrix}$$

Step $n+1$. Look at Column n. If there is a leading entry, use Row Scaling to make that leading entry a 1, and then Row Addition to clear everything above it.

$$\begin{bmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 19 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Step $n+2$. Look at Column $n-1$. If there is a leading entry, use Row Scaling to make that leading entry a 1, and then Row Addition to clear everything above it.

$$\begin{bmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & * & * & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Steps $n + 3$ ... $2n$. Repeat the above steps for all remaining columns. Now the matrix is in reduced row echelon form.

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

Remark. As a human, you can inject convenient steps into the algorithm above. You do not have to wait to scale until the end – you can clear fraction denominators/shrink large numbers at any point. You also have some choice when row swapping – if you can choose between multiple rows for a leading entry, pick one that already has a leading 1.

Example 2.2.7

Row reduce the following augmented matrix.

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

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).

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right] \xrightarrow{R_2 - 3R_1 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{array}\right] \tag{2.3}$$

$$\xrightarrow{R_3 - 2R_1 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{array}\right] \tag{2.4}$$

$$\tag{2.5}$$

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.

$$\xrightarrow{R_2 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right] \tag{2.6}$$

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 2.6 is in row echelon form!)
8. Now scale every row so that the leading term is a 1.

$$\xrightarrow{\frac{1}{5}R_3 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.7}$$

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

$$\xrightarrow{R_1+R_2 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 2 & 7 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.8}$$

$$\xrightarrow{R_1-2R_3 \mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.9}$$

$$\xrightarrow{R_2-3R_3 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.10}$$

Theorem 2.2.8

Matrices A and B are row equivalent if and only if they can be row reduced to the same echelon form.

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be the sequence of row operations that row reduces A, and let $\beta_1, \beta_2, \ldots, \beta_\ell$ be the sequence of row operations that row reduces B. In other words,

$$\alpha_k \circ \cdots \circ \alpha_2 \circ \alpha_1(A) = \text{RREF}(A) \qquad \text{and} \qquad \beta_\ell \circ \cdots \circ \beta_2 \circ \beta_1(B) = \text{RREF}(B).$$

If A and B are row equivalent, then there is a sequence of row operations $\sigma_1, \sigma_2, \ldots, \sigma_n$ for which $\sigma_n \circ \cdots \circ \sigma_2 \circ \sigma_1(A) = B$ and therefore

$$\begin{aligned} &\beta_\ell \circ \cdots \circ \beta_1 \circ \sigma_n \circ \cdots \circ \sigma_1 \circ \alpha_1^{-1} \circ \cdots \alpha_k^{-1}(\text{RREF}(A)) \\ &= \beta_\ell \circ \cdots \circ \beta_1 \circ \sigma_n \circ \cdots \circ \sigma_1(A) \\ &= \beta_\ell \circ \cdots \circ \beta_1(B) \\ &= \text{RREF}(B) \end{aligned}$$

Since any row operation to a matrix in reduced row echelon form will take it out of reduced row echelon form, then it must be that $\text{RREF}(A) = \text{RREF}(B)$. Conversely, if $\text{RREF}(A) = \text{RREF}(B)$, then we have a equence of row operations which changes A into B:

$$\begin{aligned} &\beta_1^{-1} \circ \cdots \circ \beta_\ell^{-1} \circ \alpha_k \circ \cdots \circ \alpha_1(A) \\ &= \beta_1^{-1} \circ \cdots \circ \beta_\ell^{-1}(\text{RREF}(A)) \\ &= \beta_1^{-1} \circ \cdots \circ \beta_\ell^{-1}(\text{RREF}(B)) \\ &= B \end{aligned}$$

hence A and B are row equivalent. □

2.2.3 Gaussian Elimination and Gauss–Jordan Elimination

Definition: leading/free variables

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

Definition: Gaussian/Gauss–Jordan elimination

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.

Gauss–Jordan elimination is the same process with the second step is replaced by the reduced row echelon form.

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

Example 2.2.9

Use Gaussian or Gauss–Jordan elimination to find the solution set for the given system. If there are infinitely-many solutions, put the answer in parametric form.

$$\begin{cases} \phantom{x_1 +{}} 4x_2 + x_3 = 21 \\ x_1 + 4x_2 + x_3 = 24 \\ x_1 + 2x_2 + x_3 = 16 \end{cases}$$

INCOMPLETE - ROW STEPS NOT EXPLAINED Applying Gauss–Jordan, we rewrite the system as an augmented matrix and row reduce.

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & \end{matrix} \\ \left[\begin{array}{ccc|c} 0 & 4 & 1 & 21 \\ 1 & 4 & 1 & 24 \\ 1 & 2 & 1 & 16 \end{array}\right] \end{array} \to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 16 \\ 1 & 4 & 1 & 24 \\ 0 & 4 & 1 & 21 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 16 \\ 0 & 2 & 0 & 8 \\ 0 & 4 & 1 & 21 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 16 \\ 0 & 1 & 0 & 4 \\ 0 & 4 & 1 & 21 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 16 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 0 & 11 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

$$\xrightarrow{\text{RREF}} \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & \end{matrix} \\ \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right] \end{array}$$

We have leading entries in columns 1, 2, and 3, so Our leading variables are x_1, x_2, x_3. Thus we solve the simpler system

$$\begin{cases} x_1 & & & = 3 \\ & x_2 & & = 4 \\ & & x_3 & = 5 \end{cases}$$

To get a solution of $[x_1, x_2, x_3] = [3, 4, 5]^T$.

Example 2.2.10

Use Gaussian or Gauss–Jordan elimination to find the solution set for the given system. If there are infinitely-many solutions, put the answer in parametric form.

$$\begin{cases} -4x_1 + x_2 - 4x_3 = -7 \\ 2x_1 + 4x_3 = 6 \\ -2x_1 + x_2 = -1 \end{cases}$$

INCOMPLETE - ROW STEPS NOT EXPLAINED Applying Gauss–Jordan, we rewrite the system as an augmented matrix and row reduce.

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & \end{matrix} \\ \left[\begin{array}{ccc|c} -4 & 1 & -4 & -7 \\ 2 & 0 & 4 & 6 \\ -2 & 1 & 0 & -1 \end{array}\right] \end{array} \rightarrow \left[\begin{array}{ccc|c} -4 & 1 & -4 & -7 \\ 1 & 0 & 2 & 3 \\ -2 & 1 & 0 & -1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ -4 & 1 & -4 & -7 \\ -2 & 1 & 0 & -1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ -2 & 1 & 0 & -1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 1 & 4 & 5 \end{array}\right]$$

$$\xrightarrow{\text{RREF}} \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & \end{matrix} \\ \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

We have leading variables in columns 1 and 2, so our leading variables are x_1 and x_2. We have free variable x_3. We thus solve the simpler system, writing our leading variables in terms of our free variables

$$\begin{cases} x_1 \quad + 2x_3 = 3 \\ x_2 + 4x_3 = 5 \end{cases} \rightarrow \begin{cases} x_1 = -2x_3 + 3 \\ x_2 = -4x_3 + 5 \end{cases}$$

And thus our solutions are vectors of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 & +3 \\ -4x_3 & +5 \\ x_3 & \end{bmatrix} = \begin{bmatrix} -2x_3 \\ -4x_3 \\ x_3 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ -4 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}.$$

Since x_3 can be any real number t, our solution set is thus all of the vectors in the parametric form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -2 \\ -4 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}.$$

Example 2.2.11

Use Gaussian or Gauss–Jordan elimination to find the solution set for the given system. If there are infinitely-many solutions, put the answer in parametric form.

$$\begin{cases} -2x_1 \quad\quad - 4x_3 = -6 \\ x_1 - x_2 - 2x_3 = 4 \\ x_2 + 4x_3 = 5 \end{cases}$$

INCOMPLETE - ROW STEPS NOT EXPLAINED Applying Gauss–Jordan, we rewrite the system as an augmented matrix and row reduce.

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & \end{matrix} \\ \left[\begin{array}{ccc|c} -2 & 0 & -4 & -6 \\ 1 & -1 & -2 & 4 \\ 0 & 1 & 4 & 5 \end{array}\right] \end{array} \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 1 & -1 & -2 & 4 \\ 0 & 1 & 4 & 5 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & -1 & -4 & 1 \\ 0 & 1 & 4 & 5 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & -1 & -4 & 1 \\ 0 & 0 & 0 & 6 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & -1 & -4 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & -1 & -4 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{\text{RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Without determining leading/free variables, we examing this new equivalent system

$$\begin{cases} x_1 \quad\quad + 2x_3 = 0 \\ \quad\quad x_2 + 4x_3 = 0 \\ \quad\quad\quad\quad\quad 0 = 1 \end{cases}$$

and there are no values of x_1, x_2, x_3 which can ever make the third equation true. Hence there are no solutions to this system.

2.2.4 Rank and Number of Solutions

Definition: rank

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

After row reduction, we saw the following in the last three examples:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

Example 2.2.9
Unique Solution

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Example 2.2.10
Infinitely-Many Solutions

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 6 \end{array}\right]$$

Example 2.2.11
No Solution

With the rank of each of these matrices in mind, we see the following result.

Theorem 2.2.12: Using rank to determine consistency

A linear system $[A|\mathbf{b}]$ is consistent if and only if

$$\text{rank}(A) = \text{rank}([A|\mathbf{b}]).$$

Corollary 2.2.13

A linear system $[A|\mathbf{0}]$ is always consistent.

Definition: homogeneous system

A linear system of the form $A\mathbf{x} = \mathbf{0}$ or $[A|\mathbf{0}]$ is called **homogeneous**.

Example 2.2.14

Use Gauss-Jordan elimination to find the solution set for the following homogeneous 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

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

Observation

Given any linear system,

total number of variables = leading variables + free variables.

After row reduction, we saw the we again examine three recent consistent examples:

$\left[\begin{array}{ccc\|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right]$	$\left[\begin{array}{ccc\|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right]$	$\left[\begin{array}{cccc\|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -2 & -3 & 0 \end{array}\right]$
Example 2.2.9 Unique Solution 0 free variables	Example 2.2.10 Infinitely-Many Solutions 1 free variable	Example 2.2.14 Infinitely-Many Solutions 2 free variables
0 parameter solution set	1 parameter solution set	2 parameter solution set

We thus observe the following:

Observation

The number of free variables in a linear system is equal to the number of parameters in the solution set.

With the rank of each of these matrices in mind and combining with the above observations, we obtain

Theorem 2.2.15: The Rank Theorem

If $[A|\mathbf{b}]$ is a consistent system of linear equations with n variables, then

$$n = \text{rank}(A) + \text{number of parameters in solution set.}$$

Example 2.2.16

Using the rank, determine whether or not the following system is consistent or inconsistent. If it is consistent, determine how many solutions it has. Verify your answer by finding 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

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

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

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

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

We observe that $\text{rank}([A|\mathbf{b}]) = \text{rank}(A) = 2$, hence Theorem 2.2.12 implies this system is consistent. Moreover, since we have 3 total variables, Theorem 2.2.15 implies that we have at least one parameter (i.e. infinitely-many solutions). To obtain such solutions explicitly, we work with the simpler system:

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

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

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

2.3 Spanning Sets and Linear Independence

With the lone exception of $\mathbb{R}^0 = \{\mathbf{0}\}$, every vector space has infinitely-many vectors in it, which is way too many to keep track of at all times. An effective workaround would be to store data about a finite number of objects and some procedure for recovering all important information about the vectors in $\mathbb{R}^n$. In particular, what we want to find is a finite collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ (a "basis") so that every vector in $\mathbb{R}^n$ can be written as a linear combination of these vectors (a "spanning set") and that each linear combination is unique ("linear independence"). In this way, given any vector in $\mathbb{R}^n$, our algorithmic methods for solving linear systems can be employed to recover the linear combinations.

2.3.1 Span and Spanning Sets

Definition: span, spanning set

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 <u>all</u> linear combinations of these vectors, and we write

$$\operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k}) \text{ or } \operatorname{Span}(S).$$

If $V = \operatorname{Span}(S)$, then we call S a **spanning set** for V.

Example 2.3.1

The system in Example 2.2.14 had the solution set

$$\left\{ s\begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 3 \\ 0 \\ 1 \end{bmatrix} \text{ for all real numbers } s \text{ and } t \right\}.$$

Describe this solution set using the new "Span" notation.

Observe that the description of the set yields all linear combinations of the two vectors in $\mathbb{R}^4$. The span is, by definition, the set of all linear combinations, so this set can be written more compactly as

$$\operatorname{Span}\left(\begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 0 \\ 1 \end{bmatrix} \right).$$

As we saw in Example 2.1.1, we can rewrite a linear system

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

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 2.3.2: *Poole* Theorem 2.4

A system of linear equations $\left[A \mid \mathbf{b}\right]$ is consistent if and only if

$$\mathbf{b} \in \operatorname{Span}(\mathbf{a_1}, \ldots, \mathbf{a_n}),$$

where $\mathbf{a_i}$ is the i^{th} column of A.

2.3.2 Understanding Span Geometrically

Example 2.3.3: Spans are vector spaces too

Show that, if V is a (real) vector space and $\mathbf{v_1}, \ldots, \mathbf{v_k}$ is some collection of vectors in V, then the $\operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is also (real) vector space.

Observe that vector addition and scalar multiplication behave in obvious ways. Let $\mathbf{w_1}$, $\mathbf{w_2}$ be vectors in $\operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k}$, that is

$$\mathbf{w_1} = a_1\mathbf{v_1} + \cdots + a_k\mathbf{v_k}$$
$$\mathbf{w_2} = b_1\mathbf{v_1} + \cdots + b_k\mathbf{v_k}$$

Then we have that

$$\mathbf{w_1} + \mathbf{w_2} = (a_1 + b_1)\mathbf{v_1} + \cdots + (a_k + b_k)\mathbf{v_k}.$$

Also, for any scalar r,

$$r\mathbf{w_1} = (ra_1)\mathbf{v_1} + \cdots + (ra_k)\mathbf{v_k}.$$

It's clear that the vectors $\mathbf{w_1} + \mathbf{w_2}$ and $r\mathbf{w_1}$ are contained in $\operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$. This proves the closure axioms of a vector space, and verification of the remaining axioms is a tedious (but totally straightforward) exercise.

Example 2.3.4

Let $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ be the standard basis vectors for $\mathbb{R}^2$. Give an intuitive description of $\operatorname{Span}(\mathbf{e_1}, \mathbf{e_2})$. Then justify your description using Theorem 2.3.2.

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

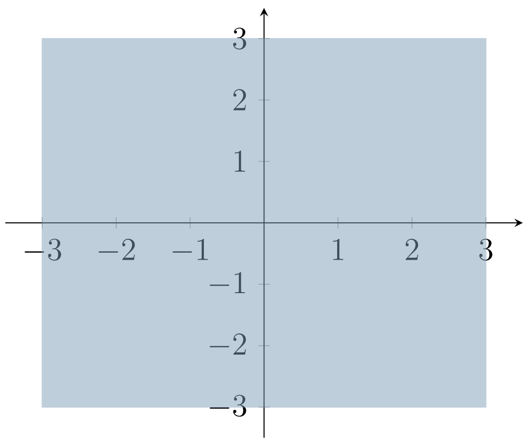


Figure 2.1: The span of $\mathbf{e_1}$ and $\mathbf{e_2}$ in $\mathbb{R}^2$.

To verify that this span is actually $\mathbb{R}^2$, we need to show that *every vector* $\begin{bmatrix}a\\b\end{bmatrix} \in \text{Span}(\mathbf{e_1}, \mathbf{e_2})$. Per *Theorem* 2.3.2, we set up the system and check for a solution.

$$\left[\begin{array}{cc|c}1 & 0 & a\\ 0 & 1 & b\\ \mathbf{e_1} & \mathbf{e_2} & \end{array}\right]$$

This system has a solution (take $x = a, y = b$ for the explicit solution, or use Theorem 2.2.12), and it has no free variables (see The Rank Theorem) so the solution is unique. This means, no matter which vector $\begin{bmatrix}x\\y\end{bmatrix}$ we pick in $\mathbb{R}^2$, we can always find a unique linear combination of $\mathbf{e_1}$ and $\mathbf{e_2}$ which results in $\begin{bmatrix}x\\y\end{bmatrix}$.

Example 2.3.5

Let $\mathbf{e_1} = \begin{bmatrix}1\\0\\0\end{bmatrix}$ and $\mathbf{e_2} = \begin{bmatrix}0\\1\\0\end{bmatrix}$ be standard basis vectors in $\mathbb{R}^3$. Give an intuitive description of $\text{Span}(\mathbf{e_1}, \mathbf{e_2})$. Then justify your description using Theorem 2.3.2.

By definition, an arbitrary vector in $\text{Span}(\mathbf{e_1}, \mathbf{e_2})$ is of the form

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

and so Span($\mathbf{e_1}, \mathbf{e_2}$) is just the xy-plane in $\mathbb{R}^3$. Notably, this set does not span all of $\mathbb{R}^3$, however, because it is missing vectors with a nonzero 3rd component (i.e. the z-direction).

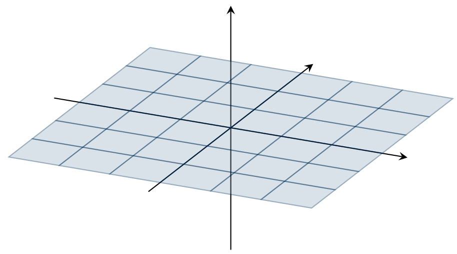


Figure 2.2: The span of $\mathbf{e_1}$ and $\mathbf{e_2}$ in $\mathbb{R}^3$.

To verify that this span is actually the xy-plane in $\mathbb{R}^3$, we need to show that every vector $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \in \text{Span}(\mathbf{e_1}, \mathbf{e_2})$. Per Theorem 2.3.2, we set up the system and check for a solution.

$$\left[\begin{array}{cc|c} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \\ \mathbf{e_1} & \mathbf{e_2} & \end{array}\right]$$

This system has a solution (Theorem 2.2.12), and it has no free variables (see The Rank Theorem) so the solution is unique. This means, no matter which vector $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$ we pick in $\mathbb{R}^2$, we can always find a unique linear combination of $\mathbf{e_1}$ and $\mathbf{e_2}$ which results in $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$.

Example 2.3.6

Let $\mathbf{v_1} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ and $\mathbf{v_3} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Give an intuitive description of Span($\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$). Then justify your description using Theorem 2.3.2.

By definition, an arbitrary vector in Span($\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$) is of the form

$$x\mathbf{v_1} + y\mathbf{v_2} + z\mathbf{v_3} = \begin{bmatrix} 2x - y + 3z \\ x + 2y + 4z \\ x + y + 3z \end{bmatrix}$$

It's a little hard to see what this is from the generic vector given above, so let's plot the vectors for a few (hundred) values of x, y, z.

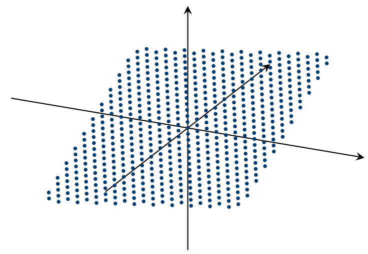


Figure 2.3: The span of $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$, which is a plane in $\mathbb{R}^3$.

Despite being the span of three vectors, the shape formed appears to only be a plane instead of a 3-dimensional objects. To verify that this span is actually a plane and not all of $\mathbb{R}^3$, we set up a linear system to see which $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$. Per Theorem 2.3.2, we set up the system and check for a solution.

$$\left[\begin{array}{ccc|c} 2 & -1 & 3 & x \\ 1 & 2 & 4 & y \\ 1 & 1 & 3 & z \\ \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \end{array}\right] \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 2 & -1 & 3 & x \\ 0 & 5 & 5 & 2y - x \\ 0 & 0 & 0 & x + 3y - 5z \end{array}\right]$$

This consistency of this system *depends on the values of* x, y, z! In fact, this system has a solution *if and only if* $x + 3y - 5z = 0$. We note that $x + 3y - 5z = 0$ is the equation of a plane in $\mathbb{R}^3$,

Example 2.3.7

Let $\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Give an intuitive description of $\text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$. Then justify your description using Theorem 2.3.2.

The span is all of $\mathbb{R}^3$. To see this, we employ the same tired technique and verify that every vector $[x, y, z]^T \in \mathbb{R}^3$ is contained in $\text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$:

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 1 & -1 & 0 & y \\ 0 & 0 & 1 & z \\ \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \end{array}\right] \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & -2 & 0 & y - x \\ 0 & 0 & 1 & z \end{array}\right]$$

By Theorem 2.2.12, this system always has a solution, hence every vector $[x, y, z]^T$ is contained in the span.

Linear (In)dependence

By now you probably have the question about what "dimension" the span might have, and whether there's any relation to the number of vectors used for the spanning set.

Definition: linear (in)dependence

A set of vectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$ in a vector space is **linearly dependent** if one or more of the vectors $\mathbf{v}_i$ can be written as a linear combination of the others. The set is called **linearly independent** otherwise.

Remark. "Linear independence" is a way of trying to capture the notion of "different directions" using only linear combinations.

Remark. The definition of linear dependence says that at least one of the vectors is a linear combination of the others (i.e. it "depends on the others"), but it does not necessarily imply that *every* vector in the set is a linear combination of the others, nor does it imply which one.

Example 2.3.8

Consider the set

$$\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

of vectors in $\mathbb{R}^2$. Show that this set is linearly dependent, but one of the vectors is not a linear combination of the others.

Certainly we have that

$$\begin{aligned} \mathbf{v}_1 &= \frac{1}{2}\mathbf{v}_2 + 0\mathbf{v}_3 \\ \mathbf{v}_2 &= 2\mathbf{v}_1 + 0\mathbf{v}_3 \end{aligned}$$

so the set is linearly independent, but it should be fairly clear to see that $\mathbf{v}_3$ is *not* a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

If we don't know which vector in a set is a linear combination of the others, how do we check for linear (in)dependence? Well, notice that if one of the vectors ($\mathbf{v_1}$, say) is a linear combination of the others

$$a_2\mathbf{v_2} + a_3\mathbf{v_3} + \cdots + a_k\mathbf{v_k} = \mathbf{v_1}$$

then we can rearrange this to write

$$-\mathbf{v_1} + a_2\mathbf{v_2} + a_3\mathbf{v_3} + \cdots + a_k\mathbf{v_k} = \mathbf{0}.$$

In other words, the homogeneous system

$$x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3} + \cdots + x_k\mathbf{v_k} = \mathbf{0}$$

has a nontrivial solution.

Theorem 2.3.9: Using systems to check for linear (in)dependence

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent if and only if the homogeneous system $\left[\begin{array}{ccc|c} \mathbf{v}_1 & \cdots & \mathbf{v}_k & \mathbf{0} \end{array}\right]$ has infinitely many solutions.
A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent if and only if the homogeneous system $\left[\begin{array}{ccc|c} \mathbf{v}_1 & \cdots & \mathbf{v}_k & \mathbf{0} \end{array}\right]$ has a unique solution (the trivial solution).

Example 2.3.10: Revisiting Example 2.3.4

Is the following set of vectors linearly independent?

$$\left\{ \mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

We employ Theorem 2.3.9 and set up the homogeneous system.

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mathbf{e_1} & \mathbf{e_2} & \mathbf{0} \end{array}\right]$$

By the The Rank Theorem, this system has a unique solution. Therefore the given set of vectors is linearly independent.

Example 2.3.11: Revisiting Example 2.3.5

Is the following set of vectors linearly independent?

$$\left\{ \mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

We employ Theorem 2.3.9 and set up the homogeneous system.

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \mathbf{e_1} & \mathbf{e_2} & \mathbf{0} \end{array}\right]$$

By the The Rank Theorem, this system has a unique solution. Therefore the given set of vectors is linearly independent.

Example 2.3.12: Revisiting Example 2.3.6

Is the following set of vectors linearly independent?

$$\left\{ \mathbf{v_1} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v_3} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix} \right\}$$

We employ Theorem 2.3.9 and set up the homogeneous system.

$$\left[\begin{array}{ccc|c} 2 & -1 & 3 & 0 \\ 1 & 2 & 4 & 0 \\ 1 & 1 & 3 & 0 \\ \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \mathbf{0} \end{array}\right] \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 2 & -1 & 3 & 0 \\ 0 & 5 & 5 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

By the The Rank Theorem, this system has infinitely-many solutions. Therefore the given set of vectors is linearly dependent.

Example 2.3.13: Revisiting Example 2.3.7

Is the following set of vectors linearly independent?

$$\left\{ \mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We employ Theorem 2.3.9 and set up the homogeneous system.

$$\underset{\mathbf{v_1}\quad\mathbf{v_2}\quad\mathbf{v_3}\quad\mathbf{0}}{\left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]} \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

By the The Rank Theorem, this system has a unique solution. Therefore the given set of vectors is linearly independent.

It seems like linear (in)dependence is playing a role in our geometric understanding of the span. Naively, one would guess that the span of linearly dependent set of k vectors is strictly smaller than dimension k.

Proposition 2.3.14: Throw Away Redundancies

If $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \ldots, \mathbf{v}_k$, then

$$\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \text{Span}(\mathbf{v}_2, \ldots, \mathbf{v}_k).$$

Proof. Suppose $\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$ and consider an arbitrary linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$:

$$\begin{aligned} &a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \\ &= a_1(c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \\ &= (a_1c_2 + a_2)\mathbf{v}_2 + (a_1c_3 + a_3)\mathbf{v}_3 + \cdots + (a_1c_k + a_k)\mathbf{v}_k \end{aligned}$$

Every combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ can be rewritten as a linear combination of just $\mathbf{v}_2, \ldots, \mathbf{v}_k$. ☐

Proposition 2.3.15

Given the vector equation

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

the "redundant" vectors are precisely the vectors corresponding to free variables in the system.

Proof. Suppose $\mathbf{v_j}$ is a linear combination of the preceding vectors $\mathbf{v_1}, \ldots, \mathbf{v_{j-1}}$. Then after row reduction, we have that column j (of the row reduced matrix) is the same linear combination of columns $1, \ldots, j-1$ (of the row reduced matrix), and therefore cannot contain a new pivot. ☐

Remark. The moral of the story is that linear dependence tells us about redundant information in our set. Moreover, Proposition 2.3.14 suggests that, if we can find these redundant vectors, we can iteratively throw them away without affecting the span, and ultimately leave ourselved with a linearly independent set.

Example 2.3.16: A maximal linearly independent set

Consider the following set of vectors:

$$\left\{\mathbf{v_1} = \begin{bmatrix}1\\1\end{bmatrix}, \mathbf{v_2} = \begin{bmatrix}-1\\1\end{bmatrix}, \mathbf{v_3} = \begin{bmatrix}3\\-1\end{bmatrix}, \mathbf{v_4} = \begin{bmatrix}4\\-2\end{bmatrix}\right\}.$$

Use Proposition 2.3.14 to reduce the following dependent set to a linearly dependent set whose span is the same as $\text{Span}(\mathbf{v_1}, \dots, \mathbf{v_4})$.

This is secretly just Example 2.2.14 again. We have

$$\left[\begin{array}{cccc|c} 1 & -1 & 3 & 4 & 0 \\ 1 & 1 & -1 & -2 & 0 \\ \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \mathbf{v_4} & \mathbf{0} \end{array}\right] \xrightarrow{R_2 - R_1 \mapsto R_2} \left[\begin{array}{cccc|c} 1 & -1 & 3 & 4 & 0 \\ 0 & 2 & -4 & -6 & 0 \end{array}\right]$$

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

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

We see that this set of vectors is linearly dependent, and specifically that $\mathbf{v_3}$, $\mathbf{v_4}$ are "redundant" per Proposition 2.3.15. Thus $\mathbf{v_1}, \mathbf{v_2}$ is a linearly independent set and

$$\text{Span}(\mathbf{v_1}, \mathbf{v_2}) = \text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}).$$

Definition: basis

Let V be a vector space. A **basis** for V is a set of vectors from V, call it $\mathcal{B}$, satisfying both of the following:

(1) $\mathcal{B}$ is a spanning set for V.
(2) $\mathcal{B}$ is linearly independent.

Remark. Given $\mathbf{v} \in V$ and a basis $\mathcal{B} = \{\mathbf{b_1}, \dots, \mathbf{b_n}\}$, the spanning condition guarantees that

$$x_1\mathbf{b_1} + \cdots + x_n\mathbf{b_n} = \mathbf{v}$$

always has a solution, and inear independence guarantees that the solution is unique.

Moreover, Proposition 2.3.14 assures us that $\mathcal{B}$ is the smallest set of vectors with this property.

Example 2.3.17

The vectors $\{\mathbf{e_1}, \mathbf{e_2}\}$ from Example 2.3.4 are a basis for $\mathbb{R}^2$ because we showed that they are linearly independent and that they span $\mathbb{R}^2$.

Example 2.3.18

The vectors $\{\mathbf{e_1}, \mathbf{e_2}\}$ from Example 2.3.5 are *not* a basis for $\mathbb{R}^3$ because they do not span $\mathbb{R}^3$.

Example 2.3.19

The vectors $\{\mathbf{e_1}, \mathbf{e_2}\}$ from Example 2.3.5 are a basis for $\text{Span}(\mathbf{e_1}, \mathbf{e_2})$ because we showed that they were linearly independent and they pretty clearly span $\text{Span}(\mathbf{e_1}, \mathbf{e_2})$.

Example 2.3.20

The vectors $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ from Example 2.3.6 are *not* a basis for $\mathbb{R}^3$ because we showed that they were neither linearly independent nor did they span $\mathbb{R}^3$.

Example 2.3.21

The vectors $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ from Example 2.3.6 are *not* a basis for $\text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$ because they are not linearly independent.

Example 2.3.22: Polynomial basis

Show that the following set of polynomials

$$\mathcal{B} = \{1,\, x,\, x^2,\, \ldots, x^n\}$$

is a basis for $\mathcal{P}_n(x)$, the space of all polynomials in the variable x with degree at most n.

Every polynomial in $\mathcal{P}_n(x)$ is of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

and is thus clearly a linear combination of the polynomials in $\mathcal{B}$. Hence $\text{Span}(\mathcal{B}) = \mathcal{P}_n(x)$.
To see linear independence, write down the 0 polynomial as a linear combination of the polynomials in $\mathcal{B}$:

$$t_0 1 + t_1x + t_2x^2 + \cdots + t_nx^n = 0 = 0 + 0x + 0x^2 + \cdots + 0x^n.$$

Clearly this only has a solution when $t_0 = t_1 = \cdots = t_n = 0$, hence is linearly independent.

Example 2.3.23: Polynomial basis

Show that the following set of polynomials

$$\mathcal{B} = \{1, 1 + x, 1 + x^2\}$$

is a basis for $\mathcal{P}_2(x)$, the space of all polynomials in the variable x with degree at most 2.

Every polynomial in $\mathcal{P}_n(x)$ is of the form

$$a_0 + a_1x + a_2x^2$$

and our goal is to show that this generic polynomial can be written as a linear combination of the polynomials in $\mathcal{B}$. In other words, we need to find scalars t_0, t_1, t_2 so that

$$a_0 + a_1x + a_2x^2 = t_0(1) + t_1(1 + x) + t_2(1 + x^2) = (t_0 + t_1 + t_2) + t_1x + t_2x^2.$$

This produces for us the following linear system:

$$\begin{cases} t_0 + t_1 + t_2 = a_0 \\ t_1 = a_1 \\ t_2 = a_2 \end{cases}$$

This sytem can be solved in our favorite way to get the unique solution

$$t_0 = a_0 - a_1 - a_2, t_1 = a_1, t_2 = a_2,$$

and it follows that every polynomial can be written as a linear combination. Hence $\mathcal{B}$ spans $\mathcal{P}_2(x)$.
To see linear independence, write down the 0 polynomial as a linear combination of the polynomials in $\mathcal{B}$

$$0 = t_0(1) + t_1(1+x) + t_2(1+x^2) = (t_0 + t_1 + t_2) + t_1 x + t_2 x^2$$

which produces for us the homogeneous system

$$\begin{cases} t_0 + t_1 + t_2 = 0 \\ t_1 = 0 \\ t_2 = 0 \end{cases}$$

and this has only the trivial solution. Thus $\mathcal{B}$ is linearly independent.
Therefore $\mathcal{B}$ is a basis for $\mathcal{P}_2(x)$.

Remark. There is nothing unique about a basis.

Example 2.3.24: Infinitely-many bases for $\mathbb{R}^2$

Let m and n be any nonzero numbers. Show that

$$\mathcal{B} = \left\{ \begin{bmatrix} m \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ n \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^2$.

INCOMPLETE

Theorem 2.3.25

If V is some vector space with two bases, $\mathcal{B}_1$ and $\mathcal{B}_2$, then both $\mathcal{B}_1$ and $\mathcal{B}_2$ contain the same number of vectors.

Proof Sketch. Set up a system to write $\mathcal{B}_1$'s vectors as linear combinations of $\mathcal{B}_2$. If $\mathcal{B}_2$ has more vectors, then you'll find that this system has free variables, i.e., the set $\mathcal{B}_2$ will be linearly dependent. □

Definition: dimension

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

Proposition 2.3.26

The dimension of $\mathbb{R}^n$ is

Proof. The standard basis $\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ is a basis for $\mathbb{R}^n$ and it contains n vectors. □

Chapter 3

Matrices

3.1 Matrix Operations

3.1.1 Matrix Basics

Definition: matrix, size, zero matrix

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.

Notably, rows are numbered top-to-bottom and columns are numbered left-to-right.

Definition: zero matrix

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

Definition: matrix equality

Two matrices A and B are **equal** if and only if

1. their sizes are equal, and
2. their corresponding entries are all equal.

In this case we write $A = B$.

Example 3.1.1

Which of the following matrices are equal?

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$$

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

$A = C$. A and B (and similarly A and D) are not equal because their sizes disagree. B and D are not equal because their corresponding entries disagree.

Definition: square matrix, diagonal matrix

A matrix is **square** if it has size $n \times n$ and the **diagonal of the matrix** $A = [a_{ij}]$ (circled below) are 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 **diagonal** if the only nonzero entries are along the diagonal. You may see this written as $A = \text{diag}(a_{11}, \ldots, a_{nn})$.

Remark. 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

Which of the matrices below are diagonal matrices?

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

A and B are both diagonal because the off-diagonal entries are all zero. C is not diagonal because it is not a square matrix.

Definition: scalar matrix

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: Kroenecker delta

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:

matrix!identity 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 in terms of the Kroenecker delta 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: matrix sum

Given two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$, the **sum** of A and B is a new $m \times n$ matrix given by

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

Example 3.1.4

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 15 & 14 & 13 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 5 & 6 \\ 12 & 11 & 110 \end{bmatrix}$. Compute $A + B$.

$$A + B = \begin{bmatrix} 1+4 & 2+5 & 3+6 \\ 15+12 & 14+11 & 13+10 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 11 \\ 27 & 25 & 23 \end{bmatrix}.$$

Definition: scalar multiplication

For an $m \times n$ matrix $A = [a_{ij}]$ and a scalar r, the **scalar multiple** of A is the $m \times n$ matrix

$$rA = [ra_{ij}] = \begin{bmatrix} ra_{11} & \cdots & ra_{1n} \vdots & \vdots \\ ra_{m1} & \cdots & ra_{mn} \end{bmatrix}.$$

Example 3.1.5

Compute rA 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

Subtraction of matrices is then defined in the obvious way: $A - B = A + (-1)B$.

3.1.3 Matrix Multiplication

Recall that we said a linear system

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

Could be written succinctly as a

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

where A is the matrix of coefficients, $\mathbf{x}$ is the vector of variables, and $\mathbf{b}$ is a vector of constants. With our new matrix perspective, however, the above equation looks like the product of an $m \times n$ matrix and an $n \times 1$ matrix, and the resulting output is a $m \times 1$ matrix.

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{m\times 1}$$

As such, we'll define the product of a matrix and a vector.

Definition: Matrix-vector multiplication

The **product** of an $m \times n$ matrix A and a vector $\mathbf{x} \in \mathbb{R}^n$ is a vector in $\mathbb{R}^m$ given by

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

Given that matrix is comprised of column vectors, we can extend this product to a product of matrices in the following way.

Definition: matrix product

The **product** of an $m \times n$ matrix A and an $n \times p$ matrix B is the $m \times p$ matrix given by

$$AB = \underbrace{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} | & & | \\ \mathbf{b_1} & \cdots & \mathbf{b_p} \\ | & & | \end{bmatrix}}_{n\times p} = \underbrace{\begin{bmatrix} | & & | \\ A\mathbf{b_1} & \cdots & A\mathbf{b_p} \\ | & & | \end{bmatrix}}_{m\times p}$$

where $\mathbf{b_i}$ is the i^{th} column of B. The above form is called the form is called the **matrix-column representation** of the product AB.

Observation: Matrix Multiplication and Dot Products

The product of matrices can also be written with dot products. The (i, j)-entry of the product matrix AB is given by the dot product:

$$\mathbf{row_i}(A) \cdot \mathbf{col_j}(B).$$

Example 3.1.6

Let A and B be the following matrices

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

Compute the product AB.

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

Example 3.1.7

Compute the following matrix products. Is there anything special about the resulting product?

(a) $A\mathbf{e_2} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \\ k & \ell & m \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

(b) $\mathbf{e_2}^T A = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \\ k & \ell & m \end{bmatrix}$

Observation

Let A be any $m \times n$ matrix and $\mathbf{e_i}$ the standard basis vector for $\mathbb{R}^m$ or $\mathbb{R}^n$ (whichever is appropriate for the following products to work). Then

- $A\mathbf{e_i}$ returns the i^{th} column of A, and
- $\mathbf{e_i}^T A$ returns the i^{th} row of A.

Example 3.1.8

Compute the following matrix products. Is there anything special about the resulting product?

(a) $A(Diag) = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \\ k & \ell & m \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

(b) $(Diag)A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \\ k & \ell & m \end{bmatrix}$

Observation

Let A be any $m \times n$ matrix and D a diagonal (either $m \times m$ or $n \times n$, whichever is appropriate for the following products to work). Let λ_i denote the i^{th} diagonal entry in D. Then

- AD scales the i^{th} column of A by λ_i, and
- DA scales the i^{th} row of A by λ_i.

3.1.4 Matrix Powers

Definition

If A is a square matrix, then for any positive integers k we can define the k^{th} **power** of the matrix A as repeated multiplication:

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

We take the convention that $A^0 = I$, the identity matrix.

Proposition 3.1.9

The obvious rules of matrix powers hold. Let k, ℓ be nonnegative integers. Then

- $A^k A^\ell = A^{k+\ell}$, and
- $\left(A^k\right)^\ell = A^{k\ell}$.

Example 3.1.10

Compute A^2 and A^3, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

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

and

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

Notably, $A^2 \neq \begin{bmatrix} 1^2 & 2^2 \\ 3^2 & 4^2 \end{bmatrix}$ and $A^3 \neq \begin{bmatrix} 1^3 & 2^3 \\ 3^3 & 4^3 \end{bmatrix}$.

3.1.5 Transpose

Definition: matrix transpose

If $A = [a_{ij}]$ is an $m \times n$ matrix, then its **transpose**, denoted A^T is the $n \times m$ matrix who $(i, j)^{\text{th}}$ entry is a_{ji}. 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

DRAW LINE USING nicematrix tikz codeafter

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

Remark. We don't really give much motivation for the matrix transpose at this level. It's related to something called the *dual space* for a vector space, but this is beyond the scope of this course. The important feature is encoded in the following exercise.

Exercise 3.1.11

Let A be an aribtrary $m \times n$ matrix, $\mathbf{v}$ an arbitrary vector in $\mathbb{R}^n$, and $\mathbf{w}$ an arbitrary vector in

$\mathbb{R}^m$. Show that

$$(A\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (A^T\mathbf{w}).$$

Definition: symmetric matrix

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

Example 3.1.12

Determine which of the following matrices is/are symmetric.

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

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

Since $A = A^T$, then A is symmetric. Since $B \neq B^T$, then B is not symmetric.

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 3.1.13

Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$. Compute A^TA and AA^T. Do you notice anything interesting?

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

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

What's interesting to notice is that, while A is not symmetric (and not even square), both AA^T and A^TA are symmetric (and hence also a square matrix). These particular matricesOB 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 be $m \times n$ matrices 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 $m \times n$ matrices form a real vector space (see page **??**). For this reason, it is sometimes denoted $\mathbb{R}^{m\times n}$.

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.

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) $I_mA = A = AI_n$ (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 Theorem 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,
- (this is nonstandard notation) $\text{row}_i(M)$ denotes the i^{th} row of M, and
- (this is nonstandard notation) $\text{col}_j(M)$ 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 **??**), we note that I_mA and A have the same size, hence they are equal if they have the same corresponding elements.

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

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 $C = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$. Compute $A(BC)$ and $(AB)C$ to verify that the products are equal.

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

and

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

Example 3.2.4

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ -1 & -2 \end{bmatrix}$. Compute $(AB)^T$, $B^T A^T$, and $A^T B^T$ to check which of these are equal.

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

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 $(AB)^T \neq A^T B^T$.

Theorem 3.2.5: *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) $(A^r)^T = (A^T)^r$ for all nonnegative integers r.

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

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

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).

□

3.3 The Inverse of a Matrix

Recall that division of real numbers $\frac{r}{a}$ is actually just multiplication of real numbers $a^{-1}r$ where a^{-1} is a real number satisfying $aa^{-1} = a^{-1}a = 1$.

Definition: matrix inverse

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 – the zero matrix is an obvious example. We'll devote the latter half of this section to exploring when a matrix is invertible.

Remark. We only define inverses for square matrices.

Theorem 3.3.1: *Poole* Theorem 3.6

The inverse is unique.

Proof. Suppose X, Y are both inverses of A. Then

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

□

Theorem 3.3.2: *Poole* Theorem 3.9 - Properties of Inverses

If A, B are invertible $n \times n$ matrices and $c \in \mathbb{R}$ is some nonzero scalar, then

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

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

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

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 is essentially proven by merely checking that the multiplication checks out.

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 Computing the Inverse

Given an invertible matrix $A = [A_{ij}]$, our goal is to solve for a matrix $X = [x_{ij}]$ is a matrix of indeterminates in the equation $AX = I_n$. This equation yields a linear system (with n^2 equations and n^2 unknowns)

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

and you can use standard techniques to solve this system.

Example 3.3.3

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

Let $X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$ be a matrix of indeterminates. Then the matrix equation

$$AX = I_2$$
$$\begin{bmatrix} x_1 + 2x_3 & x_2 + 2x_4 \\ 3x_1 + 4x_3 & 3x_2 + 4x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

yields the system

$$\begin{cases} x_1 + 2x_3 = 1 \\ x_2 + 2x_4 = 0 \\ 3x_1 + 4x_3 = 0 \\ 3x_2 + 4x_4 = 1 \end{cases}$$

And solving this in the usual way, we get that there is a unique solution $x_1 = -2$, $x_2 = 1$, $x_3 = \frac{3}{2}$, $x_4 = -\frac{1}{2}$. So

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}.$$

Writing $X = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_n} \\ | & | & | & | \end{bmatrix}$ and $I_n = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \\ | & | & | & | \end{bmatrix}$, we get

$$AX = \begin{bmatrix} | & | & & | \\ A\mathbf{x_1} & A\mathbf{x_2} & \cdots & A\mathbf{x_n} \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \\ | & | & | & | \end{bmatrix}$$

and so solving for X in the matrix equation $AX = I_n$ is akin to solving n linear systems $A\mathbf{x_j} = \mathbf{e_j}$. But wait, since we'd be doing the same row operations in every system, we can make life even easier and just make a really wide augmented matrix encoding all of these $\mathbf{e_j}$'s to the right of the vertical line.

$$\left[\begin{array}{cccc|cccc} | & | & | & | & | & | & | & | \\ \mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_n} & \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \\ | & | & | & | & | & | & | & | \end{array}\right].$$

Theorem 3.3.4: Gauss–Jordan for finding inverses

If A is an invertible $n \times n$ matrix, then A^{-1} is achieved by row reducing the system $\left[A \mid I\right]$. In particular

$$\left[A \mid I\right] \xrightarrow{RREF} \left[I \mid A^{-1}\right].$$

We apply this technique to the matrix we saw previously

Example 3.3.5: Revisiting Example 3.3.3

Compute the inverse for $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Employing the Gauss–Jordan technique,

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

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

Theorem 3.3.6: *Poole* Theorem 3.8

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad - bc \neq 0$, then A is invertible and the inverse is given by

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

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.

Exercise 3.3.7: Inverting a 3×3

Assuming $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \end{bmatrix}$ is invertible, verify that

$$A^{-1} = \frac{1}{aej + bfg + cdh - afh - bdj - ceg}\begin{bmatrix} (ej - fh) & (-bj + ch) & (bf - ce) \\ (-dj + fg) & (-aj + cg) & (-af + cd) \\ (dh - eg) & (-ah + bg) & (ae - bd) \end{bmatrix}.$$

Definition: elementary matrix

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

Fact. Elementary matrices are always invertible (and the inverse is an elementary matrix obtained by performing the inverse row operation.

Elementary matrices perform elementary row operations via matrix multiplication

Example 3.3.8

Let k be a scalar. Describe the row operations that each of the following matrices encodes:

$$\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\quad\begin{bmatrix}1&0&k\\0&1&0\\0&0&1\end{bmatrix}\quad\begin{bmatrix}1&0&0\\0&k&0\\0&0&1\end{bmatrix}$$

We try the computations:

$$\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{31}&a_{32}&a_{33}&a_{34}\\a_{21}&a_{22}&a_{23}&a_{24}\end{bmatrix}$$

so the first matrix encodes the row operation $R_2 \leftrightarrow R_3$.

$$\begin{bmatrix}1&0&k\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\end{bmatrix}=\begin{bmatrix}(a_{11}+ka_{31})&(a_{12}+ka_{32})&(a_{13}+ka_{33})&(a_{14}+ka_{34})\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\end{bmatrix}$$

so the second matrix encodes the row operation $R_1 + kR_3 \mapsto R_1$.

$$\begin{bmatrix}1&0&0\\0&k&1\\0&1&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\a_{21}&a_{22}&a_{23}&a_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\ka_{21}&ka_{22}&ka_{23}&ka_{24}\\a_{31}&a_{32}&a_{33}&a_{34}\end{bmatrix}$$

so the third matrix encodes the row operation $kR_2 \mapsto R_2$.

As such, if one can row reduce A to the identity, then the sequence of row operations must be encoded in a sequence of elementary matrices E_i. That is

$$E_nE_{n-1}\cdots E_2E_1A = I$$

It follows then that

$$A^{-1} = (E_nE_{n-1}\cdots E_2E_1).$$

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.

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

$$R_2-3R_1\mapsto R_2\begin{bmatrix}1&2\\0&-2\end{bmatrix}=\begin{bmatrix}1&0\\-3&1\end{bmatrix}A$$

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

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

3.3.2 Using inverse to solve systems

If A is an invertible $n \times n$ matrix, then

$$A\mathbf{x} = \mathbf{b} \Longrightarrow A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \Longrightarrow \mathbf{x} = A^{-1}\mathbf{b}$$

Hence the system $A\mathbf{x} = \mathbf{b}$ has the solution $\mathbf{x} = A^{-1}\mathbf{b}$. Moreover, since A^{-1} is unique, we expect this solution to be unique.

Theorem 3.3.10: *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
- $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Example 3.3.11

Use the inverse to solve the system $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix}1 & 4\\3 & 13\end{bmatrix}$ 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 3.3.6 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 3.3.12

Can we use the inverse to solve the following 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 3.3.6 we would be dividing by 0).

Remark. The astute reader may be wondering if it's worth it to find the inverse. After all, the same row operations will yield the solution anyway, so finding the inverse and then doing matrix multiplication seems like extra steps. While that's true, the crucial difference is now we can solve other systems involving the same coefficient matrix without row reducing multiple times. That is, by computing A^{-1} only once, it is effortless to solve three different $A\mathbf{x} = \mathbf{b_1}$, $A\mathbf{x} = \mathbf{b_2}$, and $A\mathbf{x} = \mathbf{b_2}$. Employing Gauss–Jordan on each system separately would take roughly three times as long finding the inverse.

3.3.3 When a matrix is invertible – The Fundamental Theorem

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

Theorem 3.3.13: *Poole Theorem 3.12* - The Fundamental Theorem of Invertible Matrices: Pt I

Let A be 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.

For part (d) 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 $\text{rank}(A) = \text{rank}(\left[A \mid \mathbf{b}\right]) = n$ – hence it's consistent – and the system only has n variables, so by the rank theorem there is a unique solution.

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: (real) vector space

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]

With this in mind, we introduce the following definition:

Definition: vector subspace

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).

Proposition 3.5.1: 3-Step Subspace Test

Let V be a vector space with zero vector $\mathbf{0}$. A subset U of V is a subspace of V if and only if the following three conditions are satisfied:

1. U contains $\mathbf{0}$.
2. If U contains two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, then u also contains the sum $\mathbf{u}_1 + \mathbf{u}_2$.
   We say that U is "closed under addition."
3. If U contains a vector $\mathbf{u}$, then for any scalar k, W also contains the scalar multiple $k\mathbf{u}$.
   We say that U is "closed under scalar multiplication."

Remark. U must satisfy *all* of the above to be a subspace. If U fails to satisfy *one or more* of the above criteria, then it is not a subspace.

Example 3.5.2

Every vector space V is a subspace of itself.

Example 3.5.3: Trivial Vector Space

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

This is straightforward to check using the 3-Step Subspace Test.

1. U contains $\mathbf{0}$ by definition.
2. Since $\mathbf{0} + \mathbf{0} = \mathbf{0}$, then U contains $\mathbf{0} + \mathbf{0}$.
3. For any scalar k, we have that $k\mathbf{0} = \mathbf{0}$, so U contains $k\mathbf{0}$.

Example 3.5.4: xy-Plane in $\mathbb{R}^3$

Let U be the set of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. Then U is a subspace of $\mathbb{R}^3$.

1. Clearly U contains $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.
2. Suppose U contains both $\mathbf{u}_1 = \begin{bmatrix} x_1 \\ y_1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} x_2 \\ y_2 \\ 0 \end{bmatrix}$. Since $\mathbf{u}_1 + \mathbf{u}_2 = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 0 \end{bmatrix}$, then U contains $\mathbf{u}_1 + \mathbf{u}_2$.
3. Suppose U contains $\mathbf{u} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$ and k is any scalar. Since $k\mathbf{u} = \begin{bmatrix} kx \\ ky \\ 0 \end{bmatrix}$ then U contains $k\mathbf{u}$.

By the 3-Step Subspace Test, U is a subspace of $\mathbb{R}^3$.

Example 3.5.5

Let U be the set of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} x \\ y \\ x + y \end{bmatrix}$. Is U a subspace of $\mathbb{R}^3$?

1. Clearly U contains $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

2. Suppose U contains both $\mathbf{u}_1 = \begin{bmatrix} x_1 \\ y_1 \\ x_1 + y_1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} x_2 \\ y_2 \\ x_2 + y_2 \end{bmatrix}$. Since $\mathbf{u}_1 + \mathbf{u}_2 = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ (x_1 + x_2) + (y_1 + y_2) \end{bmatrix}$, then U contains $\mathbf{u}_1 + \mathbf{u}_2$.
3. Suppose U contains $\mathbf{u} = \begin{bmatrix} x \\ y \\ x + y \end{bmatrix}$ and k is any scalar. Since $k\mathbf{u} = \begin{bmatrix} kx \\ ky \\ kx + ky \end{bmatrix}$ then U contains $k\mathbf{u}$.

By the 3-Step Subspace Test, U is a subspace of $\mathbb{R}^3$.

Example 3.5.6: $z = 1$ Plane in $\mathbb{R}^3$

Let U be the set of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$. Then U is <u>not</u> a subspace of $\mathbb{R}^3$.

So long as U fails one or more criteria of the 3-Step Subspace Test, it will fail to be a subspace of $\mathbb{R}^3$.

1. U does not contains $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.
2. Suppose U contains both $\mathbf{u}_1 = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$. Since $\mathbf{u}_1 + \mathbf{u}_2 = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ 2 \end{bmatrix}$, then U does not contains $\mathbf{u}_1 + \mathbf{u}_2$.
3. Suppose U contains $\mathbf{u} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ and k is any scalar. Since $k\mathbf{u} = \begin{bmatrix} kx \\ ky \\ k \end{bmatrix}$, then whenever $k \neq 1$, U does not contain $k\mathbf{u}$.

As a matter of fact, U fails *every* criterion of the 3-Step Subspace Test.

Example 3.5.7: unit disk in $\mathbb{R}^2$

Let W be the unit disk in $\mathbb{R}^2$. That is, $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \; : \; x^2 + y^2 \leq 1 \right\}$. Is W a subspace of $\mathbb{R}^2$?

W contains $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, but for any $k > 1$, the vector $k \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} k \\ 0 \end{bmatrix}$ is not in W. Therefore U is <u>not</u> a subspace.

Example 3.5.8: Polynomial Subspaces

Let $\mathcal{P}^2$ be the set of polynomials of degree at most 2 and $\mathcal{P}^3$ the set of all polynomials of degree at most 3. Every polynomial $p(x) = a_2x^2 + a_1x + a_0$ in $\mathcal{P}^2$ is also a polynomial in $\mathcal{P}^3$:

$$p(x) = 0x^3 + a_2x^2 + a_1x + a_0.$$

Is $\mathcal{P}^2$ is a subspace of $\mathcal{P}^3$?

With the usual operations of polynomial addition and scalar multiplication (see Example 1.1.3), one can use the 3-Step Subspace Test to check that $\mathcal{P}^2$ is a subspace of $\mathcal{P}^3$. The idea comes down to the fact that the sum and scalar multiple of degree-2 polynomials always results in a degree-2 polynomial (hence it is closed under both addition and salar multiplication). In fact, this same argument shows that $\mathcal{P}^m$ is a subspace of $\mathcal{P}^n$ whenever $m < n$.

Example 3.5.9

Let U be the set of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} x \\ y \\ \sin(x) \end{bmatrix}$. Is U a subspace of $\mathbb{R}^3$?

Consider $\mathbf{u} = \begin{bmatrix} \pi/2 \\ 0 \\ \sin(\pi/2) \end{bmatrix} = \begin{bmatrix} \pi/2 \\ 0 \\ 1 \end{bmatrix}$. Then $2\mathbf{u} = \begin{bmatrix} \pi \\ 0 \\ 2 \end{bmatrix}$, and this is not in U because $\sin(\pi) \neq 2$. This subset fails to be closed under addition, hence by the 3-Step Subspace Test, U is not a subspace of $\mathbb{R}^3$.

Example 3.5.10

Let W be the finite set $W = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix} \right\}$. W is <u>not</u> a subspace of $\mathbb{R}^2$.

Let $\mathbf{w} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. Notice that W does not contain $k\mathbf{w}$ unless $k = 0$ or $k = 1$, hence by the 3-Step Subspace Test, W is not a subspace of $\mathbb{R}^2$.

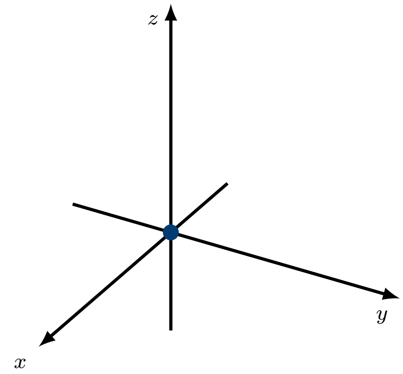
3.5.1 Geometry of Subspaces of $\mathbb{R}^n$

At this point, subspaces have only been described purely algebraically, so what do they look like geometrically?

Example 3.5.11: Subspace in Example 3.5.3

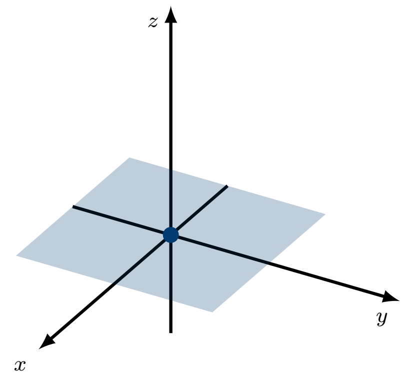
The set $\{\mathbf{0}\}$, i.e. the trivial subspace (c.f. Example 3.5.3). This *is* a subspace of $\mathbb{R}^3$ (in fact, of $\mathbb{R}^n$ for an n).

This is just a point – the origin.



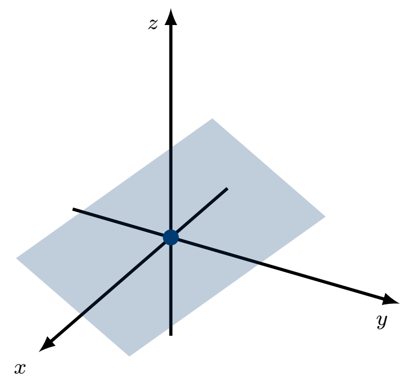
Example 3.5.12: Subspace in Example 3.5.4

Example 3.5.4 is the xy-plane. This *is* a subspace of $\mathbb{R}^3$.



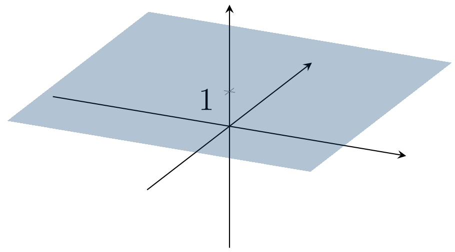
Example 3.5.13: Subspace in Example 3.5.5

Example 3.5.5 is the some skewed plane through the origin. This *is* a subspace of $\mathbb{R}^3$.



Example 3.5.14: Non-Subspace in Example 3.5.6

Example 3.5.6 is parallel to the xy-plane but does not pass through the origin. This *is not* a subspace of $\mathbb{R}^3$.



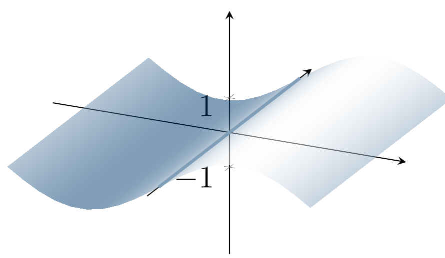
Example 3.5.15: Non-Subspace in Example 3.5.7

Example 3.5.7 passes through the origin, but it doesn't extend infinitely-far in any direction. This *is not* a subspace of $\mathbb{R}^2$.

Picture Here

Example 3.5.16: Non-Subspace in Example 3.5.9

Example 3.5.9 passes through the origin, but is this wonky curvy surface. This *is not* a subspace of $\mathbb{R}^3$.



So what we're gathering is that nontrivial subspaces have to look like infinite, "flat" objects through the origin. In fact, the three conditions of the 3-Step Subspace Test essentially tell us this:

1. The subspace must pass through the origin.
2. The subspace must be "flat".
3. The subspace must contain infinitely-many vectors and extend infinitely-far in every possible direction (trivial subspace excepted).

For now we'll simply state this as a fact, and the proof will follow from our later discussion of bases:

Theorem 3.5.17

Every subspace of $\mathbb{R}^n$ is a copy of $\mathbb{R}^m$ where $0 \leq m \leq n$.

3.5.2 The Four Fundamental Subspaces Associated with Matrices

With the notion of a "subspace" in mind, let's try to revisit some ideas involving matrices. First, a simple result.

Theorem 3.5.18: *Poole* Theorem 3.19

Let $\mathbf{v}_1, \ldots, \mathbf{v_k} \in \mathbb{R}^n$. Then $\text{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is a subspace of $\mathbb{R}^n$.

Proof. For simplicitly, let $W = \text{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$.

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 $c \in \mathbb{R}$ be some scalar. Then
$$c\mathbf{u} = (cc_1)\mathbf{v_1} + \cdots + (cc_k)\mathbf{v_k}.$$
Since $c\mathbf{u}$ is a linear combination of the vectors $\mathbf{v_i}$, then $c\mathbf{u} \in W$.

□

Definition: column space, row space

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

1. The **column space of A** is a subspace of $\mathbb{R}^m$ 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 and compute with column vectors whenever possible, it may be more useful to define $\text{Row}(A) := \text{Col}(A^T)$.

Theorem 3.5.19: *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. We appeal to the 3-Step Subspace Test.

1. Clearly $\mathbf{x} = \mathbf{0}$ is a solution to the homogeneous system.
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: null space, left null space

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

1. The **null space of A** is the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$. It is a subspace of $\mathbb{R}^n$ and is denoted $\text{Null}(A)$.
   (In some texts, it is called the *kernel of* A and is denoted $\ker(A)$.)
2. The **left null space** of A is the set of solutions to the homogeneous system $A^T\mathbf{x} = \mathbf{0}$. It is a subspace of $\mathbb{R}^m$ and is denoted $\text{Null}(A^T)$.
   (In some texts, it is called the *cokernel of* A and is denoted $\text{coker}(A)$.)

Remark. We won't be overly concerned with the left null space of A in this class. I'll mention it again in Chapter 5 to try to justify why it's considered a "fundamental subspace."

Example 3.5.20

Compute $\text{Col}(A)$, $\text{Row}(A)$, and $\text{Null}(A)$ 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 (hey wait, that's how many vectors span both $\text{Row}(A)$ and $\text{Col}(A)$... weird), 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: basis for a subspace

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

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

Remark. This definition is the same as a basis for a vector space, as previously defined. We've simply restated in terms of subspaces to highlight the fact that the basis vectors *have to come from the subspace.*

Example 3.5.21

Given $A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \end{bmatrix}$, the matrix in Example 3.5.20, find a basis for $\text{Col}(A)$, for $\text{Row}(A)$, and for $\text{Null}(A)$.

INCOMPLETE

3.5.3 Strategies For Finding Bases

Finding a basis for the column space

1. Row reduce the matrix (just row-echelon form is fine)
2. Identify the "pivot columns" (those which contain a leading entry/pivot)
3. Your basis is the pivot columns from the original matrix.

Remark. It's important that we take as a basis the vectors from the original matrix. Our row operations fundamentally alter the columns (but preserve relationships between columns), so almost certainly the column space of A is different from the column space of $\text{RREF}(A)$. For example

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad \text{RREF}(A) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

The column space of A is the y-axis in $\mathbb{R}^2$, whereas the column space of $\text{RREF}(A)$ is the x-axis in $\mathbb{R}^2$.

Example 3.5.22

Find basis for the column space of $A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{bmatrix}$.

By definition $\text{Col}(A) = \text{Span}(\text{columns of } A)$, so the "span" condition of the basis is checked off. All we have to do now is find a linearly independent set of vectors in this span. Using the "Proposition 2.3.14" result, we can do this by checking for linear independence in A's columns, and simply removing all vectors which do not correspond to a leading entry in $\text{RREF}(A)$.

$$\text{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 & \mathbf{c}_5 \end{bmatrix}$$

Using the reduced row echelon form, we see that columns $\mathbf{c}_3$ and $\mathbf{c}_5$ are linear combinations of the other columns. Notably,

$$\mathbf{c_3} = 1\mathbf{c_1} + 2\mathbf{c_2} \quad \text{and} \quad \mathbf{c_5} = -1\mathbf{c_1} + 3\mathbf{c_2} + 4\mathbf{c_4}$$

Letting $\mathbf{v_1}, \ldots, \mathbf{v_5}$ be the columns of A, we have the same linear dependencies

$$\mathbf{v_3} = 1\mathbf{v_1} + 2\mathbf{v_2} \quad \text{and} \quad \mathbf{v_5} = -1\mathbf{v_1} + 3\mathbf{v_2} + 4\mathbf{v_4}$$

(and you can check that this is true). Hence $\text{Col}(A) = \text{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_4})$ and $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_4}\}$ is a linearly independent set. Hence

$$\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_4}\} = \left\{ \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

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

Finding a basis for the null space

1. Row reduce the matrix (just row-echelon form is fine)
2. Solve the system $A\mathbf{x} = \mathbf{0}$.
3. Write the solution set as a parameterized linear combination of k vectors (where k is the number of free variables).
4. Take these k vectors as a basis for $\text{Null}(A)$.

Finding a basis for the left null space

1. Transpose the matrix.
2. Find a basis for the null space of the transpose.

Remark. It is actually always true that the vectors one gets from the parametric form are linearly independent, and this is easy to see from the previous example. We specifically chose our vectors to correspond to the free variables x_3 and x_5. The vector corresponding to x_3 has a 1 in the third component and a 0 in the fifth component. Similarly, the vector corresponding to x_5 has a 1 in the fifth component and a 0 in the third component. There is no way that a scalar multiple of one vector can be transformed into the other (because we can't turn 0's into 1's).

More generally, there will be exactly one vector on each list with a number other than 0 in a prescribed component, so it cannot be a linear combination of the others.

Example 3.5.23

Find basis for the null space of $A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{bmatrix}$, the same matrix from Example 3.5.22.

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 $\text{Null}(A)$.

Example 3.5.24

Find a basis for the subspace $W = \text{Span}\left(\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}-2\\2\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}2\\0\\1\end{bmatrix}\right)$.

Put these vectors into the columns of a matrix A. Now, by definition, $W = \text{Col}(A)$. Employ the usual technique.

Example 3.5.25

Find a basis for the subspace $W = \left\{\begin{bmatrix}x\\y\\z\end{bmatrix} \in \mathbb{R}^3 \text{ such that } x+y+z=0\right\}$.

Notice that the defining equation is a linear system, which we can write as

$$A\mathbf{x} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}0\end{bmatrix} = \mathbf{0}.$$

In this way, we see that $W = \text{Null}(A)$. Now find a basis for this null space.

3.5.4 Dimension and Rank

Theorem 3.5.26: *Poole* Theorem 3.23 - The Basis Theorem

Let W be a subspace of a vector space V 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.

In other words, the number of basis vectors is unique, and we have the obvious definition from before, applied to subspaces:

Definition: dimension of a subspace

The **dimension** of a subspace W is the number of vectors in a basis for W. We denote this $\dim(W)$.

What's the dimension of the column space and the dimension of the row space? Observe the following correspondence:

$$\begin{aligned}\dim(\text{Col}(A)) &= \#\text{ linearly independent columns in } A\\ &= \#\text{ leading entries in } \text{RREF}(A)\\ &= \#\text{ nonzero rows in } \text{RREF}(A)\\ &= \text{rank}(A).\end{aligned}$$

Definition: rank, again

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

Now let's think about what row operations are actually doing to the rows: they are just adding/subtracting linear combinations of rows to get rid of dependencies. For example

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Looking at the transpose of this last matrix, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}$$

and clearly the first and second columns are linearly independent! In other words,

Theorem 3.5.27: *Poole* Theorem 3.24

For any matrix A, $\dim(\text{Row}(A)) = \dim(\text{Col}(A))$.

or equivalently

Theorem 3.5.28: *Poole* Theorem 3.25

For any matrix A, $\text{rank}(A) = \text{rank}(A^T)$

The row space and column space dimensions get a cool name, but what about the null space? Well...

Definition: nullity

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

Theorem 3.5.29: *Poole* Theorem 3.26 - Rank–Nullity

If A is an $m \times n$ matrix, then $\text{rank}(A) + \text{nullity}(A) = n$.

It would be good to compare this to the original The Rank Theorem. Every "free variable" in our linear system produces a basis vector for the null space. So if the original rank theorem said "leading variables + free variables = total variables," then one can see how this relates to the rank, nullity, and number of columns.

This also allows us to add to the Fundamental Theorem of Invertible Matrices.

Theorem 3.5.30: Fundamental Theorem of Invertible Matrices, Pt II

Suppose A is an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

$\vdots$

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

3.5.5 Coordinates

Recall that span and linear independence culminate in the following:

Theorem 3.5.31

Let V be a vector space with an ordered basis $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$. For every vector $\mathbf{v} \in V$, there is a unique linear combination of $\mathcal{B}$-basis vectors such that

$$\mathbf{v} = k_1\mathbf{b_1} + \cdots + a_n\mathbf{b_n}.$$

Definition: coordinates, coordinate vector (with respect to a basis)

The coefficients k_i in the previous theorem are called the **coordinates of v with respect to the $\mathcal{B}$** and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix}$$

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

The whole idea is that a basis is just providing a reference for how one can travel in a space, sothe coordinates are communicating the directions.

Example 3.5.32

Let $P(0,0)$ and $Q(3,1)$ be points in the plane and consider the vector $\mathbf{v} = \overrightarrow{PQ}$. Write the coordinate vector of $\mathbf{v}$ with respect to the standard basis for $\mathbb{R}^2$ (which you might denote as $\mathcal{E} = \{\mathbf{e_1}, \mathbf{e_2}\}$).

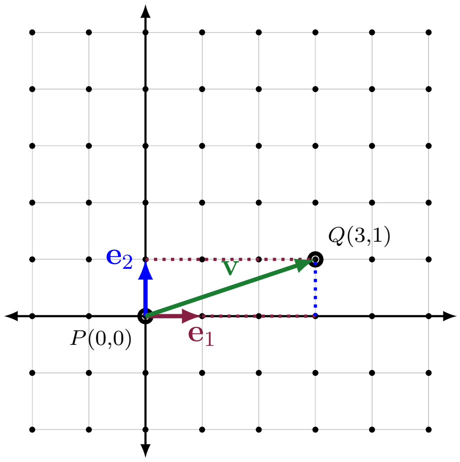


Figure 3.1: $\mathbf{v}$ in the $\mathcal{E}$-basis.

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

hence

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

Example 3.5.33

Let $P(0,0)$ and $Q(3,1)$ be points in the plane and consider the vector $\mathbf{v} = \overrightarrow{PQ}$. Write the coordinate vector of $\mathbf{v}$ with respect to 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 $\mathbb{R}^2$.

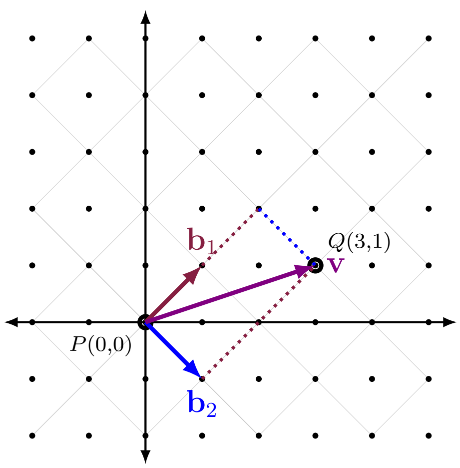


Figure 3.2: $\mathbf{v}$ in the $\mathcal{B}$-basis.

$$\mathbf{v} = 2\mathbf{b_1} + 1\mathbf{b_1}$$

then

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

Remark. We typically don't write the subscript $\mathcal{E}$ for vectors when they are written in the standard basis. In fact, unless otherwise specified, all vectors should be thought of as being written in the standard basis.

Example 3.5.34

Let $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$. Find $\left[\mathbf{v}\right]_{\mathcal{B}}$ where $\mathcal{B} = \left\{ \underbrace{\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}}_{\mathbf{b_1}}, \underbrace{\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}}_{\mathbf{b_2}}, \underbrace{\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}}_{\mathbf{b_3}} \right\}$.

We just need to solve the system $k_1\mathbf{b_1} + k_2\mathbf{b_2} + k_3\mathbf{b_3} = \mathbf{v}$. Indeed,

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ -1 & 0 & -1 & 2 \\ 2 & 3 & 6 & 1 \end{array}\right] \xrightarrow{RREF} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & 15 \\ 0 & 0 & 1 & -10 \end{array}\right]$$

and thus $\left[\mathbf{v}\right]_{\mathcal{B}} = \begin{bmatrix} 8 \\ 15 \\ -10 \end{bmatrix}$.

3.6 Introduction to Linear Transformations

Definition: linear transformation

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

Remark. The "codomain" is where the range lives; these are generally not the same. For example, $f(x) = x^2$ is a function from $\mathbb{R}$ to $\mathbb{R}$, but the range of the function is only nonnegative real numbers.

Example 3.6.1: Identity transformation

Show that $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{v}) = \mathbf{v}$ is a linear transformation.

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ and k a scalar. Then

$$T(\mathbf{u} + \mathbf{v}) = \mathbf{u} + \mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$
$$T(k\mathbf{u}) = k\mathbf{u} = kT(\mathbf{u})$$

Example 3.6.2: Trivial transformation

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

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ and k a scalar. Then

$$T(\mathbf{u} + \mathbf{v}) = \mathbf{0} = \mathbf{0} + \mathbf{0} = T(\mathbf{u}) + T(\mathbf{v})$$

and

$$T(k\mathbf{u}) = \mathbf{0} = k\mathbf{0} = T(\mathbf{u})$$

Example 3.6.3: dot product/linear functional

Let $\mathbf{w}$ be some fixed vector in $\mathbb{R}^n$. Show that the function

$$\varphi : \mathbb{R}^n \to \mathbb{R}$$
$$\varphi(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w}$$

is a linear transformation.

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ and let k be any scalar. It follows from the Properties of the Dot

Product that

$$\varphi(\mathbf{u}+\mathbf{v}) = (\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = (\mathbf{u}\cdot\mathbf{w}) + (\mathbf{v}\cdot\mathbf{w}) = \varphi(\mathbf{u}) + \varepsilon_j(\mathbf{v})$$

and

$$\varphi(k\mathbf{v}) = (k\mathbf{v})\cdot\mathbf{w} = k(\mathbf{v}\cdot\mathbf{w}) = k\,\varphi(\mathbf{v})$$

Exercise 3.6.4: derivative

Let $\mathcal{P}^n(x)$ be the space of degree-n polynomials. Show that the derivative function

$$\frac{d}{dx} : \mathcal{P}^n(x) \to \mathcal{P}^{n-1}(x)$$

is a linear transformation.

Recall that polynomials are differentiable. Let $p(x)$ and $q(x)$ be degree-n polynomials and let k be any scalar. Then, from Calc I

$$\frac{d}{dx}\Big(p(x)+q(x)\Big) = \frac{d}{dx}\Big(p(x)\Big) + \frac{d}{dx}\Big(q(x)\Big)$$

and

$$\frac{d}{dx}\Big(k\,p(x)\Big) = k\,\frac{d}{dx}\Big(p(x)\Big)$$

Example 3.6.5: matrix transformation

Let A be any 3×2 matrix. Show that the following map is a linear transfmroation:

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

Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^2$, and let k be a scalar. Then the result follows from previously established facts about matrix operations (Theorems 3.2.1 and 3.2.2):

$$T(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$

and

$$T(k\mathbf{v}) = A(k\mathbf{v}) = k(A\mathbf{v}) = kT(\mathbf{v}).$$

The work done in the previous example proves a more general result.

Theorem 3.6.6

If A is an $m \times n$ matrix, then the function

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

is a linear transformation.

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for $\mathbb{R}^n$ and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. For any vector $\mathbf{v}$ in $\mathbb{R}^n$, where

$$\mathbf{v} = a_1\mathbf{b_1} + \cdots + a_n\mathbf{b_n}$$

we have

$$T(\mathbf{v}) = a_1T(\mathbf{b_1}) + \cdots + a_nT(\mathbf{b_n})$$

and so in this way

$$\text{range}(T) = \text{Span}(T(\mathbf{b_1}), \ldots, T(\mathbf{b_n}))$$

that is, everything about the function is determined by what happens to the basis $\mathcal{B}$. Let's look at the particular case of the standard basis.

Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ and write $\mathcal{E} = \{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ for the standard basis in $\mathbb{R}^n$. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and write $T(\mathbf{e_i}) = \mathbf{a_i}$, where $\mathbf{a_i}$ is a vector in $\mathbb{R}^m$. Then we have

$$\begin{aligned} T(\mathbf{v}) = T\left(\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\right) &= T(v_1\mathbf{e_1} + \cdots + v_n\mathbf{e_n}) \\ &= v_1T(\mathbf{e_1}) + \cdots + v_nT(\mathbf{e_n}) \\ &= v_1\mathbf{a_1} + \cdots + v_n\mathbf{a_n} \\ &= v_1\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + v_n\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = A\mathbf{v} \end{aligned}$$

Woah. That means that *every single linear transformation* is just multiplication by a matrix.

Theorem 3.6.7: Linear Transformations *Are* Matrices

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.

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

Remark. This is why we like column vectors. If we took row vectors as the standard, we'd have to write $T(\mathbf{x}) = \mathbf{x}A$, and that just looks silly.

Definition: standard matrix

The matrix in the above theorem is called the **standard matrix for** T. We may sometimes write $[T]$ to denote the standard matrix of T.

Remark. This matrix is "standard" because it uses the standard basis. As we've seen, one can obtain coordinate vectors for different bases, and you'll still be able to get a matrix representing the linear transformation. As you can imagine, notation for this can get kind of clunky.

Theorem 3.6.8: matrix multiplication is function composition

If $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^p \to \mathbb{R}^n$ are two linear transformations, then

$$[T_1 \circ T_2] = [T_1][T_2]$$

That is, matrix multiplication is just a composition of functions.

Proof. Let $\mathbf{x} \in \mathbb{R}^p$. Then

$$(T_1 \circ T_2)(\mathbf{x}) = T_1(T_2(\mathbf{x})) = T_1\big([T_1]\mathbf{x}\big) = [T_1][T_2]\mathbf{x}$$

and thus we get the following:

1. $(T_1 \circ T_2) : \mathbb{R}^p \to \mathbb{R}^m$ is a linear map and
2. the standard matrix is given by $[T_1 \circ T_2] = [T_1][T_2]$, an $m \times p$ matrix.

□

Example 3.6.9

Let S and T be the linear transformations below:

$$S : \mathbb{R}^2 \to \mathbb{R}^2 \qquad\qquad T : \mathbb{R}^2 \to \mathbb{R}^2$$

$$S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ x-y \end{bmatrix} \qquad\qquad T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 3x-y \\ 2y \end{bmatrix}$$

Find the standard matrix for S, the standard matrix for T, and the standard matrix for $S \circ T$. Use this to verify Theorem 3.6.8.

The standard matrix for S is

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

The standard matrix for T is

$$[T] = \left[T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) \;\; T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \right] = \begin{bmatrix}3 & -1\\0 & 2\end{bmatrix}$$

The product of these matrices is

$$[T][S] = \begin{bmatrix}2 & 4\\2 & -2\end{bmatrix}$$

Now, the composition is

$$T \circ S\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = T\left(\begin{bmatrix}x+y\\x-y\end{bmatrix}\right) = \begin{bmatrix}3(x+y)-(x-y)\\2(x-y)\end{bmatrix} = \begin{bmatrix}2x+4y\\2x-2y\end{bmatrix}$$

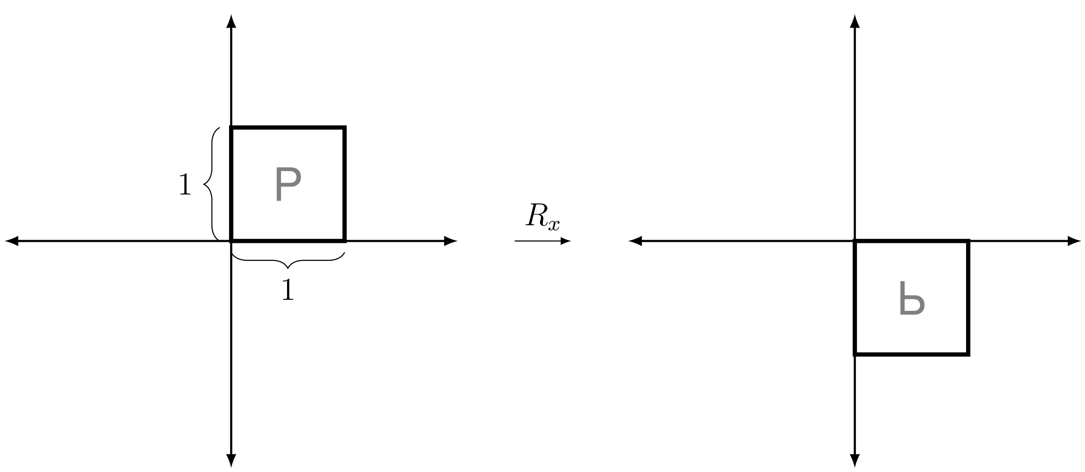
and thus the standard matrix for $T \circ S$ is

$$[T \circ S] = \left[(T \circ S)\left(\begin{bmatrix}1\\0\end{bmatrix}\right) \;\; (T \circ S)\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \right] = \begin{bmatrix}2 & 4\\2 & -2\end{bmatrix}.$$

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

Example 3.6.10: reflection about x-axis

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

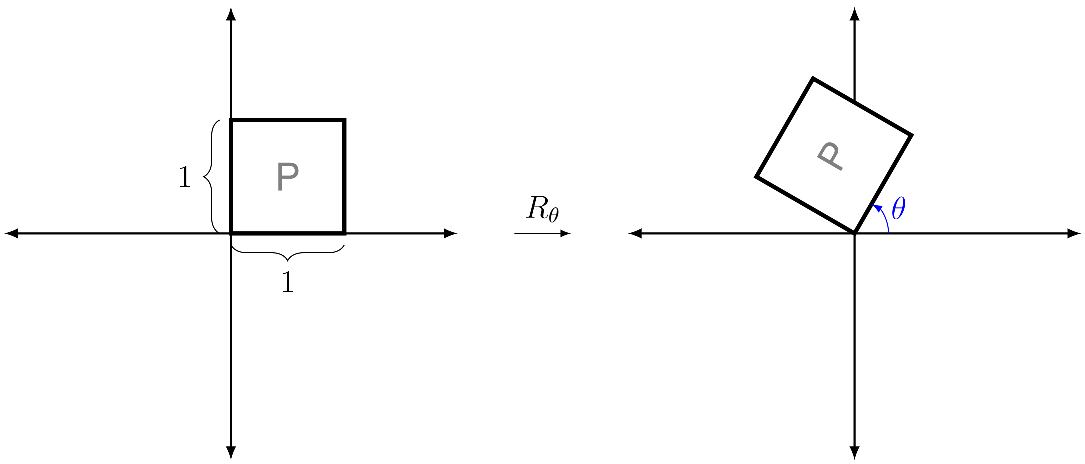
$$[R_x] = \left[R_x\left(\begin{bmatrix}1\\0\end{bmatrix}\right) \;\; R_x\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \right] = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$$

Exercise 3.6.11

Show that reflection about the y-axis is also a linear transformation and find the standard matrix for this.

Example 3.6.12: rotation

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.

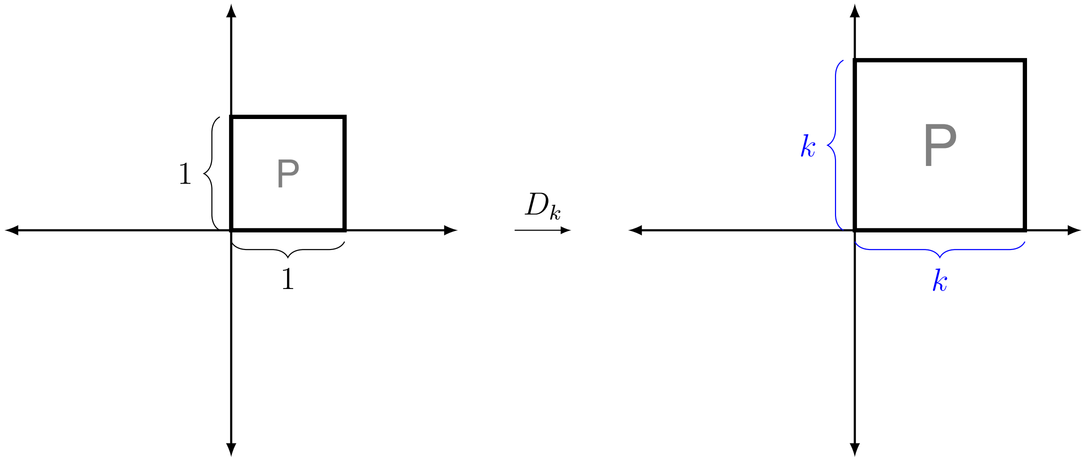
$$\begin{aligned}[R_\theta] &= \left[R_\theta\left(\begin{bmatrix}\cos(0)\\ \sin(0)\end{bmatrix}\right) \quad R_\theta\left(\begin{bmatrix}\cos(\frac{\pi}{2})\\ \sin(\frac{\pi}{2})\end{bmatrix}\right)\right] \\ &= \begin{bmatrix}\cos(0+\theta) & \cos(\frac{\pi}{2}+\theta)\\ \sin(0+\theta) & \sin(\frac{\pi}{2}+\theta)\end{bmatrix}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

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 3.6.13

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, ky) so for vectors,

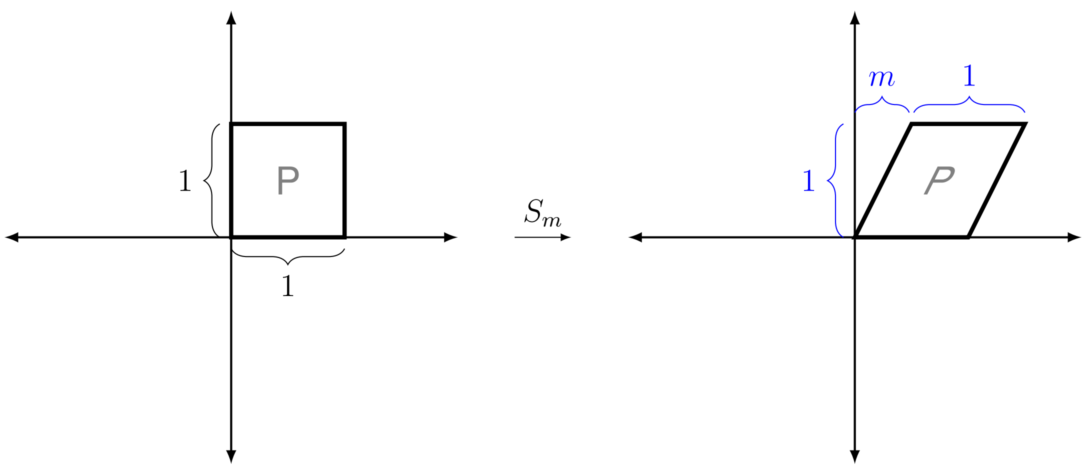
$$D_k\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} kx \\ ky \end{bmatrix}$$

The standard matrix for D_k is given by

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

Example 3.6.14: horizontal shear

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 $(x + my, y)$,

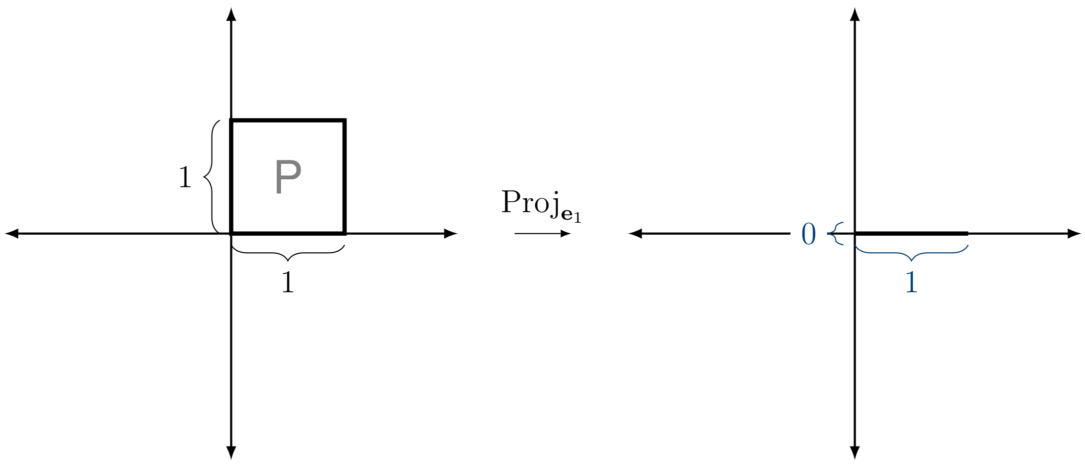
$$S_m\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + my \\ y \end{bmatrix}$$

The standard matrix for S_m is given by

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

Example 3.6.15: projection onto the x-axis

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



Since

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

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

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

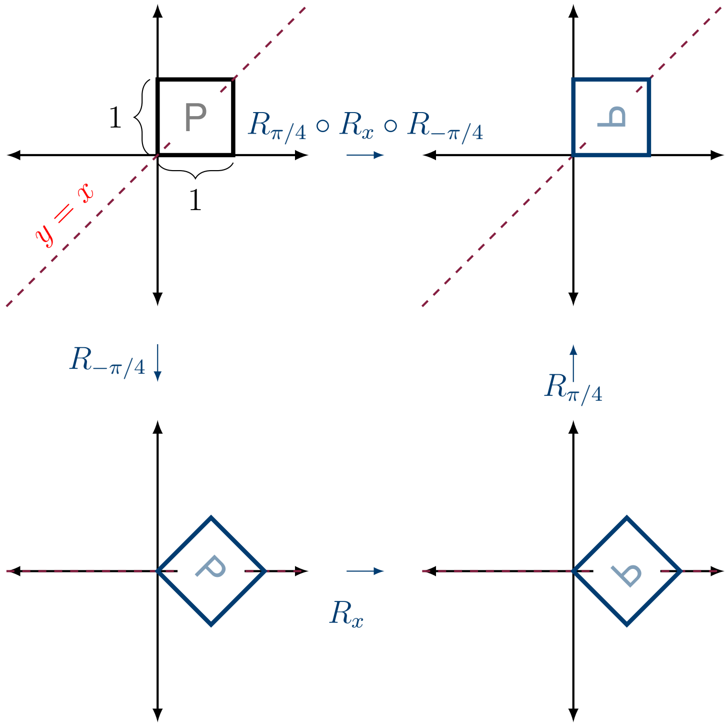
Example 3.6.16: composition of simpler transformations

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 $y = x$. This latter interpretation is somehow easier because we know all of the component matrices in the product.



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

$$[R_{\pi/4} \circ R_x \circ R_{-\pi/4}] = [R_{\pi/4}][R_x][R_{-\pi/4}] = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Exercise 3.6.17

Give a geometric description of the following matrices' behavior on $\mathbb{R}^3$ (i.e. on the vectors $[x, y, z]^T$).

1. $\begin{bmatrix} 1 & 3 & \\ 0 & 1 & \\ & & 2 \end{bmatrix}$
2. $\begin{bmatrix} \cos\theta & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$
3. $\begin{bmatrix} -1 & & \\ & 2 & \\ & & 3 \end{bmatrix}$

One-to-One, Onto, Invertibility

Not every matrix is square, so visualizing their behavior in a smooth way isn't really feasible. We aim to say something, then, about them in general.

Definition: range, kernel

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **range** of T is the set of all possible outputs

$$\text{range}(T) = \{T(\mathbf{x}) \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n\}$$

and the **kernel** of T is the collection of vectors in the domain which are sent to the zero vector:

$$\ker(T) = \{\mathbf{x} \text{ in } \mathbb{R}^n \text{ so that } T(\mathbf{x}) = \mathbf{0}\}$$

Remark. The range can loosely be interpreted as "vector information retained by T" and the kernel represents "vector information lost by T."

We make the following connection with the standard matrix, we have

Proposition 3.6.18

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ have standard matrix A. Then

- $\text{range}(T) = \text{Col}(A)$.
- $\ker(T) = \text{Null}(A)$.

Definition: one-to-one, onto

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 $\mathbf{v} \in \mathbb{R}^n$ for which $T(\mathbf{v}) = \mathbf{w}$.

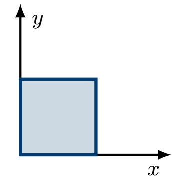
Remark. Maybe the right way to think about it is this: "one-to-one" means that the range of T is a copy of $\mathbb{R}^n$ living inside of $\mathbb{R}^m$, and "onto" means that the range of T is all of $\mathbb{R}^m$. In either case, this suggests that the range of T is as large as can be

For linear transformations, we have a more convenient way of thinking about these notions.

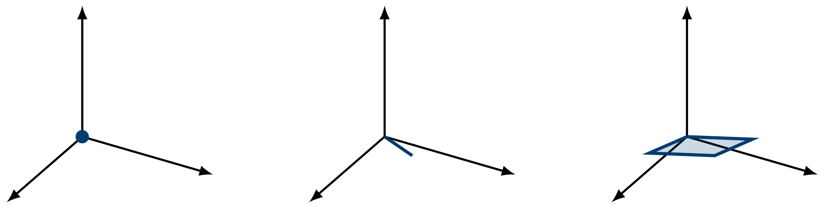
Example 3.6.19

Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ with standard matrix A. Sketch a diagram showing all possible ranges of T. When is T one-to-one? When is T onto?

Since $\text{Col}(A) = \text{range}(T)$, then $\text{range}(T)$ is a subspace of $\mathbb{R}^3$ whose dimension is equal to $\text{rank}(A)$. Since A is a 3×2 matrix, there are only three options for the rank: 0, 1, and 2. In turn, there are only three possible types of subspaces representing the range:



domain(T) = $\mathbb{R}^2$



range(T) $\cong \mathbb{R}^0$ rank(A) $= 0$ | range(T) $\cong \mathbb{R}^1$ rank(A) $= 1$ | range(T) $\cong \mathbb{R}^2$ 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 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

$$\begin{aligned} c_1T(\mathbf{e}_1) + c_2T(\mathbf{e}_2) &= \mathbf{0} \\ c_2T(\mathbf{e}_2) &= -c_1T(\mathbf{e}_1) \\ T(c_2\mathbf{e}_2) &= T(-c_1\mathbf{e}_1) \end{aligned}$$

and hence we have two different vectors with the same output - T is *not* one-to-one in these cases.

T clearly cannot be onto since the range is, as most, a 2-dimensional subspace of $\mathbb{R}^3$.

This idea motivates the following

Theorem 3.6.20: one-to-one and rank

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with standard matrix A is one-to-one if and only if rank(A) $= n$.

And following from Rank–Nullity,

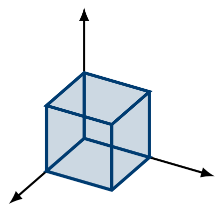
Corollary 3.6.21

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with standard matrix A is one-to-one if and only if nullity(A) $= 0$.

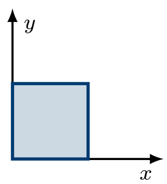
Example 3.6.22

Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with standard matrix A. Sketch a diagram showing all possible ranges of T. When is T one-to-one? When is T onto?

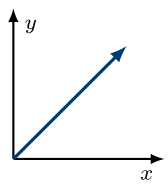
Since $\text{Col}(A) = \text{range}(T)$, then $\text{range}(T)$ is a subspace of $\mathbb{R}^2$ whose dimension is equal to $\text{rank}(A)$. Since A is a 2×3 matrix, there are only three options for the rank: 0, 1, and 2. In turn, there are only three possible types of subspaces representing the range:



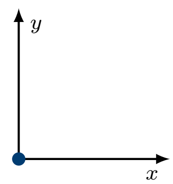
$\text{domain}(T) \cong \mathbb{R}^3$



$\text{range}(T) = \mathbb{R}^2$
$\text{rank}(A) = 2$



$\text{range}(T) \cong \mathbb{R}^1$
$\text{rank}(A) = 1$



$\text{domain}(T) \cong \mathbb{R}^0$
$\text{rank}(A) = 0$

Let $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$ be the standard basis for $\mathbb{R}^3$, the domain of T. Note that $\{T(\mathbf{e_1}), T(\mathbf{e_2}), T(\mathbf{e_3})\}$ 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 in these cases.

T clearly cannot be onto since the range is, as most, a 2-dimensional subspace of $\mathbb{R}^3$.

Theorem 3.6.23: onto and rank

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto if and only if $\text{rank}(A) = m$.

Example 3.6.24

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto the first coordinate:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ 0 \end{bmatrix} \text{ with standard matrix } A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Is T one-to-one? Is T onto?

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.

Corollary 3.6.25

A is invertible if and only if $T : \mathbb{R}^n \to \mathbb{R}^n$ 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 3.6.20 and 3.6.23, this is true if and only if $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is both one-to-one and onto. □

Theorem 3.6.26: *Poole Theorem 3.12* - The Fundamental Theorem of Invertible Matrices: Pt III

Let $A \in \mathbb{R}^{n \times n}$. The following are equivalent:

(a A is invertible.

⋮

(n) $T(\mathbf{x}) = A\mathbf{x}$ is an invertible linear transformation.

Definition: invertible linear transformation

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **invertible** (as a function) if its standard matrix A is invertible. In this case, $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ has standard matrix A^{-1}.

This is completely reasonable:

$$(T \circ T^{-1})(\mathbf{x}) = AA^{-1}\mathbf{x} = \mathbf{x} = A^{-1}A\mathbf{x} = (T^{-1} \circ T)\mathbf{x}$$

Example 3.6.27: rotations are invertible

Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation of the plane by an angle of θ. Find the inverse transformation.

Intuitively, the inverse transformation is $R_{-\theta}$, the rotation of the plane by an angle of $-\theta$. We can also see this by Looking at the standard matrices

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$[R_\theta]^{-1} = \frac{1}{\cos^2\theta + \sin^2\theta} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = [R_{-\theta}]$$

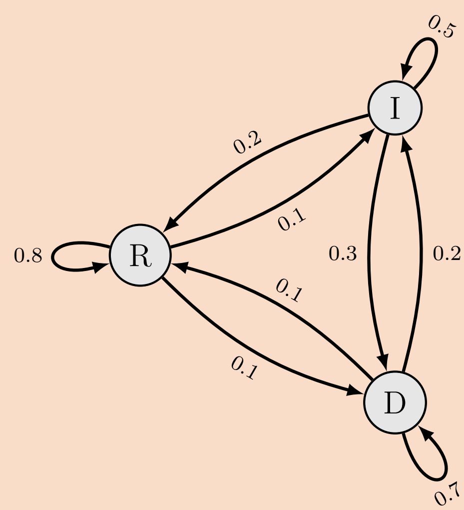
3.7 Applications

3.7.1 Markov Chains

Remark. I have chosen to handle this section *after* discussing eigenvalues/eigenvectors/diagonalization. As such, the presentation in these notes is slightly different than the course textbook.

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, $\mathbf{x_k} = P^k \mathbf{x_0}$.

Example 3.7.2

Given the process described in 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 $k \to \infty$?

$$\mathbf{x_1} = \begin{bmatrix} 870 \\ 800 \\ 430 \end{bmatrix}, \qquad \mathbf{x_{20}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}, \qquad \mathbf{x_{100}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}, \quad \mathbf{x_{1000}} \approx \begin{bmatrix} 764 \\ 859 \\ 477 \end{bmatrix}$$

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

$$\mathbf{x} = \lim_{k\to\infty} P^k \mathbf{x_0} = \lim_{k\to\infty} P^{k+1} \mathbf{x_0} = P \left(\lim_{k\to\infty} P^k \mathbf{x_0} \right) = P\mathbf{x}.$$

(where one of the equalities above follows from a fact we haven't proved – linear transformations are continuous).

Definition: Markov chains

The vector $\mathbf{x_0}$ in the last example is known as the **initial state vector**, A is known as the **transition matrix**, and the entire process above is called a **Markov chain** (with 3 **states**).

Since Markov chains specifically model probablistic scenarios, the types of transition matrix and state vectors are fundamentally related to probability. As such, they get special names.we have special names for these types of vectors and matrices.

Definition: stochastic matrix, probability vector

A vector with all nonnegative entries that sum to 1 is called a **probability vector**. A square matrix with probability *column vectors* is called a **stochastic matrix**. A stochastic matrix A is called a **regular stochastic matrix** if there is some power k for which A^k has all strictly-positive entries.

Remark. Every transition matrix in this class will be a regular stochastic matrix; don't worry about the formality of this. One motivation is that ensures that your transition matrix is not a block diagonal matrix, which means that you're modeling two completely unrelated/independent probablistic scenarios.

From our example, we observe the following

Proposition 3.7.3: entries of stochastic matrices

If A is a stochastic matrix, and k is some (positive) integer, then then (i, j)-entry of A^k represents the probability that object switches *to state i from state j after k iterations.* This can easily be found with the formula

$$\mathbf{e_i}^T A^k \mathbf{e_j}.$$

The motivation for this formula is that an object in state j means that it has a 100% chance of being in state j and a 0% chance of being in any other state, so this is the vector $\mathbf{e_j}$. To find the i^{th} component of a vector, one can just multiply by $\mathbf{e_i}^T$.

Theorem 3.7.4: The Fundamental Theorem of Markov Chains

For a regular stochastic matrix P, there is a unique steady state probability vector.

Corollary 3.7.5

A regular stochastic matrix P always has an eigenvalue of 1, which has geometric multiplicity 1.

Taking a limit of A^k as $k \to \infty$ is really hard because finding a Computing high powers of matrices is hard. One *could* use diagonalization, but it turns out that there's an easier way of obtaining steady state vectors. Observe 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

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 3.7.6

Find the 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 $\lim_{k\to\infty} P^k\mathbf{x}_0 = \lambda\mathbf{x}$.

Chapter 4

Eigenvalues and Eigenvectors

4.1 Introduction to Eigenvalues and Eigenvectors

4.1.1 Motivation - Geometric and Computational

In Section 3.6, we saw that we could think about the geometry of the plane after applying a 2×2 matrix (see Examples 3.6.10 and 3.6.12 to 3.6.15). In 3 dimensions, this becomes harder to "see", and in $n > 3$ dimensions, this is probably virtually impossible for most to "see." ☹ We'd like to get an understanding of what multiplication by an $n \times n$ matrix is doing to $\mathbb{R}^n$, even if we can't perfectly visualize it.

Observe the following

$$\begin{bmatrix} -3 & 2 & 2 \\ -14 & 8 & 5 \\ 2 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} -3 & 2 & 2 \\ -14 & 8 & 5 \\ 2 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

In other words, this 3×3 matrix, while hard to describe generally, could be given a somewhat intuitive description:

- it scales/dilates by a factor of 3 in the direction of $[0, -1, 1]^T$, and
- it scales/dilates by a factor of 1 in the direction of $[1, 2, 0]^T$.

Example 4.1.1

Let A be the 3×3 matrix $\begin{bmatrix} 5 & 0 & 0 \\ 21 & 5 & 3 \\ -5 & -2 & 0 \end{bmatrix}$. In which directions does A scale/dilate by 2? By 3? By 5?

We compute a bunch of null spaces.

$$\text{Null}(A-2I) = \text{Null}\begin{bmatrix} 3 & 0 & 0 \\ 21 & 3 & 3 \\ -5 & -2 & -2 \end{bmatrix} = \text{Null}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \text{Span}\left(\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right)$$
$$\text{Null}(A-3I) = \text{Null}\begin{bmatrix} 2 & 0 & 0 \\ 21 & 2 & 3 \\ -5 & -2 & -3 \end{bmatrix} = \text{Null}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix} = \text{Span}\left(\begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix}\right)$$
$$\text{Null}(A-5I) = \text{Null}\begin{bmatrix} 0 & 0 & 0 \\ 21 & 0 & 3 \\ -5 & -2 & -5 \end{bmatrix} = \text{Null}\begin{bmatrix} 7 & 0 & 1 \\ 0 & 7 & 15 \\ 0 & 0 & 0 \end{bmatrix} = \text{Span}\left(\begin{bmatrix} -1 \\ -15 \\ 7 \end{bmatrix}\right)$$

So we can say that A stretches in the direction $[0, 1, -1]^T$ by a factor of 2, it stretches in the direct $[0, 3, 2]^T$ by a factor of 3, and it stretches in the direction of $[-1, -15, 7]^T$ by a factor of 5.

The three vectors we found above actually end up being a basis for $\mathbb{R}^3$! (This is left as an exercise to check). And this is an awful convenient basis because it makes it easy to figure out images of linear transformations without having to multiply by the matrix.

Example 4.1.2

Let $\mathcal{B} = \left\{ \underbrace{\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}}_{\mathbf{b_1}}, \underbrace{\begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix}}_{\mathbf{b_2}}, \underbrace{\begin{bmatrix} -1 \\ -15 \\ 7 \end{bmatrix}}_{\mathbf{b_3}} \right\}$ be the the basis from the previous problem, and let T be the linear transformation defined as follows

$$T(\mathbf{b_1}) = 2\mathbf{b_1}, \qquad T(\mathbf{b_2}) = 3\mathbf{b_2}, \qquad T(\mathbf{b_3}) = 5\mathbf{b_3}.$$

Find

$$T\left(\begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}\right).$$

We can guess T is just multiplication by the matrix A from the previous example, and we get

$$T(\mathbf{x}) = \begin{bmatrix} 5 & 0 & 0 \\ 21 & 5 & 3 \\ -5 & -2 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 51 \\ -17 \end{bmatrix}.$$

However, since we're not told about this matrix A and finding it might be a chore, we can try another way.

It's an exercise to see that

$$\begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} = 3\mathbf{b_1} + 2\mathbf{b_2} - \mathbf{b_3}$$

so by properties of linear transformations

$$\begin{aligned} T\left(\begin{bmatrix}1\\6\\0\end{bmatrix}\right) &= T(3\mathbf{b_1} + 2\mathbf{b_2} - \mathbf{b_3}) \\ &= 3T(\mathbf{b_1}) + 2T(\mathbf{b_2}) - T(\mathbf{b_3}) \\ &= 3(2\mathbf{b_1}) + 2(3\mathbf{b_2}) - (5\mathbf{b_3}) \\ &= \begin{bmatrix}5\\51\\-17\end{bmatrix} \end{aligned}$$

So it must be that A is precisely the standard matrix for T, but we didn't have to compute it at all.

Example 4.1.3

With all of the same vectors/matrices as in the previous example, compute

$$(\underbrace{T \circ \cdots \circ T}_{1357})\left(\begin{bmatrix}1\\6\\0\end{bmatrix}\right).$$

Knowing that

$$T(\mathbf{x}) = A\begin{bmatrix}1\\6\\0\end{bmatrix}, \qquad \text{where } A = \begin{bmatrix}5 & 0 & 0\\21 & 5 & 3\\-5 & -2 & 0\end{bmatrix}$$

we must have that

$$(\underbrace{T \circ \cdots \circ T}_{1357})\left(\begin{bmatrix}1\\6\\0\end{bmatrix}\right) = A^{1357}\begin{bmatrix}1\\6\\0\end{bmatrix}.$$

Clearly computing such a large power of a matrix is unreasonable by hand, so instead we turn to thinking about this basis again.

$$\begin{aligned} A^{1357}\begin{bmatrix}1\\6\\0\end{bmatrix} &= A^{1357}\left(3\mathbf{b_1} + 2\mathbf{b_2} - \mathbf{b_3}\right) \\ &= A^{1356}\left(3A\mathbf{b_1} + 2A\mathbf{b_2} - A\mathbf{b_3}\right) \\ &= A^{1356}\left(3(2)\mathbf{b_1} + 2(3)\mathbf{b_2} - (5)\mathbf{b_3}\right) \\ &= A^{1355}\left(3(2)A\mathbf{b_1} + 2(3)A\mathbf{b_2} - (5)A\mathbf{b_3}\right) \\ &= A^{1355}\left(3(2)^2\mathbf{b_1} + 2(3)^2\mathbf{b_2} - (5)^2\mathbf{b_3}\right) \\ &\vdots \\ &= A\left(3(2)^{1355}A\mathbf{b_1} + 2(3)^{1355}A\mathbf{b_2} - 5^{1355}A\mathbf{b_3}\right) \\ &= A\left(3(2)^{1356}\mathbf{b_1} + 2(3)^{1356}\mathbf{b_2} - 5^{1356}\mathbf{b_3}\right) \\ &= 3(2)^{1356}A\mathbf{b_1} + 2(3)^{1356}A\mathbf{b_2} - 5^{1356}A\mathbf{b_3} \end{aligned}$$

$$= 3(2)^{1357}\mathbf{b_1} + 2(3)^{1357}\mathbf{b_2} - 5^{1357}\mathbf{b_3}$$

$$= \begin{bmatrix} 5^{1357} \\ 3 \cdot 2^{1357} - 6 \cdot 3^{1357} + 15 \cdot 5^{1357} \\ -3 \cdot 2^{1357} + 4 \cdot 3^{1357} - 7 \cdot 5^{1357} \end{bmatrix}$$

$$= \begin{bmatrix} \begin{array}{l} 317903920618191493426102720109864453822775435663406621422926466024775263324969463817186027196394381 3\\ 9979729989719483756313092265515479828862585088764805469841432413644922018007106716531119051690353568\\ 6219847229775824541601813631214504774699908796564679333609508326052356404602836588590278832973891486\\ 5067948676596100579342276597434628615958506999412808941001719448105647066943653202479324194864702033\\ 3771486857217721800685406316605021418695670648409019759142759857096526297844676843284681451889552748\\ 6311347143011215791981935909576642700119940346320329906238603735526254483556233818301781640172441734\\ 8484729361068016540597998832367227142643842421425027414378682811532441256749630953163132178753359819\\ 9665108413033133171029830358003193956833939698036455221540115866399293929460403624894786911416386396\\ 3333663095959220380059822475876747882198099580266289891423048436583265729962822253490467937475747721\\ 65149104702353000106995750684291124343872070312 5 \end{array} \\ \\ \begin{array}{l} 4768558809272872401391540801647966807341631534951099321343896990371628949874541957257790407945915720\\ 9969594984579225634469638398273219743293877633147208204762148620467383027010660074796678577535530352\\ 9329770844663736812402720446821757162049863194847019000414262489078534606904254882885418249460837229\\ 7431434979997918505956964022622671770857166711667206875557523500491867478971771256887319975035925942\\ 2040197089312187630398383187906603277026131763994996234630191792502223111375849448110727107351686160\\ 1805833638936869415093739699310837438461426483887598118563378236293145165827406646103046940922952574\\ 4416704942275632728302886931145806978277553705932668953970188384909417777300970904284736798856935004\\ 3241387387097013451671204236527342741706543372300920906036746498473738497047557484817494869768119696\\ 3941088428076741493566679832092178490834694640584602689603355442254840136574272273943392182639620551\\ 377233255224072883319975310049517817946469545 67113 \end{array} \\ \\ \begin{array}{l} -222532744432734045398271904076905117675942804964384634996048526217342684327478624672030219037476066\\ 9798581099280363862941916458586083588020380956213536382888900268955144541260497470157178333618324749\\ 8035389306084307717912126954185015334228993615759527553352665582823664948322198561201319518308172404\\ 0543390538368644883016862636918472592911722570795429523473783316589621753256475154384774939156454001\\ 8122868533570947521339330023028956242574205331480729156201063139928858628531522599344194429467777085\\ 7661454867833424313630303505558659234463738423491368616888392437147464237567166539975641982254136100\\ 5249208188800714061637504095402070550217852983254598917190733677456766048997797807743345790394334457\\ 9091101172742200040321841506589515110298920417288838283969245312149041002303861601599202057384416267\\ 8971919330595870079050353513176069603035131696095586125826339633416024902210812036957113752191475031\\ 5896978778798794886479081576740071870231385770 47239 \end{array} \end{bmatrix}$$

Let's give names to these "stretching" ideas.

Definition: eigenvalues, eigenvectors

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 "own" as in the sense of characterizing an intrinsic property; a less literal translation would be along the lines of"special" or "characteristic."

Remark. The eigenvectors are not unique. If $\mathbf{u}$ is an eigenvector corresponding to λ and $\mathbf{v} = k\mathbf{u}$ for some scalar k, then

$$A\mathbf{v} = A(k\mathbf{u}) = kA\mathbf{u} = k\lambda u = \lambda(k\mathbf{u}) = \lambda\mathbf{v}.$$

For any $n \times n$ matrix A, if there is a vector $\mathbf{v}$ for which $A\mathbf{v} = \lambda\mathbf{v}$ (where λ is some scalar), then we can say a little bit about this vector.

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ A\mathbf{v} - \lambda\mathbf{v} &= \mathbf{0} \\ A\mathbf{v} - \lambda I\mathbf{v} &= \mathbf{0} \\ (A - \lambda I)\mathbf{v} &= \mathbf{0} \end{aligned}$$

so $\mathbf{v}$ is in $\text{Null}(A - \lambda I)$

Definition: eigenspace

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, denoted $\boldsymbol{E_\lambda}$, is

$$E_\lambda := \operatorname{Span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$$

or equivalently

$$E_\lambda := \operatorname{Null}(A - \lambda I).$$

Remark. The eigenspace is comprised entirely of eigenvectors and the zero vector.

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

$A = \begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}$ has eigenvector $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$. Find the corresponding eigenvalue.

$$\begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ -4 \end{bmatrix} = \lambda\begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

hence the corresponding eigenvalue is $\lambda = -4$.

Example 4.1.5

Show that 3 is another eigenvalue of $A = \begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}$ and find an eigenvector corresponding to $\lambda = 3$.

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 $\operatorname{Null}(A - 3I)$.

$$\left[A - 3I \mid \mathbf{0}\right] = \left[\begin{array}{cc|c} -6 & 2 & 0 \\ 3 & -1 & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 3 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

So a vector $\mathbf{v}$ is in $\operatorname{Null}(A - 3I)$ if it has the form $\begin{bmatrix} t \\ 3t \end{bmatrix}$. As such, any nonzero vector of this form is an eigenvector of A corresponding to 3.

4.1.2 Geometry of Eigenvectors

Eigenvectors are those vectors whose directions aren't changed, but are merely scaled (and we're considering $\mathbf{v}$ and $-\mathbf{v}$ to have the same "direction"). It follows that eigenspaces are subspaces that are preserved or *stabilized* by the linear transformation.

Example 4.1.6

Using $A = \begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}$ as before, notice that the vectors $[-2, 1]^T$ and $[1, 3]^T$ do not change direction after a transformation, but are merely scaled.

Note: we consider $\mathbf{v}$ and $-\mathbf{v}$ to be in the "same direction".

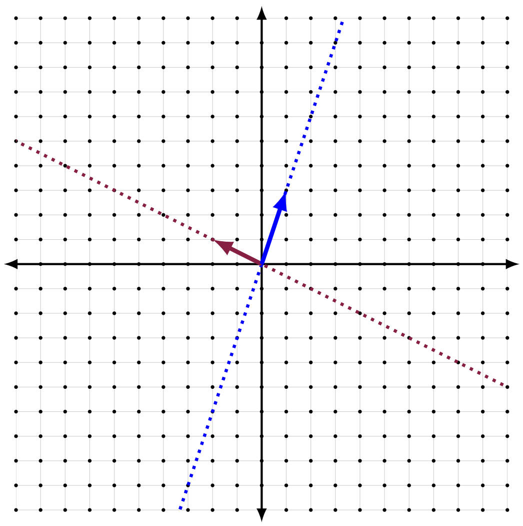


Figure 4.1: Before applying transformation A.

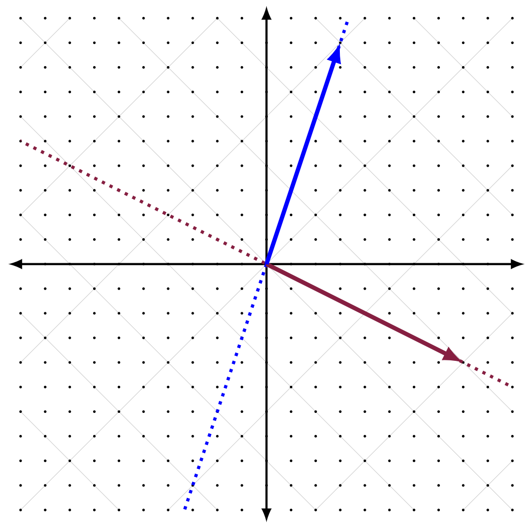


Figure 4.2: After applying transformation A.

Since eigenvectors correspond to directions that are unchanged, one can approximate them by plottinga number of vectors in all directions, apply the transformation, and look for those vectors whose directions have changed the least. Often one plots them "head-to-tail" so that the vector $\mathbf{v}$ emanates from the origin, and the vector $A\mathbf{v}$ is plotted emanating from the head of $\mathbf{v}$. Then the lengths of $\mathbf{v}$ and $A\mathbf{v}$ can be compared to estimate the eigenvalue. The next example shows this.

Example 4.1.7

Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

The image below shows number of unit vectors $\mathbf{v}$ (in black) and then the vectors $A\mathbf{v}$ (in blue) emanating from the heads of $\mathbf{v}$. Use this information to estimate the eigenvectors and eigenvalues for A, then verify algebraically.

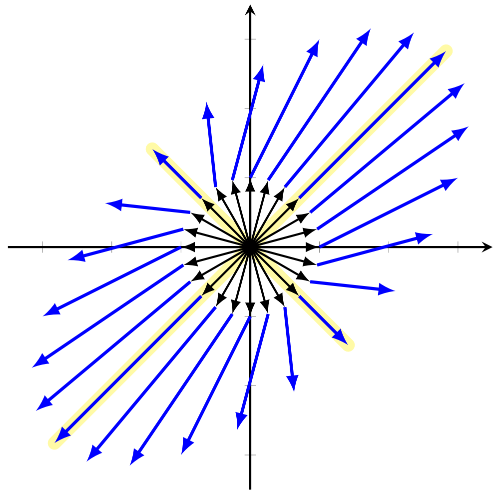


Figure 4.3: Vectors $\mathbf{v}$ and $A\mathbf{v}$ drawn head-to-tail. The directions that appear to be unchanged have been highlighted.

We see that the eigenvectors appear to be in the directions of $\mathbf{u_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{u_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Moreover, looking at the lengths of $A\mathbf{u_1}$ and $A\mathbf{u_2}$ relative to the lengths of $\mathbf{u_1}$ and $\mathbf{u_2}$ (respectively), one would guess that the corresponding eigenvalues are 1 and 3, respectively.
We verify our conjecture:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

4.1.3 Computing eigenvalues for a given matrix

An eigenvalue λ is a scalar for which $\text{Null}(A - \lambda I)$ contains nonzero vectors. In other words, if A is an $n \times n$ matrix, then λ is an eigenvalue if and only if

$$\text{nullity}(A - \lambda I) > 0 \qquad \text{or} \qquad \text{rank}(A - \lambda I) < n.$$

Example 4.1.8

Find the eigenvalues for $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$.

Looking at $A - \lambda I = \begin{bmatrix} 1-\lambda & 2 & 3 \\ 0 & 4-\lambda & 5 \\ 0 & 0 & 6-\lambda \end{bmatrix}$ the eigenvalues are $\lambda = 1, 4, 6$ because...

$(\boldsymbol{\lambda} = \mathbf{1})$ $A - I = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 3 & 5 \\ 0 & 0 & 5 \end{bmatrix}$ has only two nonzero columns, hence rank is at most 2.

($\boldsymbol{\lambda = 4}$) $A - 4I = \begin{bmatrix} -3 & 2 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 2 \end{bmatrix}$ has the first and second columns as scalar multiples of each other, hence the rank is as most 2.

($\boldsymbol{\lambda = 6}$) $A - 6I = \begin{bmatrix} -5 & 2 & 3 \\ 0 & -1 & 5 \\ 0 & 0 & 0 \end{bmatrix}$ has a row of all zeroes, hence the rank is at most 2.

$\boldsymbol{\lambda \neq 1, 4, 6}$ $A - \lambda I$ is in row echelon form with no rows of zeroes, so it must have rank 3.

Example 4.1.9

Find all eigenvalues for $A = \begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}$.

We have that

$$A - \lambda I = \begin{bmatrix} -3-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix} \xrightarrow[(-3-\lambda)R_2 \mapsto R_2]{3R_1 \mapsto R_1} \begin{bmatrix} 3(-3-\lambda) & 6 \\ 3(-3-\lambda) & (-3-\lambda)(2-\lambda) \end{bmatrix}$$
$$\xrightarrow{-R_1+R_2 \mapsto R_2} \begin{bmatrix} 3(-3-\lambda) & 6 \\ 0 & (-3-\lambda)(2-\lambda)-6 \end{bmatrix}$$

Observe that

$$(-3-\lambda)(2-\lambda) - 6 = -6 - 2\lambda + 3\lambda + \lambda^2 - 6 = \lambda^2 + \lambda - 12 = (\lambda+4)(\lambda-3)$$

so $\operatorname{rank}(A - \lambda I) < 2$ when $\lambda = -4, 3$.

As well, our first row operation implicitly assumed that $-3 - \lambda \neq 0$ (since multiplying by 0 is not a row operation), so we should check the case that $\lambda = -3$:

$$\operatorname{rank}(A - (-3)I) = \operatorname{rank}\left(\begin{bmatrix} 0 & 2 \\ 3 & 5 \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} 3 & 5 \\ 0 & 2 \end{bmatrix}\right) = 2$$

and therefore our *only* eigenvalues are

$$\lambda = -4, 3.$$

Exercise 4.1.10

Let $A = \begin{bmatrix} -4 & 2 & 6 \\ 0 & -3 & 2 \\ 0 & 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -4 & 0 & 7 \\ 0 & -3 & 2 \\ 0 & 3 & 2 \end{bmatrix}$.

1. Find all eigenvalues for A. For each eigenvalue, λ, find a basis for the eigenspace E_λ.
2. Find all eigenvalues for B. For each eigenvalue, λ, find a basis for the eigenspace E_λ.

Of course, if $A - \lambda I$ has rank less than n, then $A - \lambda I$ is not invertible, and for a 2×2 matrix

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we know that A is not invertible precisely when $ad - bc = 0$.

Example 4.1.11

Find all eigenvalues for $A = \begin{bmatrix} -3 & 2 \\ 3 & 2 \end{bmatrix}$.

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \det\left(\begin{bmatrix} -3-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix}\right) \\ &= (-3-\lambda)(2-\lambda) - 6 \\ &= \lambda^2 + \lambda - 12. \end{aligned}$$

Using our favorite method of solving for the zeroes of this polynomial, we exactly see that its zeroes are $\lambda = -4, 3$, which are precisely the eigenvalues we expected to get from the previous examples.

We ought to give a name to that $ad - bc$ quantitiy.

Definition: determinant (of a 2×2 matrix)

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the **determinant** of A is a real number, $\det(A)$, given by

$$\det(A) = ad - bc$$

4.1.4 Foreshadowing – A Basis of Eigenvectors

Example 4.1.12

Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. As we saw in Example 4.1.7, the eigenvectors for A were $[-1, 1]^T$ and $[1, 1]^T$ (with eigenvalues 1 and 3, respectively). The images below show the result of applying A to $\mathbb{R}^2$, first drawn with the coordinate grid from the standard basis, and then again drawn with the coordinate grid coming from our "eigenbasis." The eigenbasis figure shows that this matrix A behaves a bit like a diagonal matrix $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ as it simply stretches the grid in two different directions.

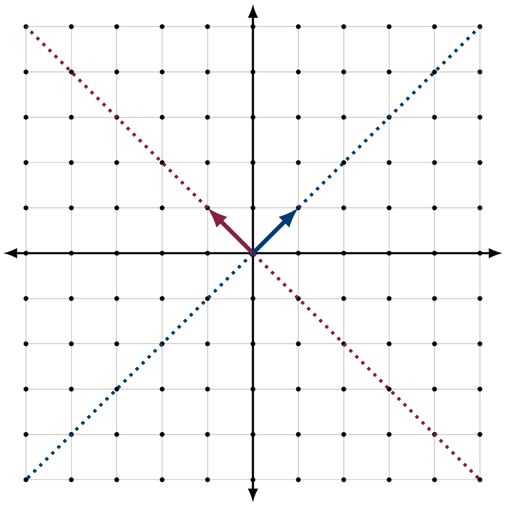


Figure 4.4: Before applying transformation A.
(Shown using the standard coordinate grid.)

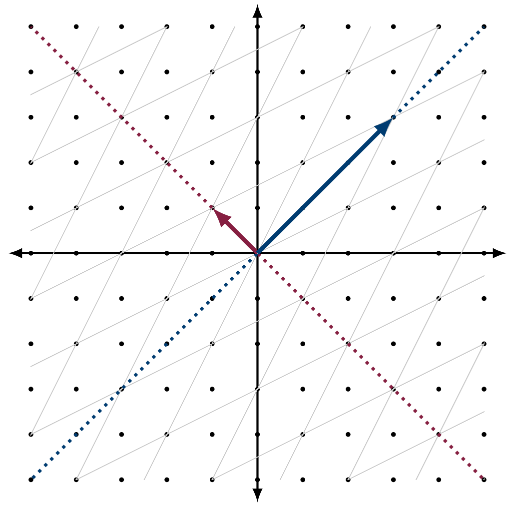


Figure 4.5: After applying transformation A.
(Shown with the transformed standard coordinate grid.)

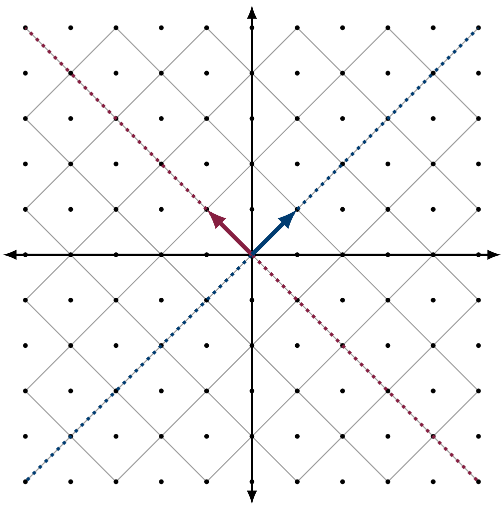


Figure 4.6: Before applying transformation A.
(Shown using the "*eigen*grid".)

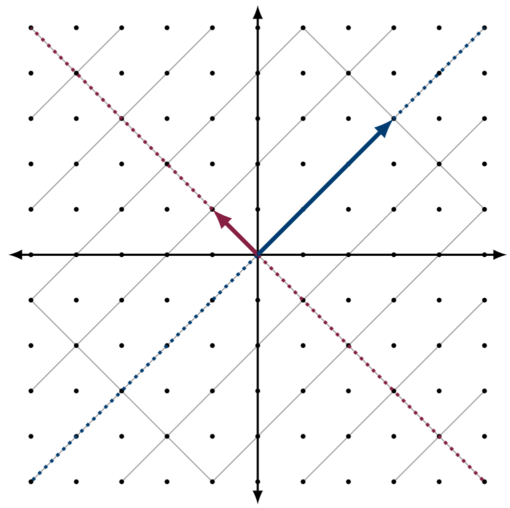


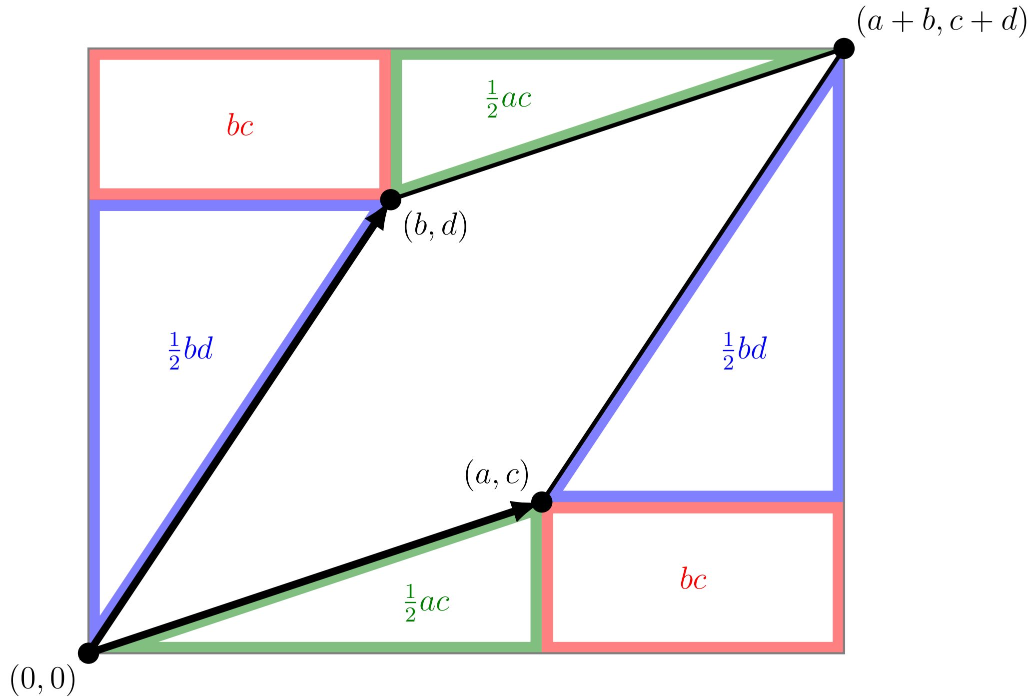
Figure 4.7: After applying transformation A.
(Shown with the transformed "*eigen*grid".)

4.2 Determinants

4.2.1 The Determinant - A Geometric Perspective

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 illustration, we will assume that $a > c > 0$ and $d > b > 0$.



The (signed) area of the parallelogram above is

$$\text{area} = (a+b)(c+d) - 2bc - bd - ac = ad - bc. = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

Armed with this geometric interpretation, we get a means of generalizing the definition to higher dimensions.

Definition: determinant

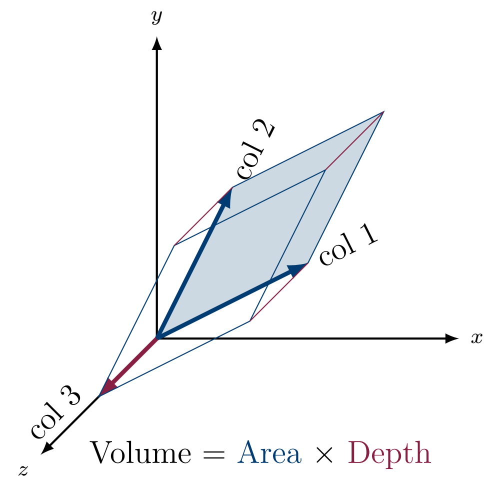
The **determinant** of an $n \times n$ matrix A the signed n-dimensional volume of the n-dimensional parallelepiped formed by the columns of A.

4.2.2 Computing the determinant of an $n \times n$ matrix, Part 1

Example 4.2.1: block diagonal 3×3

Find the determinant of the matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Drawing out the parallelepiped, we see that this is actually an "extruded object" - like a cylinder or a rectangular prism, it's just the area of the base parallelogram times the depth/height. In our case, the blue parallelogram is entirely contained in the xy-plane, and the third column is perpendicular to it.



So

$$\det(A) = \det \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = (2)\left(\det \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right) = (2)(4-1) = 6.$$

Of course, this nice setup doesn't happen in general, and so you can ask what happens when the columns are placed more generally in space. With a bit of playing around naively in $\mathbb{R}^3$, one could probably come up with the formula relating the volume to the areas of the three parallelograms forming the sides of the parallelepiped, in which case the determinant of the 3×3 should be related to the determinants of smaller matrices contained within the 3×3. In this way, one can iteratively compute the determinants of large $n \times n$ matrices.

Definition: minor, submatrix

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 4.2.2

For $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, find the $(2,1)$-minor.

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 4.2.3: 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

$$\det A = (-1)^{i+1} a_{i1} M_{i,1} + (-1)^{i+2} a_{i2} M_{i,2} + \cdots + (-1)^{i+n} a_{in} M_{i,n}.$$

or *along the* j^{th} *column* as the sum

$$\det A = (-1)^{1+j} a_{1j} M_{1,j} + (-1)^{2+j} a_{2j} M_{2,j} + \cdots + (-1)^{n+j} a_{nj} M_{n,j}.$$

The quantity $(-1)^{i+j} M_{i,j}$ is sometimes called the (i,j)**-cofactor** and the above sums are called **cofactor expansions**.

Since the determinant is always the same whether expanding along rows or columns, and row become columns via the transpose, then we quickly deduce the following important fact:

Corollary 4.2.4

$\det(A) = \det(A^T)$

Example 4.2.5

For $A = \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix}$, compute $\det A$ using cofactor expansion and expanding along the first row.

$$\begin{aligned} \det A &= a_{11} M_{1,1} - a_{12} M_{1,2} + a_{13} M_{1,3} \\ &= 3 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} - 1 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} + 4 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} \\ &= 3 \det \begin{bmatrix} 5 & -9 \\ 6 & 5 \end{bmatrix} - 1 \begin{bmatrix} -1 & -9 \\ 2 & 5 \end{bmatrix} + 4 \det \begin{bmatrix} -1 & 5 \\ 2 & 6 \end{bmatrix} \end{aligned}$$

$$= 3(79) - 1(13) + 4(-16) = 160.$$

Example 4.2.6

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 and see that this is the same number as in Example 4.2.5.

$$\begin{aligned} \det A &= -a_{12}M_{1,2} + a_{22}M_{2,2} - a_{32}M_{3,2} \\ &= -1 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} + 5 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} - 6 \det \begin{bmatrix} 3 & 1 & 4 \\ -1 & 5 & -9 \\ 2 & 6 & 5 \end{bmatrix} \\ &= -1 \det \begin{bmatrix} -1 & -9 \\ 2 & 5 \end{bmatrix} + 5 \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} - 6 \det \begin{bmatrix} 3 & 4 \\ -1 & -9 \end{bmatrix} \\ &= -1(13) + 5(7) - 6(-23) = 160. \end{aligned}$$

Remark. The powerful thing about Cofactor expansion is that the specific choice of row or column to use doesn't matter, so you get to pick the one most convenient to you (which, ideally, is the row/column with the most 0's).

Example 4.2.7

Compute $\det \begin{bmatrix} 2 & 3 & 5 & 7 \\ 0 & 11 & 13 & 17 \\ 0 & 0 & 19 & 23 \\ 0 & 0 & 0 & 29 \end{bmatrix}$.

We do cofactor expansion along the first column (twice)

$$\begin{aligned} \det \begin{bmatrix} 2 & 3 & 5 & 7 \\ 0 & 11 & 13 & 17 \\ 0 & 0 & 19 & 23 \\ 0 & 0 & 0 & 29 \end{bmatrix} &= 2 \det \begin{bmatrix} 11 & 13 & 17 \\ 0 & 19 & 23 \\ 0 & 0 & 29 \end{bmatrix} \\ &= 2(11) \det \begin{bmatrix} 19 & 23 \\ 0 & 29 \end{bmatrix} \\ &= 2(11)(19)(29) \end{aligned}$$

The last example is essentially the proof of the following:

Proposition 4.2.8: Determinants of triangular matrices

If A is an upper/lower-triangular (hence also diagonal) matrix, then $\det(A)$ is the product of the diagonal entries.

4.2.3 The Determinant: A Functional Perspective

Instead of defining the determinant purely geometrically, one can instead think of it as a multivariable function, that inputs n different vectors simultaneously, and outputs a single real number. That is, given a matrix $A = \begin{bmatrix} | & & | \\ \mathbf{a_1} & \cdots & \mathbf{a_n} \\ | & & | \end{bmatrix}$, one defines

$$F_{\det}(\mathbf{a_1}, \ldots, \mathbf{a_n}) = \det(A).$$

Proposition 4.2.9

$F_{\det}(\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}) = \det(I_n) = 1.$

Algebraic Proof of Proposition 4.2.9. The identity matrix is a diagonal matrix with all 1's along the diagonal. It follows from *Proposition* 4.2.8 that $\det(I) = \underbrace{(1)(1)\cdots(1)}_{n \text{ times}} = 1.$ □

Geometric Proof of Proposition 4.2.9. The volume of the unit n-cube is one. □

Proposition 4.2.10

$\det(A)$ is **multilinear** (it behaves like a linear transformation in each column). Specifically, for all $\mathbf{u}, \mathbf{w}, \mathbf{v_1}, \ldots, \mathbf{v_n}$ in $\mathbb{R}^n$ and all scalars k,

$$\begin{aligned} F_{\det}(k\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}) &= kF_{\det}(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}) \\ &\vdots \\ F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_{n-1}}, k\mathbf{v_n}) &= kF_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_{n-1}}, \mathbf{v_n}) \end{aligned}$$

and

$$\begin{aligned} F_{\det}(\mathbf{u} + \mathbf{w}, \mathbf{v_2}, \ldots, \mathbf{v_n}) &= F_{\det}(\mathbf{u}, \mathbf{v_2}, \ldots, \mathbf{v_n}) + F_{\det}(\mathbf{w}, \mathbf{v_2}, \ldots, \mathbf{v_n}) \\ &\vdots \\ F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_{n-1}}, k\mathbf{u} + \mathbf{w}) &= F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_{n-1}}, \mathbf{u}) + F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_{n-1}}, \mathbf{w}) \end{aligned}$$

One can prove this using the cofactor expansion formula, but it's probably easiest to see using a 2×2 matrix.

Algebraic Proof of Proposition 4.2.10. Addition: This is a straightforward computation.

$$\begin{aligned} \det \begin{bmatrix} (a_1 + a_2) & b \\ (c_1 + c_2) & d \end{bmatrix} &= (a_1 + a_2)d - b(c_1 + c_2) \\ &= a_1 d - bc_1 + a_2 d - bc_2 \\ &= \det \begin{bmatrix} a_1 & b \\ c_1 & d \end{bmatrix} + \det \begin{bmatrix} a_2 & b \\ c_2 & d \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \det \begin{bmatrix} a & (b_1 + b_2) \\ c & (d_1 + d_2) \end{bmatrix} &= a(d_1 + d_2) - (b_1 + b_2)c \\ &= ad_1 - b_1 c + ad_2 - b_2 c \end{aligned}$$

$$= det \begin{bmatrix} a & b_1 \\ c & d_1 \end{bmatrix} + \det \begin{bmatrix} a & b_2 \\ c & d_2 \end{bmatrix}$$

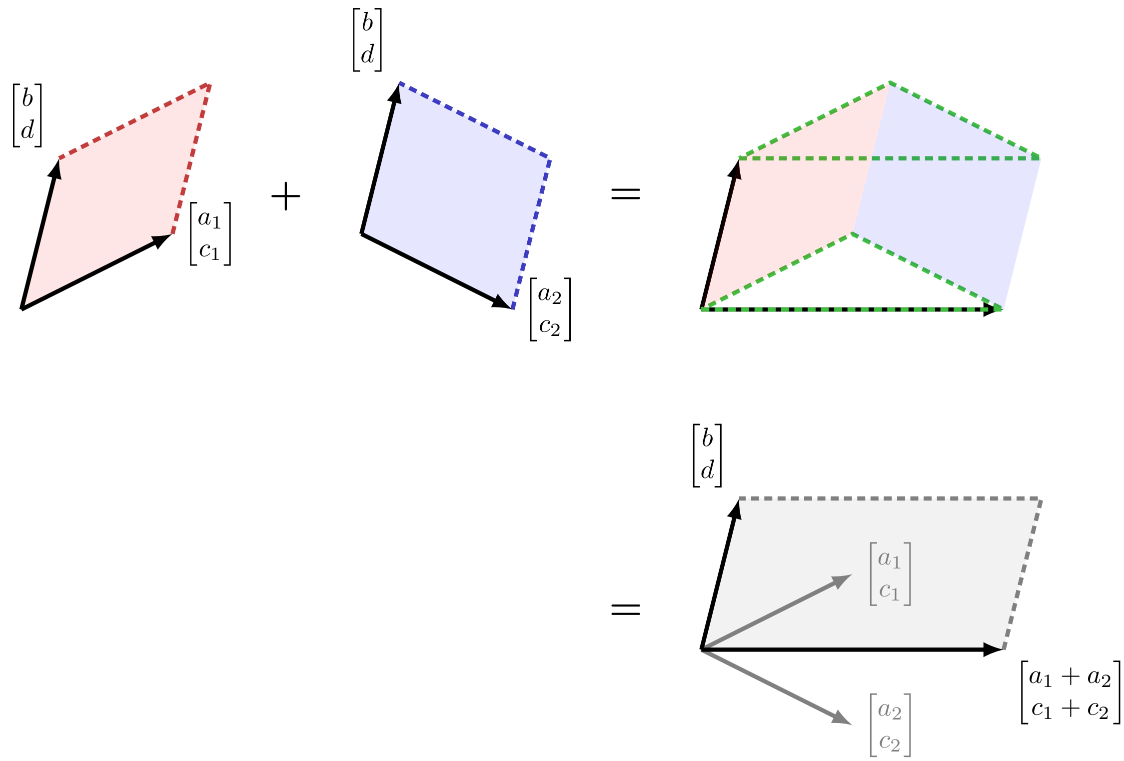
and that for any scalar k,

$$\det \begin{bmatrix} ka & b \\ kc & d \end{bmatrix} = (kad - bkc) = k(ad - bc) = k \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

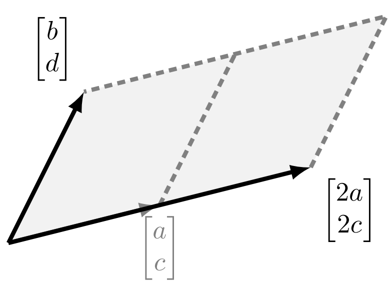
$$\text{and} \qquad \det \begin{bmatrix} a & kb \\ c & kd \end{bmatrix} = (akd - kbc) = k(ad - bc) \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

□

Geometric Proof of Proposition 4.2.10. Addition: (drawn only in the case of the first column)



scalar multiplication: (drawn in the case that $k = 2$)



□

Proposition 4.2.11

The determinant is **alternating** (that is, it switches sign whenever columns are swapped).

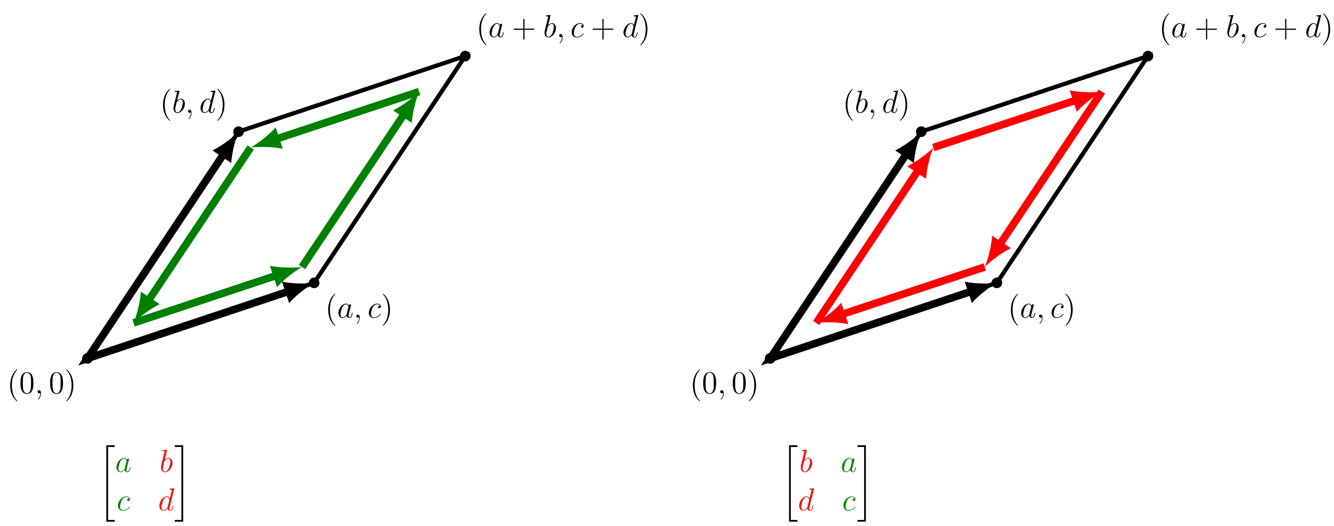
$$F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_i}, \ldots, \mathbf{v_j}, \ldots, \mathbf{v_n}) = -F_{\det}(\mathbf{v_1}, \ldots, \mathbf{v_j}, \ldots, \mathbf{v_i}, \ldots, \mathbf{v_n}).$$

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*), we essentially define this geometric meaning from the algebraic condition, so a geometric argument would be circular. But to give at least a geometric interpretation to this orientation business, the 2-dimensional perspective is that the first column tells you how to travel around your parallelogram, and a positive determinant indicates you're traveling counterclockwise, whereas a negative determinant indicates that you're traveling clockwise.



From the functional properties of the determinant we have that

Theorem 4.2.12

$\det(A) = 0$ if and only if A has linearly dependent columns.

and by Corollary 4.2.4

Theorem 4.2.13

$\det(A) = 0$ if and only if A has linearly dependent rows.

As such, we can add this to our theorem...

Theorem 4.2.14: The Fundamental Theorem of Invertible Matrices: Pt IV

Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

(a) A is invertible.

$\vdots$

(o) $\det(A) \neq 0$.

4.2.4 Computing the determinant of a $n \times n$ Matrix, Version 2

Observation

Combined with $\det(A) = \det(A^T)$ (Corollary 4.2.4), the functional properties of the determinant imply that modifying the rows of A make (tractible) modifications to $\det(A)$.

This, in turn, gives us a new strategy for computing the determinant using row operations (because once our matrix is in row echelon form, it is easy to compute the determinant.

Theorem 4.2.15: Determinants and Row Operations

Let A be an $n \times n$ matrix. For simplicity in the formulas below, we introduce the notation "$\underset{\text{row op}}{A}$" to denote the matrix equivalent A after it has had a specified row operation applied to it. Then we have the following:

1. Swapping Row i and Row j:

$$\det\left(\underset{R_i \leftrightarrow R_j}{A}\right) = -\det(A)$$

2. Multiplying Row i by a <u>nonzero</u> constant k:

$$\det\left(\underset{kR_i \mapsto R_i}{A}\right) = k\det(A)$$

3. Adding (a multiple of) Row j to Row i:

$$\det\left(\underset{R_i + kR_j \mapsto R_i}{A}\right) = \det(A)$$

Example 4.2.16

Compute the determinant of $A = \begin{bmatrix} 2 & 6 & 0 \\ -1 & 1 & 1 \\ -1 & -3 & 1 \end{bmatrix}$ using row operations.

$$\det(A) = \det \begin{bmatrix} 2 & 6 & 0 \\ -1 & 1 & 1 \\ -1 & -3 & 1 \end{bmatrix} \qquad \text{(initial matrix)}$$

$$\frac{1}{2}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ -1 & 1 & 1 \\ -1 & -3 & 1 \end{bmatrix} \qquad (\tfrac{1}{2}R_1 \mapsto R_1)$$

$$\frac{1}{2}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & 1 \\ -1 & -3 & 1 \end{bmatrix} \qquad (R_1 + R_2 \mapsto R_2)$$

$$\frac{1}{2}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad (R_1 + R_3 \mapsto R_3)$$

$$\frac{1}{2}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (R_2 - R_3 \mapsto R_2)$$

$$\frac{1}{2}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (R_2 - R_3 \mapsto R_2)$$

$$\frac{1}{8}\det(A) = \det \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (\tfrac{1}{4}R_2 \mapsto R_2)$$

$$\frac{1}{8}\det(A) = \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 1 \qquad (R_1 - 3R_2 \mapsto R_1)$$

from which it follows that $\det(A) = 8$.

4.2.5 Properties of Determinants and Determinants for Special Types of Matrices

Recall that there are three types of elementary matrices, call them

$$\underset{\text{swap}}{E}, \qquad \underset{k\ \text{scale}k}{E}, \qquad \underset{\text{add}}{E},$$

corresponding to the three types of elementary row operations – row swapping, row scaling by k, and row addition. We see that

$$\det\left(\underset{\text{swap}}{E}\right) = -1, \qquad \det\left(\underset{k\ \text{scale}}{E}\right) = k, \qquad \det\left(\underset{\text{add}}{E}\right) = 1,$$

and so since multiplication by an elementary matrix corresponds to a row operation, we get the following:

Lemma 4.2.17

If A is any $n \times n$ matrix and E is any $n \times n$ elementary matrix, then

$$\det(EA) = \det(E)\det(A)$$

Theorem 4.2.18: *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. and if A is invertible, $\det A^{-1} = \dfrac{1}{\det A}$.

Sketch of Proof.

1. If B scales the unit cube to a parallelepiped and scales the volume by $\det(B)$, then A further changes this parallelepiped and scales the volume by $\det(A)$. Thus AB scales the volume by $\det(A)\det(B)$. Alternatively, one can see this using Lemma 4.2.17:
   If A is not invertible, then it has linearly dependent columns and $\det(A) = 0$. Moreover, AB also has linearly dependent columns (think about the matrix-column representation), so $\det(AB) = 0$.
   If B is not invertible, essentially the same argument applies.
   If both A and B are invertible, then they are products of elementary matrices, so Lemma 4.2.17 applies.
2. If each edge of the paralellepiped is scaled by a factor of k, then the whole volume is scaled by a factor of k^n.
3. Given the first part, $1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1})$.

□

Example 4.2.19

Verify each part of the Theorem 4.2.18 using

$$A = \begin{bmatrix} 3 & -1 \\ 8 & -2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \text{and} \qquad k = 5.$$

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: eigenvalues, eigenvectors, eigenspace

If A is a square matrix, then $\det(A - xI)$ is a polynomial with indeterminate x and is called the **characteristic polynomial** of A (which we'll denote as $\mathrm{Char}_A(x)$). The **eigenvalues** of A are precisely the roots of the characteristic polynomial. For each eigenvalue λ, the corresponding **eigenspace** is $E_\lambda = \mathrm{Null}(A - \lambda I)$ and the nonzero vectors in E_λ are **eigenvectors**.

Remark. Non-square matrices do not have eigenvalues/eigenvectors, because if $v \in \mathbb{R}^n$ is an eigenvector for A, then $Av = \lambda v$ implies that $v \in \mathbb{R}^n$ as well, hence A is $n \times n$. 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: Revisiting Example 4.1.8

Find the eigenvalues and eigenvectors for $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$.

Using Proposition 4.2.8, upper/lower-triangular (and hence also diagonal) matrices have simple determinants:

$$\mathrm{Char}_A(x) = \det(A - xI) = \begin{bmatrix} 1-x & 2 & 3 \\ 0 & 4-x & 5 \\ 0 & 0 & 6-x \end{bmatrix} = (1-x)(4-x)(6-x)$$

so the eigenvalues are $\lambda = 1, 4, 6$.

Proposition 4.3.2: *Poole* - Theorem 4.15

The diagonal entries of an upper/lower-triangular (and hence also diagonal) matrix are the diagonal entries.

Proof. If A is triangular, then by Proposition 4.2.8, the characteristic polynomial is

$$\mathrm{Char}_A(\lambda) = \det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

□

Proposition 4.3.3: The determinant is the product of the eigenvalues

If A is an $n \times n$ square matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ (not necessarily all distinct), then $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$, the product of all of the eigenvalues.

Proof. Since

$$\mathrm{Char}_A(x) = \det(A - xI) = (\lambda_1 - x)(\lambda_2 - x) \cdots (\lambda_n - x)$$

then

$$\det(A) = \det(A - 0I) = (\lambda_1 - 0)(\lambda_2 - 0)\cdots(\lambda_n - 0) = \lambda_1\lambda_2\cdots\lambda_n.$$

□

From this it follows that we have yet another test for invertibility:

Theorem 4.3.4: The Fundamental Theorem of Invertible Matrices: Pt V

Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

(a) A is invertible.

⋮

(p) 0 is <u>not</u> an eigenvalue of A.

Example 4.3.5

Find the eigenvalues and eigenvectors for $A = \begin{bmatrix} 2 & 12 & 10 \\ 0 & -4 & -4 \\ 1 & 2 & 1 \end{bmatrix}$.

We first compute $\det(A - xI)$ via cofactor expansion along the first column.

$$\begin{aligned}
\mathrm{Char}_A(x) = \det(A - xI) &= \det \begin{bmatrix} 2-x & 12 & 10 \\ 0 & -4-x & -4 \\ 1 & 2 & 1-x \end{bmatrix} \\
&= (2-x)\det\begin{bmatrix} -4-x & -4 \\ 2 & 1-x \end{bmatrix} + 1\det\begin{bmatrix} 12 & 10 \\ -4-x & -4 \end{bmatrix} \\
&= (2-x)\left((-4-x)(1-x)+8\right) + 1\left(-48 - 10(-4-x)\right) \\
&= -\left(x^3 + x^2 - 12x\right) \\
&= -x(x+4)(x-3)
\end{aligned}$$

The characteristic polynomial factors nicely and the eigenvalues are $-4, 0, 3$. The corresponding eigenspaces are

$$\begin{aligned}
E_{-4} = \mathrm{Null}(A + 4I) &= \mathrm{Null}\left(\begin{bmatrix} 6 & 12 & 10 \\ 0 & 0 & -4 \\ 1 & 2 & 5 \end{bmatrix}\right) = \mathrm{Span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}\right), \\
E_0 = \mathrm{Null}(A - 0I) &= \mathrm{Null}\left(\begin{bmatrix} 2 & 12 & 10 \\ 0 & -4 & -4 \\ 1 & 2 & 1 \end{bmatrix}\right) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right), \\
E_3 = \mathrm{Null}(A - 3I) &= \mathrm{Null}\left(\begin{bmatrix} -1 & 12 & 10 \\ 0 & -7 & -4 \\ 1 & 2 & -2 \end{bmatrix}\right) = \mathrm{Span}\left(\begin{bmatrix} 22 \\ -4 \\ 7 \end{bmatrix}\right).
\end{aligned}$$

Example 4.3.6

Let $A = \begin{bmatrix} 3 & -2 & -4 \\ 0 & 2 & 7 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -2 & -4 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Both matrices have the same characteristic polynomials:

$$\text{Char}_A(x) = \text{Char}_B(x) = (3-x)(2-x)^2.$$

Do they have the same eigenvectors?

We first compute the eigenspaces for the matrix A.

$$E_3(A) = \text{Null}(A - 3I) = \text{Null}\left(\begin{bmatrix} 0 & -2 & -4 \\ 0 & -1 & 7 \\ 0 & 0 & -1 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right),$$

$$E_2(A) = \text{Null}(A - 2I) = \text{Null}\left(\begin{bmatrix} 1 & -2 & -4 \\ 0 & 0 & 7 \\ 0 & 0 & 0 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}\right)$$

Now we compute the eigenspaces for the matrix B.

$$E_3(B) = \text{Null}(A - 3I) = \text{Null}\left(\begin{bmatrix} 0 & -2 & -4 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right),$$

$$E_2(B) = \text{Null}(A - 2I) = \text{Null}\left(\begin{bmatrix} 1 & -2 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}\right)$$

They do *not* share the same eigenvectors.

Notice that the above 3×3 matrices each had eigenvalues 3, 2, 2 (counted with multiplicity). But in one of them, the eigenspace E_2 was 2-dimensional, and in the other the eigenspace E_2 was 1-dimensional. We should give some names to what we've witnessed.

Definition: algebraic multiplicity, geometric multiplicity

The **algebraic multiplicity** of an eigenvalue λ is the multiplicity as a root of the characteristic polynomial (i.e., the number of times that the factor $(\lambda - x)$ appears in the characteristic polynomial $\text{Char}_A(x)$), and the **geometric multiplicity** is the dimension of the eigenspace E_λ, i.e., $\text{nullity}(A - \lambda I)$.

Remark. There's no good notation for algebraic and geometric multiplicity, so for simplicity we'll write $\text{AlgMult}(\lambda)$ and $\text{GeoMult}(\lambda)$, respectively.

Example 4.3.7

What are the algebraic and geometric multiplicities of the eigenvalues of matrix A and B in Example 4.3.6?

$$\text{AlgMult}_A(3) = 1 \qquad \text{AlgMult}_B(3) = 1$$
$$\text{GeoMult}_A(2) = 2 \qquad \text{GeoMult}_B(2) = 2$$

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.

Theorem 4.3.8

If $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues with eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_m}$ (respectively), then the set $\{\mathbf{v_1}, \ldots, \mathbf{v_m}\}$ is linearly independent.

Proof. We prove this only in the case that $m = 2$. Consider

$$x_1\mathbf{v_1} + x_2\mathbf{v_2} = \mathbf{0}$$

The goal is to show that $x_1, x_2 = 0$. We can do two different things to this equation: multiply it all by A, and multiply it all by λ_1.

$$A\left(x_1\mathbf{v_1} + x_2\mathbf{v_2}\right) = x_1\lambda_1\mathbf{v_1} + x_2\lambda_2\mathbf{v_2} = \mathbf{0}$$
$$\lambda_1\left(x_1\mathbf{v_1} + x_2\mathbf{v_2}\right) = x_1\lambda_1\mathbf{v_1} + x_2\lambda_1\mathbf{v_2} = \mathbf{0}$$

Subtracting these equations from each other yields

$$x_2(\lambda_1 - \lambda_2)\mathbf{v_2} = \mathbf{0}.$$

So if $\lambda_2 \neq \lambda_1$ and $\mathbf{v_2} \neq \mathbf{0}$, then it must be that $x_2 = 0$ (and this forces $x_1 = 0$). Therefore $\{\mathbf{v_1}, \mathbf{v_2}\}$ is a linearly independent set. □

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.

Write

$$A = \begin{bmatrix} 0 & 18 \\ 8 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} = LU$$

It is easy to see (although the reader should verify) that A, L, and U have the following eigenvalues:

A	L	U
$\alpha_1 = -12$	$\lambda_1 = -2$	$\mu_1 = 4$
$\alpha_2 = 12$	$\lambda_2 = 2$	$\mu_2 = 9$

$$\det(A) = -144 \qquad \det(L) = -4 \qquad \det(U) = 36$$

Given that $\det(A) = \det(L)\det(U)$, we might have naively hoped that a product of matrices would have eigenvalues that are products of eigenvalues – that $\alpha_i = \lambda_j\mu_k$ – but that's not the case. *sad trombone noises.*

However, things are nice with matrix powers!

Example 4.3.9

Consider $A = \begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$. Find the eigenvalues and eigenvectors for both A^{-1} and A^2.

The characteristic polynomial is

$$\text{Char}_A(x) = (2-x)(3-x)$$

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 $A^{-1} = \dfrac{1}{6}\begin{bmatrix} 4 & 2 \\ -1 & 1 \end{bmatrix}$ 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!

One may wonder if the results above hold for 3×3 matrices

Exercise 4.3.10

Let $A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix}$. The reader is invited to check that

$$A^2 = \begin{bmatrix} 4 & 21 & 54 \\ 0 & 25 & 72 \\ 0 & 0 & 49 \end{bmatrix} \qquad \text{and} \qquad A^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{3}{10} & -\frac{1}{35} \\ 0 & \frac{1}{5} & -\frac{6}{35} \\ 0 & 0 & \frac{1}{7} \end{bmatrix}.$$

Find the eigenvalues for A and A^2 and A^{-1}.

INCOMPLETE Eigenvalues of A are $2, 5, 7$ with eigenvectors $[1, 0, 0]$, $[1, 1, 0]$, $[13, 15, 5]$, respectively.
Eigenvalues of A^2 are $2^2, 5^2, 7^2$ with eigenvectors $[1, 0, 0]$, $[1, 1, 0]$, $[13, 15, 5]$, respectively.

Eigenvalues of A^{-1} are $2^{-1}, 5^{-1}, 7^{-1}$ with eigenvectors $[1, 0, 0]$, $[1, 1, 0]$, $[13, 15, 5]$, respectively.

Exercise 4.3.11

Let $A = \begin{bmatrix} 0 & 1 & 1 \\ -3 & 4 & 3 \\ 3 & -1 & 2 \end{bmatrix}$. The reader is invited to check that

$$A^2 = \begin{bmatrix} 0 & 3 & 5 \\ -3 & 10 & 15 \\ 9 & -3 & 4 \end{bmatrix} \qquad \text{and} \qquad A^{-1} = \frac{1}{6}\begin{bmatrix} 11 & -3 & -1 \\ 15 & -3 & -3 \\ -9 & 2 & 3 \end{bmatrix}.$$

Find the eigenvalues and eigenvectors for A, A^2, A^{-1}.

$$\begin{aligned}
\det(A - xI) &= -x^3 + 6x^2 - 11x + 6 = (1-x)(2-x)(3-x) \\
\det(A^2 - xI) &= -x^3 + 14x^2 - 49x = (1-x)(4-x)(9-x) \\
\det(A^{-1} - xI) &= \frac{1}{6}\left(-6x^3 + 11x^2 - 6x + 1\right) = \frac{1}{6} = \frac{1}{6}(1-x)(1-2x)(1-3x) \\
&= (1-x)\left(\frac{1}{2} - x\right)\left(\frac{1}{3} - x\right)
\end{aligned}$$

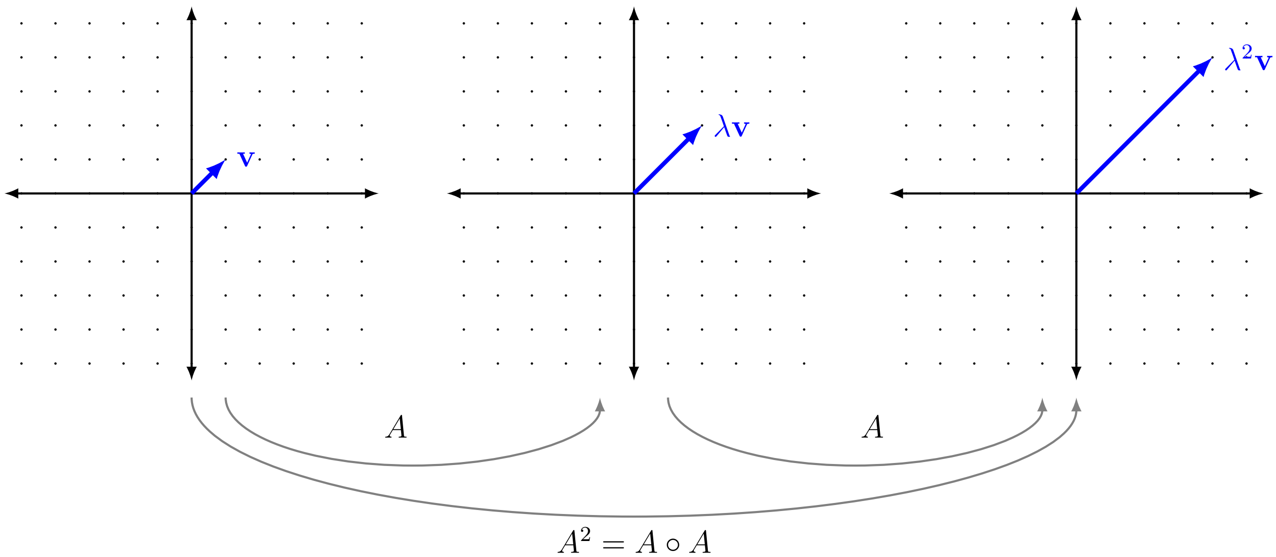
INCOMPLETE

Theorem 4.3.12

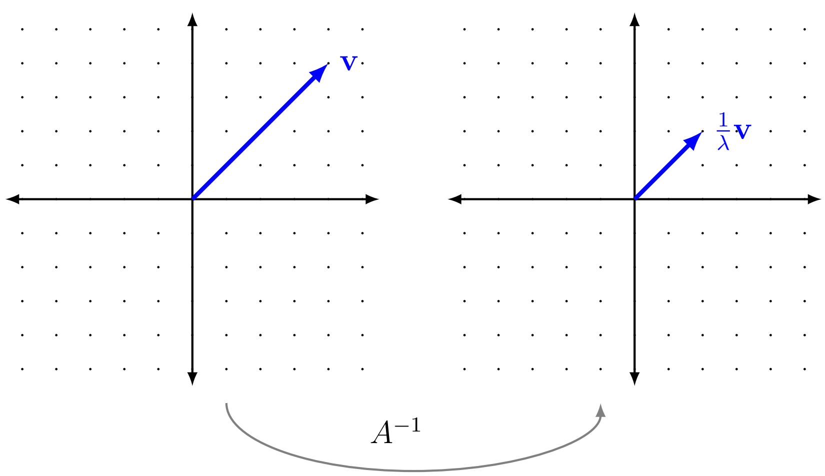
Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\mathbf{v}$.

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

Before jumping in, it would be good to see what our target goal is. Recall the following figures drawn in Example 4.1.12. The figures below show the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ applied to $\mathbb{R}^2$, first using the coordinate grid from the standard basis, and then again using the coordinate grid formed by the eigenvectors $[-1, 1]^T$ and $[1, 1]^T$. The second set of figures shows that this matrix A behaves a bit like a diagonal matrix $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ as it simply stretches the grid in two different directions.

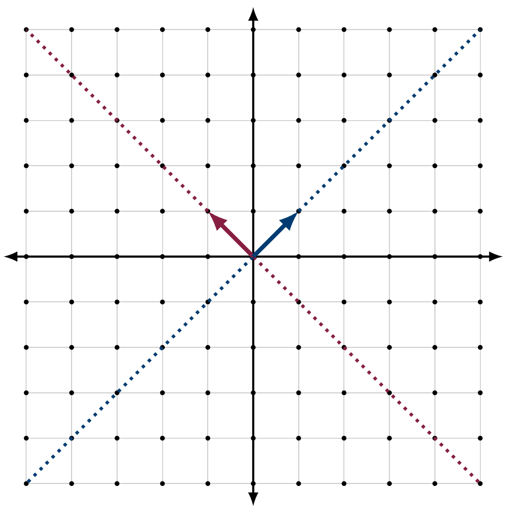


Figure 4.8: Before applying transformation A.
(Shown using the standard coordinate grid.)

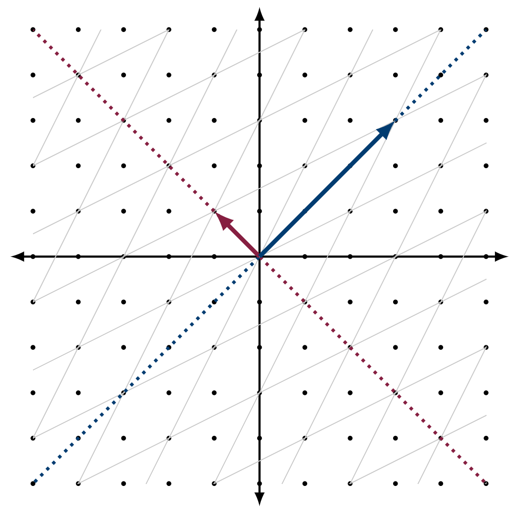


Figure 4.9: After applying transformation A.
(Shown with the transformed standard coordinate grid.)

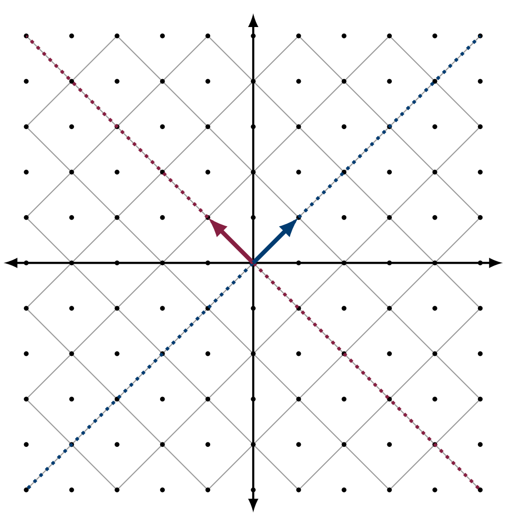


Figure 4.10: Before applying transformation A.
(Shown using the "*eigen*grid".)

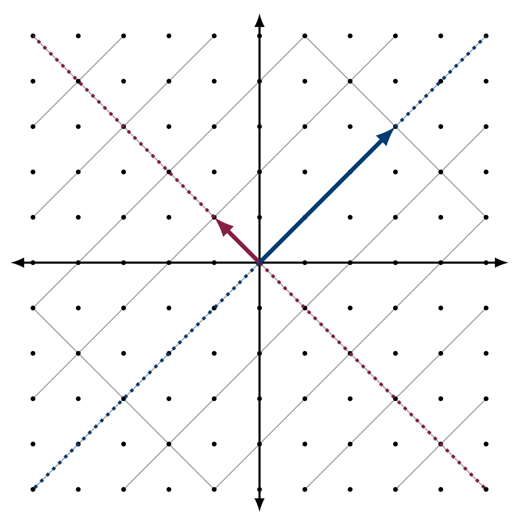


Figure 4.11: After applying transformation A.
(Shown with the transformed "*eigen*grid".)

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: similar matrices

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

Show that $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$ are similar.

With $P = \begin{bmatrix} -3 & 1 \\ 1 & 1 \end{bmatrix}$, we have

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} -3 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{-4} \begin{bmatrix} 1 & -1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} = B \end{aligned}$$

Example 4.4.2

Show that matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are <u>not</u> 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 it's impossible that both $(wx+yz) = 1$ and $(-wx-yz) = 1$ (otherwise $wx+yz = -(wx+yz)$, hence $wx + yz = 0$).

Example 4.4.3

Show that the matrices $R = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ are siilar.

Just as before, we solve

$$A\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} x & y \\ z & w \end{bmatrix}R$$

which yields $x = z$ and $y = -w$. So we can pick our favorite x and y-values satisfying these conditions so long as we don't accidentally pick our determinant to be 0. Forcing the determinant to be 1

$$P = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix}.$$

Remark. Think of what the above suggests. R is the matrix that reflects over the x-axis, and A is the matrix that reflects over the line $y = x$. Given that they're both simple reflections, they are similar from a geometric perspective. But moreover, P is precisely the matrix that rotates counter-clockwise by $\frac{\pi}{4}$, i.e., the matrix that takes the x-axis to the line $y = x$. So the relationship

$$A = PRP^{-1}$$

means that, to perform A, we first rotate $y = x$ to the x-axis, reflect across the x-axis, and rotate the x-axis back up to $y = x$.

4.4.1 Properties of similarity and similar matrices

Theorem 4.4.4: *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 4.4.5: *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.

Partial proof sketch. Let P be a matrix for which $B = P^{-1}AP$.

(a) $\det B = \det(P^{-1}AP) = (\det P^{-1})\,(\det A)\,(\det P) = \left(\dfrac{1}{\det P}\right)(\det A)\,(\det P) = \det A$

(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) $\text{Char}_B(x) = \det(B - xI) = \det(P^{-1}AP - xP^{-1}IP) = \det(P^{-1}(A - xI)P) = \det(A - xI) = \text{Char}_A(x)$

(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 Example 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 4.4.6

Use Theorem 4.4.5 to argue why matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ from Example 4.4.2 are not similar.

A and B are not similar because $\det A = -1$ and $\det B = 1$.

Example 4.4.7

Use Theorem 4.4.5 to argue why matrices $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}$ are not similar.

Although both matrices have the same rank ($\text{rank}\,A = \text{rank}\,B = 2$) and determinant ($\det A = \det B = 6$), they are not similar because their characteristic polynomials are different ($\text{Char}_A(x) = (x-2)(x-3)$ and $\text{Char}_B(x) = (x-1)(x-6)$).

4.4.2 Diagonalization

Definition: diagonalizable

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, $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ is diagonalizable since it is similar to the diagonal matrix

$B = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$.

Notice that the characteristic polynomial of A above is

$$\text{Char}_A(x) = \det(A - xI) = (2-x)(4-x) - 3 = x^2 - 6x + 5 = (x-1)(x-5)$$

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

$$AP = PD$$

$$A \begin{bmatrix} | & & | \\ \mathbf{p_1} & \cdots & \mathbf{p_n} \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{p_1} & \cdots & \mathbf{p_n} \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\begin{bmatrix} | & & | \\ A\mathbf{p_1} & \cdots & A\mathbf{p_n} \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\mathbf{p_1} & \cdots & \lambda_n\mathbf{p_n} \\ | & & | \end{bmatrix}$$

and so $A\mathbf{p_i} = \lambda_i \mathbf{p_i}$, hence the λ_i are eigenvalues for A with corresponding eigenvectors $\mathbf{p_i}$. We also see that P is invertible if and only if all n of the eigenvectors $\mathbf{p_i}$ are linearly independent.

Example 4.4.10

Determine whether or not the following matrix is diagonalizable: $A = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$.

The characteristic polynomial for A is

$$\det(A - \lambda I) = \det \begin{bmatrix} 3-\lambda & 1 & 0 \\ 0 & 3-\lambda & 1 \\ 0 & 0 & 3-\lambda \end{bmatrix} = (3-\lambda)^3$$

so A has an single eigenvalue of 3 with algebraic multiplicity 3. The corresponding eigenspace is

$$E_3 = \text{Null}(A - 3I) = \text{Span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right).$$

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

Determine whether or not the following matrix is diagonalizable: $A = \begin{bmatrix}4 & -3 & -3\\3 & -2 & -3\\-1 & 1 & 2\end{bmatrix}$.

$$\det(A - \lambda I) = \det\begin{bmatrix}4-\lambda & -3 & -3\\3 & -2-\lambda & -3\\-1 & 1 & 2-\lambda\end{bmatrix} = -(\lambda-1)^2(\lambda-2)$$

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:

$$P^{-1}AP = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 2\end{bmatrix}.$$

What these examples have highlighted is that a matrix may only fail to be diagonalizable if it has repeated eigenvalues.

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

For every eigenvalue λ,

$$1 \leq \text{GeoMult}(\lambda) \leq \text{AlgMult}(\lambda)$$

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, $\text{GeoMult}(\lambda_i) = \text{AlgMult}(\lambda_i)$.

Example 4.4.16

Let $A = \begin{bmatrix} 3 & -1 & 2 \\ 3 & -1 & 6 \\ -2 & 2 & -2 \end{bmatrix}$. Determine whether or not A is diagonalizable. If it is, find an invertible matrix P and a diagonal matrix D for which $P^{-1}AP = D$.

A 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

Let $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 4 \\ 0 & 0 & 3 \end{bmatrix}$. Determine whether or not A is diagonalizable. If it is, find an invertible matrix P and a diagonal matrix D for which $P^{-1}AP = D$.

A 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.

4.4.3 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

$$D^k = (P^{-1}AP)^k = P^{-1}A^kP \qquad \implies \qquad A^k = PD^kP^{-1}.$$

In this way, computing the k^{th} power of A is only as computationally difficult as diagonalizing A. (For the record, this is actually very fast.)

Example 4.4.18

Let $A = \begin{bmatrix} 11 & -6 \\ 15 & -8 \end{bmatrix}$. Find a general formula for A^k.

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,

$$A^k = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 1^k & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 5 & -3 \end{bmatrix} = \begin{bmatrix} -9 + 10(2^k) & 6 - 6(2^k) \\ -15 + 15(2^k) & 10 - 9(2^k) \end{bmatrix}.$$

Chapter 5

Orthogonality

5.1 Orthogonality in $\mathbb{R}^n$

Recall from Corollary 1.2.7 that two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are perpendicular/orthogonal if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

Definition: orthogonal set

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called **orthogonal** if $\mathbf{v_i} \cdot \mathbf{v_j} = 0$ whenever $i \neq j$. An orthogonal set of vectors is said to be **orthonormal** if, additionally, $\mathbf{v_i} \cdot \mathbf{v_i} = 1$ for every i.

Example 5.1.1

The standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$ is an orthogonal set.

This is straightforward and left as an exercise

Example 5.1.2

Verify that the set $\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right\}$ is an orthogonal set of vectors in $\mathbb{R}^3$.

This is straightforward to verify:

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

Theorem 5.1.3: Orthogonal $\implies$ Linearly Independent

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 so that

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

We aim to show that each $a_i = 0$. Toward that goal, notice that for each $i = 1, \ldots, k$, we have

$$\begin{aligned} 0 = \mathbf{0} \cdot \mathbf{v}_i &= (a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k) \cdot \mathbf{v}_i \\ &= a_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + a_k(\mathbf{v}_k \cdot \mathbf{v}_i) \qquad \text{[distributive property]} \end{aligned}$$

Since the $\mathbf{v}_i$'s form an orthogonal set, we have that most of the dot products above are 0, so the equation above reduces to

$$0 = a_i(\mathbf{v}_i \cdot \mathbf{v}_i)$$

And since each $\mathbf{v}_i$ is nonzero, then $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$, so it must be that $a_i = 0$. □

Definition: orthogonal basis

A basis $\mathcal{B}$ for $\mathbb{R}^n$ is an **orthogonal** (*resp.* **orthonormal**) **basis** if it is also an orthogonal (*resp.* orthonormal) set.

Example 5.1.4

The sets in Examples 5.1.1 and 5.1.2 are orthogonal bases.

Remark. The definition of an orthogonal basis also applies to subspaces.

Since every finite-dimensional vector space has a basis, one may be led to ask the following:

Question 5.1.5: Motivation for Chapter 5

Does every vector (sub)space have an <u>orthogonal</u> basis?

To answer, we need to revisit the dot product.

5.1.1 The Dot Product (Revisited)

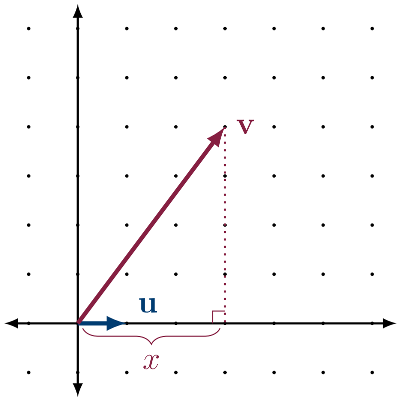
Up until now, the dot product has been a purely algebraic operation. Since orthogonality is fundamentally a geometric condition, then we should find a geometric description of the dot product by way of two similar examples.

Example 5.1.6

Let

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Compute $\mathbf{u} \cdot \mathbf{v}$.



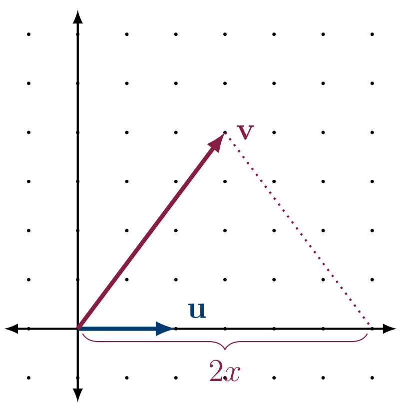
$\mathbf{u} \cdot \mathbf{v} = x$, which is just the first component of $\mathbf{v}$.

Example 5.1.7

Let

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Compute $\mathbf{u} \cdot \mathbf{v}$.



$\mathbf{u} \cdot \mathbf{v} = 2x$, which is the first component of $\mathbf{v}$, scaled by $\|\mathbf{u}\| = 2$.

Loosely-speaking: the dot product $\mathbf{u} \cdot \mathbf{v}$ is the "length of the shadow that $\mathbf{v}$ casts on $\mathbf{u}$, scaled by the length of $\mathbf{u}$." (Of course, since $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$, then the symmetric interpretation is also valid.) One can also see this interpretation using the angle formulation

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

where $|\mathbf{v}| \cos\theta$ is interpreted as "the amount of $\mathbf{v}$ in the direction of $\mathbf{u}$."

5.1.2 Orthogonal Projections Onto Vectors

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: (orthogonal) projection onto a vector

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ (with $\mathbf{u} \neq \mathbf{0}$), the **(orthogonal) projection of v onto u** is the vector $\text{proj}_{\mathbf{u}}(\mathbf{v})$ given by

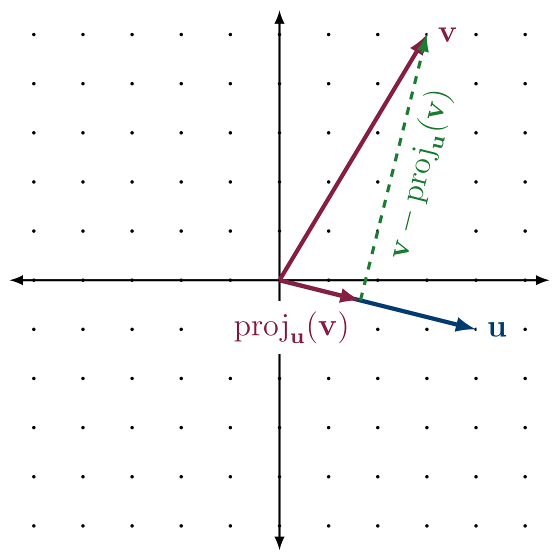
$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}}{\|\mathbf{u}\|} \cdot \mathbf{v}\right)\frac{\mathbf{u}}{\|\mathbf{u}\|} = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}.$$

Example 5.1.8

Let $\mathbf{u} = [4, -1]^T$ and $\mathbf{v} = [3, 5]^T$. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$.

Using the formula from the definition we have

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} = \left(\frac{7}{17}\right)\mathbf{u} = \begin{bmatrix} 28/17 \\ -7/17 \end{bmatrix}.$$



Remark. As the notation suggests, for any nonzero vector $\mathbf{u}$, the projection onto $\mathbf{u}$ is a function:

$$\text{proj}_{\mathbf{u}} : \mathbb{R}^n \to \mathbb{R}^n.$$

Exercise 5.1.9

Show that the function $\text{proj}_{\mathbf{u}} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation.
Hint: this follows from the bilinearity of the dot product.

Example 5.1.10

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Compute

$$\text{proj}_{\mathbf{u}}(\mathbf{e_1}) \qquad \text{and} \qquad \text{proj}_{\mathbf{u}}(\mathbf{e_2})$$

where $\mathbf{e_1}, \mathbf{e_2}$ are the standard basis vectors for $\mathbb{R}^2$. Then find the standard matrix for the function $\text{proj}_{\mathbf{u}}(\,)$.

INCOMPLETE WORK

$$\text{proj}_{\mathbf{u}}(\mathbf{e_1}) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} \qquad \text{and} \qquad \text{proj}_{\mathbf{u}}(\mathbf{e_2}) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}.$$

Standard matrix:

$$[\text{proj}_{\mathbf{u}}(\ )] = \begin{bmatrix} | & | \\ \text{proj}_{\mathbf{u}}(\mathbf{e_1}) & \text{proj}_{\mathbf{u}}(\mathbf{e_2}) \\ | & | \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}.$$

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: orthogonal complement

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 vectors orthogonal to W is a subspace called the **orthogonal complement of W** and is denoted $W^\perp$.

Theorem 5.2.1: *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 $\mathbf{w_i}$.

Proof.

1. 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.
2. [INCOMPLETE]
3. [INCOMPLETE]
4. [INCOMPLETE]

□

Example 5.2.2

Suppose W is the xy-plane in $\mathbb{R}^3$ (i.e. the set of vectors $[x, y, 0]^T$). Find $W^\perp$

Item #4 in Theorem 5.2.1 tells us that, if we can find a basis for W, then we just have to figure out when a given vector is perpendicular to our basis vectors. In this case, we have

$$W = \text{Span}\left(\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{\mathbf{w_1}}, \underbrace{\begin{bmatrix}0\\1\\0\end{bmatrix}}_{\mathbf{w_2}}\right)$$

. Suppose $\mathbf{v} = [v_1, v_2, v_3]$ is an element of $W^\perp$. Then

$$\begin{aligned}\mathbf{v} \cdot \mathbf{w_1} = 0 &\implies v_1 = 0\\ \mathbf{v} \cdot \mathbf{w_2} = 0 &\implies v_2 = 0\end{aligned}$$

So all vectors $\mathbf{v}$ in $W^\perp$ have the form

$$\mathbf{v} = \begin{bmatrix}0\\0\\v_3\end{bmatrix}.$$

In other words, $W^\perp$ is the z-axis in $\mathbb{R}^3$.

When finding the basis for a subspace W, we found it convenient to encode it as the column space of a matrix A. Can we use this matrix to find the orthogonal complement as well?

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.

Exercise 5.2.4

Suppose $W = \text{Col}\left(\begin{bmatrix}1 & 0\\0 & 1\\0 & 0\end{bmatrix}\right)$ is the xy-plane in $\mathbb{R}^3$ (i.e. the set of vectors $[x, y, 0]^T$). Verify that $W^\perp$ is the z-axis using this new null space technique.

Example 5.2.5

Let $W = \text{Span}\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\-1\\0\end{bmatrix}\right)$ be a plane in $\mathbb{R}^3$. Find $W^\perp$.

We observe that $W = \begin{bmatrix}1 & 2\\2 & -1\\3 & 0\end{bmatrix}$. Applying **??**,

$$W^\perp = \text{Null}\left(\begin{bmatrix}1 & 2 & 3\\2 & -1 & 0\end{bmatrix}\right).$$

Since

$$\text{RREF}\left(\begin{bmatrix}1 & 2 & 3\\2 & -1 & 0\end{bmatrix}\right) = \begin{bmatrix}1 & 0 & 3/5\\0 & 1 & 6/5\end{bmatrix}$$

we deduce that $W^\perp = \text{Span}\left(\begin{bmatrix}3/5\\6/5\\-1\end{bmatrix}\right)$.

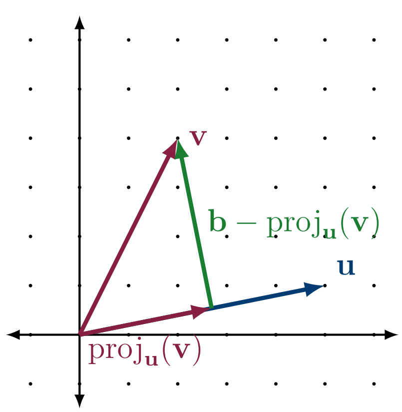
Example 5.2.6

Let $W = \text{Span}\left(\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\\0\end{bmatrix}\right)$ be a plane in $\mathbb{R}^4$. Find $W^\perp$.

INCOMPLETE

5.2.2 Orthogonal Projections Onto Subspaces

Notice that the vector $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is perpendicular to $\mathbf{u}$,



which we can verify algebraically:

$$\begin{aligned}(\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})) \cdot \mathbf{u} &= \mathbf{v} \cdot \mathbf{u} - \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u} \cdot \mathbf{u} \\ &= \mathbf{v} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v} \\ &= \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} \\ &= 0.\end{aligned}$$

Since the dot product of these vectors is 0, by Corollary 1.2.7 they must be perpendicular.

So, for two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, we have that $\text{proj}_{\mathbf{u}}(\mathbf{v})$ is a vector contained in $\text{Span}(\mathbf{u})$, and $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is contained in $\mathbf{u}^{\perp}$. In this way, we can "decompose" $\mathbf{v}$ into a sum of two vectors. If $W = \text{Span}(\mathbf{u})$ then we have

$$\mathbf{v} = \underbrace{\text{proj}_{\mathbf{u}}(\mathbf{v})}_{\text{in } W} + \underbrace{(\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v}))}_{\text{in } W^{\perp}}$$

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 v onto W** is

$$\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{w_1}}(\mathbf{v}) + \cdots + \text{proj}_{\mathbf{w_k}}(\mathbf{v})$$

and the **component of v orthogonal to W** is

$$\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(\mathbf{v}) = \text{proj}_{W^{\perp}}(\mathbf{v}).$$

The **orthogonal decomposition of v with respect to W** is the formula

$$\mathbf{v} = \underbrace{\text{proj}_W(\mathbf{v})}_{\text{in } W} + \underbrace{\text{perp}_W(\mathbf{v})}_{\text{in } W^{\perp}}.$$

Remark. While $\text{perp}_W(\mathbf{v})$ is reasonable notation, I can't say it's particularly common, and certainly "$\mathbf{v} - \text{proj}_W(\mathbf{v})$" or "$\text{proj}_{W^{\perp}}(\mathbf{v})$" are less ambiguous.

Example 5.2.7

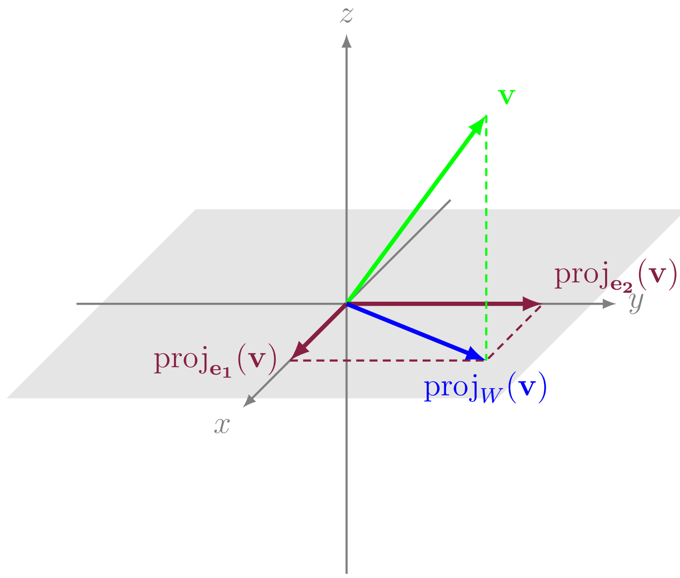
Suppose W is the xy-plane in $\mathbb{R}^3$ and let $\mathbf{v} = [3, 4, 5]^T$. Find the orthogonal decomposition of $\mathbf{v}$.

$$\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{e_1}}(\mathbf{v}) + \text{proj}_{\mathbf{e_2}}(\mathbf{v}) = 3\mathbf{e_1} + 4\mathbf{e_2} = [3, 4, 0]^T.$$

hence

$$\perp_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(\mathbf{v}) = [0, 0, 5]^T.$$

Visually,



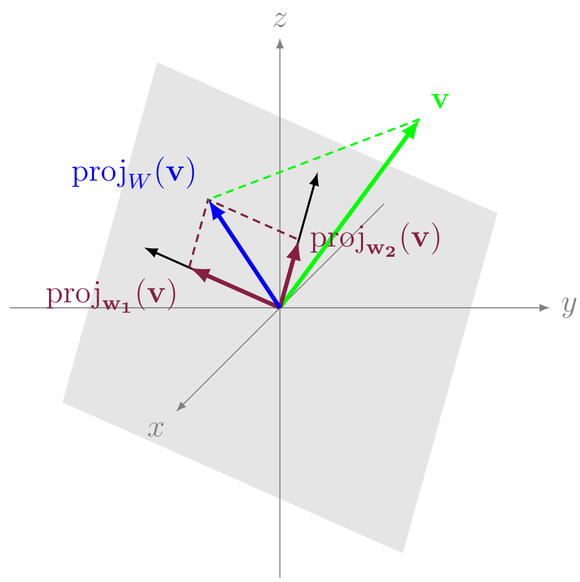
Example 5.2.8

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 $W = \text{Span}(\mathbf{w_1}, \mathbf{w_2})$. Compute $\text{proj}_W(\mathbf{v})$.

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

$$\begin{aligned}
\text{proj}_W(\mathbf{v}) &= \text{proj}_{\mathbf{w_1}}(\mathbf{v}) + \text{proj}_{\mathbf{w_2}}(\mathbf{v}) \\
&= \left(\frac{\mathbf{v} \cdot \mathbf{w_1}}{\mathbf{w_1} \cdot \mathbf{w_1}}\right) \mathbf{w_1} + \left(\frac{\mathbf{v} \cdot \mathbf{w_2}}{\mathbf{w_2} \cdot \mathbf{w_2}}\right) \mathbf{w_2} \\
&= \left(\frac{6 - 8 + 10}{12}\right) \mathbf{w_1} + \left(\frac{-6 + 10}{8}\right) \mathbf{w_2} \\
&= \frac{2}{3}\mathbf{w_1} + \frac{3}{4}\mathbf{w_2} = \frac{1}{6}[-1, -8, 17]^T
\end{aligned}$$

Visually,



Proposition 5.2.9: Projections to find linear combinations

If a subspace W of $\mathbb{R}^n$, $\mathbf{w} \in W$, then $\text{proj}_W(\mathbf{w}) = \mathbf{w}$.
Moreover, if $\{\mathbf{b_1}, \ldots, \mathbf{b_k}\}$ is an orthogonal basis for W, then

$$\mathbf{w} = \text{proj}_{\mathbf{b_1}}(\mathbf{w}) + \cdots + \text{proj}_{\mathbf{b_k}}(\mathbf{w}) = \left(\frac{\mathbf{w} \cdot \mathbf{b_1}}{\mathbf{b_1} \cdot \mathbf{b_1}}\right)\mathbf{b_1} + \cdots + \left(\frac{\mathbf{w} \cdot \mathbf{b_k}}{\mathbf{b_k} \cdot \mathbf{b_k}}\right)\mathbf{b_k}.$$

Example 5.2.10

Let $W = \text{Span}\left(\overset{\mathbf{b_1}}{\begin{bmatrix}1\\0\\1\end{bmatrix}}, \overset{\mathbf{b_2}}{\begin{bmatrix}1\\1\\-1\end{bmatrix}}\right)$ be a plane in $\mathbb{R}^2$ and $\mathbf{w} = \begin{bmatrix}1\\-1\\3\end{bmatrix}$ a vector in W. Write $\mathbf{w}$ as a linear combination of $\mathbf{b_1}$ and $\mathbf{b_2}$.

INCOMPLETE

$$\mathbf{w} = 2\mathbf{u_1} - \mathbf{u_2}.$$

Example 5.2.11

Show that the equation Proposition 5.2.9 doesn't work if one has a non-orthogonal basis.

$$W = \text{Span}\left(\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}\right) \qquad \text{and} \qquad \mathbf{w} = \begin{bmatrix}2\\2\\4\end{bmatrix}.$$

Proposition 5.2.12

$\text{perp}_W(\mathbf{v})$ is orthogonal to W

Proof. Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ and let v be any vector in $\mathbb{R}^n$. We have to show that $\perp_W(\mathbf{v})$ is perpendicular to *every* vector in W, so let $\mathbf{w}$ be an arbitrary vector in W. By a previous theorem, we can write

$$\mathbf{w} = \text{proj}_{\mathbf{w}_1}(\mathbf{w}) + \cdots + \text{proj}_{\mathbf{w}_k}(\mathbf{w})$$

so

$$\text{perp}_W(\mathbf{v}) \cdot \mathbf{w} = \big(\mathbf{v} - \text{proj}_{\mathbf{w}_1}(\mathbf{v}) - \cdots - \text{proj}_{\mathbf{w}_k}(\mathbf{v})\big) \cdot \big(\text{proj}_{\mathbf{w}_1}(\mathbf{w}) + \cdots + \text{proj}_{\mathbf{w}_k}(\mathbf{w})\big)$$

Since the $\mathbf{w}_i$'s are an orthogonal set, then $\text{proj}_{\mathbf{w}_i}(\mathbf{v}) \cdot \text{proj}_{\mathbf{w}_j}(\mathbf{w}) = 0$ whenever $i \neq j$, so the only remaining terms in the above expansion are of the form

$$\begin{aligned} \mathbf{v} \cdot \text{proj}_{\mathbf{w}_i}(\mathbf{w}) &= \mathbf{v} \cdot \left(\frac{\mathbf{w} \cdot \mathbf{w}_i}{\mathbf{w}_i \cdot \mathbf{w}_i}\right) \mathbf{w}_i = \frac{(\mathbf{w} \cdot \mathbf{w}_i)(\mathbf{v} \cdot \mathbf{w}_i)}{\mathbf{w}_i \cdot \mathbf{w}_i} \\ \text{and} \quad \text{proj}_{\mathbf{w}_i}(\mathbf{v}) \cdot \text{proj}_{\mathbf{w}_i}(\mathbf{w}) &= \left(\frac{\mathbf{v} \cdot \mathbf{w}_i}{\mathbf{w}_i \cdot \mathbf{w}_i}\right) \mathbf{w}_i \cdot \left(\frac{\mathbf{w} \cdot \mathbf{w}_i}{\mathbf{w}_i \cdot \mathbf{w}_i}\right) \mathbf{w}_i = \frac{(\mathbf{w} \cdot \mathbf{w}_i)(\mathbf{v} \cdot \mathbf{w}_i)}{\mathbf{w}_i \cdot \mathbf{w}_i} \end{aligned}$$

and these terms all cancel. □

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? Maybe finally answer Question 5.1.5?

Observation 5.3.1.1

Suppose $\{\mathbf{b_1}, \mathbf{b_2}\}$ is a basis for W, then the vector

$$\mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2}) = \mathbf{b_2} - \left(\frac{\mathbf{b_1} \cdot \mathbf{b_2}}{\mathbf{b_1} \cdot \mathbf{b_1}}\right) \mathbf{b_1}$$

is

1. a linear combination $\{\mathbf{b_1}, \mathbf{b_2}\}$, and
2. is perpendicular to $\mathbf{b_1}$

and therefore

$$\left\{\mathbf{b_1},\, \mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2})\right\}$$

is an *orthogonal* basis for W...

Exercise 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 $W = \text{Span}(\mathbf{b_1}, \mathbf{b_2})$. Verify that

$$\left\{\mathbf{b_1},\, \mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2})\right\}$$

is an orthogonal basis for W.

First we compute:

$$\mathbf{b_2} - \text{proj}_{\mathbf{b_1}}(\mathbf{b_2}) = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 21/25 \\ 28/25 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/25 \\ -3/25 \\ 2 \end{bmatrix}$$

1. INCOMPLETE - check linear independence. Suppose
$$x \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} + y \begin{bmatrix} 4/25 \\ -3/25 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$
2. INCOMPLETE - check span independence. Let
$$\mathbf{v} = a \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

be an arbitrary vector in W. We attempt to solve

$$x\begin{bmatrix}3\\4\\0\end{bmatrix}+y\begin{bmatrix}4/25\\-3/25\\2\end{bmatrix}=\mathbf{v}.$$

3. We check the dot product

$$\begin{bmatrix}3\\4\\0\end{bmatrix}\cdot y\begin{bmatrix}4/25\\-3/25\\2\end{bmatrix}=\frac{12}{25}-\frac{12}{25}+0=0$$

and therefore the vectors are orthogonal.

It follows that

$$\left\{\mathbf{b_1},\,\mathbf{b_2}-\operatorname{proj}_{\mathbf{b_1}}(\mathbf{b_2})\right\}$$

is a basis for W, as desired.

As it turns out, the above observation/exercise can be extended into any dimension, and this iterative process is known as the *Gram-Schmidt orthogonalization*.

Theorem 5.3.2: Gram–Schmidt orthogonalization

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 $W_1=\operatorname{Span}(\mathbf{x_1})$.
2. For each $i=2,...,k$, set

$$\mathbf{x_i}=\mathbf{b_i}-\operatorname{proj}_{W_{i-1}}(\mathbf{b_i})\qquad\text{and}\qquad W_i=\operatorname{Span}(\mathbf{x_1},\ldots,\mathbf{x_i}).$$

For each i, $\{\mathbf{x_1},\ldots,\mathbf{x_i}\}$ is an orthogonal basis for W_i and the process terminates after, at most, k steps, where $W_k=W$.

Remark. The basis produced by the Gram–Schmidt orthogonalization process is not unique, as every step required a choice of basis vector $\mathbf{b_i}$. In $\mathbb{R}^2$ for example, running the procedure on the basis

$$\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix}\right\}$$

will produce either the standard basis or the basis

$$\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1/2\\-1/2\end{bmatrix}\right\},$$

depending on which vector you used in step 1.

Remark. One can always scale the basis elements to have norm 1, further producing an orthonormal basis. Sometimes Gram–Schmidt orthogonalization is defined by insisting one normalize each vector along the way, in which case the word' "orthogonalization" should be replaced by "orthonormalization."

Example 5.3.3

Find an ortho<u>normal</u> basis for $W = \text{Span}\left(\overset{\mathbf{b_1}}{\begin{bmatrix}1\\2\\2\\0\end{bmatrix}}, \overset{\mathbf{b_2}}{\begin{bmatrix}0\\1\\2\\2\end{bmatrix}}, \overset{\mathbf{b_3}}{\begin{bmatrix}2\\0\\1\\2\end{bmatrix}}\right)$.

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

$$\begin{aligned}\mathbf{x_3} &= [2,0,1,2]^T - \text{proj}_{W_2}([2,0,1,2]^T)\\ &= [2,0,1,2]^T - \text{proj}_{\mathbf{x_1}}([2,0,1,2]^T) - \text{proj}_{\mathbf{x_2}}([2,0,1,2]^T)\\ &= [2,0,1,2]^T - \left[\frac{4}{9}, \frac{8}{9}, \frac{8}{9}, 0\right]^T - \left[-\frac{4}{9}, -\frac{2}{9}, \frac{4}{9}, \frac{4}{3}\right]^T\\ &= \left[2, -\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right]^T.\end{aligned}$$

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

$$\left\{\begin{bmatrix}1/3\\1/3\\2/3\\0\end{bmatrix}, \begin{bmatrix}-2/3\sqrt{5})\\-1/3\sqrt{5}\\2/3\sqrt{5}\\2/\sqrt{5}\end{bmatrix}, \begin{bmatrix}2/\sqrt{5}\\-2/3\sqrt{5}\\-1/3\sqrt{5}\\2/3\sqrt{5}\end{bmatrix}\right\}.$$

Remark. If at any step in the Gram-Schmidt process, you find that

$$\mathbf{x_i} = \mathbf{b_i} - \text{proj}_{W_{i-1}}(\mathbf{b_i}) = \mathbf{0},$$

then you should completely skip $\mathbf{x_i}$ and move on. *The zero vector cannot be a basis vector.* This also tells you that $\mathbf{b_i}$ is a linear combination of $\mathbf{b_1}, \ldots, \mathbf{b_{i-1}}$.

Example 5.3.4

Find an ortho<u>normal</u> basis for $W = \text{Span}\left(\overset{\mathbf{b_1}}{\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}}, \overset{\mathbf{b_2}}{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}}, \overset{\mathbf{b_3}}{\begin{bmatrix} 5 \\ 1 \\ -1 \\ 3 \end{bmatrix}}, \overset{\mathbf{b_4}}{\begin{bmatrix} 2 \\ 0 \\ 1 \\ 2 \end{bmatrix}} \right)$.

INCOMPLETE $\mathbf{x_3} = \mathbf{0}$, so the orthonormal basis will only involve $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_4}$.

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.

Theorem 5.3.5: QR Factorization

Let A be an $m \times n$ matrix with linearly independent columns. Then there exists an $m \times n$ matrix Q and an $n \times n$ matrix R so that

- Q has orthogonal/orthonormal columns,
- R is upper-triangular,
- and $A = QR$.

The proof is constructive. Let $\{\mathbf{A_1}, \dots, \mathbf{A_n}\}$ be the columns of A and let $\{\mathbf{Q_1}, \dots, \mathbf{Q_n}\}$ 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

$$\begin{aligned}
\mathbf{A_1} &= r_{11} \mathbf{Q_1} \\
\mathbf{A_2} &= r_{12} \mathbf{Q_1} + r_{22} \mathbf{Q_2}
\end{aligned}$$

$$\mathbf{A_3} = r_{13}\mathbf{Q_1} + r_{23}\mathbf{Q_2} + r_{33}\mathbf{Q_3}$$
$$\vdots$$

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 $Q^TQ = I_n$, so in fact $R = Q^TA$, saving us some time in computing R.

Example 5.3.6

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

$$R = Q^TA = \begin{bmatrix} 14 & 21 & 14 \\ 0 & 175 & 70 \\ 0 & 0 & 35 \end{bmatrix}.$$

Chapter 6

Vector Spaces

6.3 Change of Basis

Let's start with a real-world scenario: A hiker leaving basecamp and headed up a mountain to a rendezvous point. He realizes partway through his trip that he forgot his compass (oh no!) and isn't sure which way is North, thus can't navigate to the particular rendezvous point he had in mind. He has been keeping track of how many steps he's taken forward/backward and left/right, but he doesn't know his precise location. Using his satellite phone, he calls back to basecamp and relays this information to them. Basecamp detects his precise location, but their coordinates are in terms of miles and North/East/South/West cardinal directions (as well as vertically upward). How can the person at basecamp tell the hiker which way to go and how many steps to take to get to the rendezvous point?

If basecamp and the hiker could just agree on a set of directions and an appropriate notion of distance, then this would be easy. Our goal is to figure out how to translate between their various coordinate systems (bases). This will literally save a fictional hiker's life.

Example 6.3.1

Basecamp is using the standard basis directions (and miles):

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \qquad \mathbf{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

The hiker is using the following relative (**f**orward, **r**ight, **u**p) directions (also in miles):[a]

$$\mathbf{f} = \frac{1}{3000}\begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \mathbf{r} = \frac{1}{3000}\begin{bmatrix}1\\-2\\1\end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

If basecamp says that the rendezvous point is at location $5\mathbf{e}_1 - \mathbf{e}_2 + 5\mathbf{e}_3$, how many steps and in which direction(s) will the hiker need to go to get from basecamp to the rendezvous point?

[a] the scale factor $\frac{1}{3000}$ is because one takes about 3000 steps per mile.

This question really comes down to finding a solution to x, y, z for which

$$x\mathbf{f} + y\mathbf{r} + z\mathbf{u} = 5\mathbf{e}_1 - \mathbf{e}_2 + 5\mathbf{e}_3$$

or, written as a matrix equation,

$$\begin{bmatrix} | & | & | \\ \mathbf{f} & \mathbf{r} & \mathbf{u} \\ | & | & | \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 5 \end{bmatrix}$$

The fundamental theorem of invertible matrices tells us that since the hiker's directions $\mathcal{H} = \{\mathbf{f}, \mathbf{r}, \mathbf{u}\}$ are a basis, then the matrix

$$\begin{bmatrix} | & | & | \\ \mathbf{f} & \mathbf{r} & \mathbf{u} \\ | & | & | \end{bmatrix}$$

is invertible. Hence the system can be solved with the inverse:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} | & | & | \\ \mathbf{f} & \mathbf{r} & \mathbf{u} \\ | & | & | \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 9000 \\ 6000 \\ 0 \end{bmatrix}$$

This means that the hiker has to walk 9000 steps forward and 6000 steps to the right to get from basecamp to the rendezvous point.

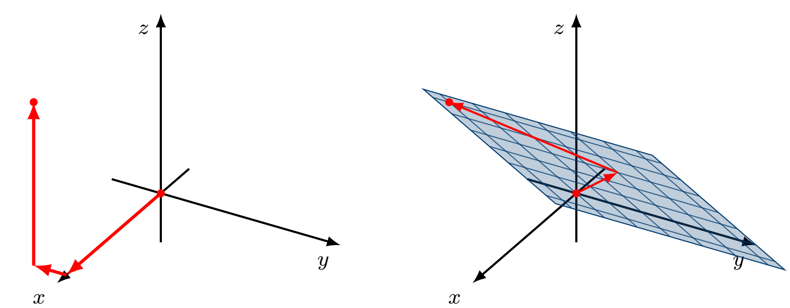


Figure 6.1: On the left, traveling to the rendezvous point via the standard basis. On the right, traveling to the rendezvous point in the $\mathcal{H}$ basis (i.e., on the mountain plane).

Exercise 6.3.2

If the hiker in Example 6.3.1 has already walked 7000 steps forward and 1234 steps to the right, how many steps (and in which directions) should he walk to get to the rendezvous point?

Let's examine Example 6.3.1 while thinking about coordinates:

- The vector $\begin{bmatrix} 5 \\ -1 \\ 5 \end{bmatrix}$ provides the coordinates for the location, written in the standard basis.

- Each of the vectors **f**, **r**, **u** are written in the standard basis.
- The vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9000 \\ 6000 \\ 0 \end{bmatrix}$ provides the coordinates for the location, written in the hiker's coordinate basis $\mathcal{H}$.
- The matrix $\begin{bmatrix} | & | & | \\ \mathbf{f} & \mathbf{r} & \mathbf{u} \\ | & | & | \end{bmatrix}^{-1}$ has the effect of converting the location from the standard basis into the $\mathcal{H}$-basis.

Definition: change of basis matrix

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

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

is called the **change-of-basis matrix** from $\mathcal{B}$ to $\mathcal{C}$. It has the effect of every vector **v** in $\mathbb{R}^n$:

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

Remark. There is no standard notation for the standard basis, so we'll use $\mathcal{E}$.

Proposition 6.3.3: Basic Properties of Change-of-Basis Matrices

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.

Remark. The important takeaway is the following: almost certainly the vectors within two bases $\mathcal{B}$ and $\mathcal{C}$ are communicated to you *written in the standard basis.* So changing bases to/from the standard basis is quite easy:

$$\underset{\mathcal{E}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} | & | & | \\ \mathbf{b_1} & \cdots & \mathbf{b_n} \\ | & | & | \end{bmatrix} \qquad \underset{\mathcal{E}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} | & | & | \\ \mathbf{c_1} & \cdots & \mathbf{c_n} \\ | & | & | \end{bmatrix}$$

Thus the change of basis matrix is given by a simple matrix product.

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left(\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}\right)^{-1}\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right) = \begin{bmatrix} | & | & | \\ \mathbf{c_1} & \cdots & \mathbf{c_n} \\ | & | & | \end{bmatrix}^{-1} \begin{bmatrix} | & | & | \\ \mathbf{b_1} & \cdots & \mathbf{b_n} \\ | & | & | \end{bmatrix}$$

Theorem 6.3.4: A formula for a change-of-basis matrix

Let $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ and $\mathcal{C} = \{\mathbf{c_1}, \ldots, \mathbf{c_n}\}$ be basis for $\mathbb{R}^n$. Then

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} | & | & | \\ \mathbf{c_1} & \cdots & \mathbf{c_n} \\ | & | & | \end{bmatrix}^{-1} \begin{bmatrix} | & | & | \\ \mathbf{b_1} & \cdots & \mathbf{b_n} \\ | & | & | \end{bmatrix}.$$

Example 6.3.5

Let $\mathcal{B} = \left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{b_1}}, \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\mathbf{b_2}} \right\}$ and $\mathcal{C} = \left\{ \underbrace{\begin{bmatrix} 1 \\ -2 \end{bmatrix}}_{\mathbf{c_1}}, \underbrace{\begin{bmatrix} -3 \\ 3 \end{bmatrix}}_{\mathbf{c_2}} \right\}$ be bases for $\mathbb{R}^2$.

1. Let $\mathbf{v} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$. Find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathbf{C}}$, the coordinate representation of $\mathbf{v}$ with respect to the $\mathcal{B}$ and $\mathcal{C}$-bases, respectively.
2. Compute the change of basis matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$.
3. Verify that your change of basis matrix in the previous does what you expect it to do:

$$[\mathbf{v}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{v}]_{\mathcal{B}}$$

1.
2.
3. By the above remark,

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} 1 & -3 \\ -2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{-3} \begin{bmatrix} 3 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ -1 & -\frac{1}{3} \end{bmatrix}$$

This makes sense algebraically: Notice that

$$\underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{b_1}} = -2 \underbrace{\begin{bmatrix} 1 \\ -2 \end{bmatrix}}_{\mathbf{c_1}} - \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\mathbf{c_2}} \qquad \text{and} \qquad \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\mathbf{b_2}} = 0 \underbrace{\begin{bmatrix} 1 \\ -2 \end{bmatrix}}_{\mathbf{c_1}} - \frac{1}{3} \underbrace{\begin{bmatrix} -3 \\ 3 \end{bmatrix}}_{\mathbf{c_2}}$$

and we have

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

Chapter 7

Distances and Approximation

7.2.1 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.

Example 7.2.1

Find the equation of the line passing through the points $P_1(0,1)$ and $P_2(2,5)$.

PICTURE HERE

Writing down the equation of a line as $y = mx + b$, we can plug in the point values, and then we'll be left finding the slope and intercept of this line.

$$\begin{cases} y_1 &= mx_1 + b \\ y_2 &= mx_2 + b \end{cases} \longrightarrow \begin{cases} 1 &= m(0) + b \\ 5 &= m(2) + b \end{cases} \longrightarrow \underbrace{\begin{bmatrix} 1 \\ 5 \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} m \\ b \end{bmatrix}}_{\mathbf{x}}$$

Using standard techniques, one sees that the solution is

$$\begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

i.e. the line $y = 2x + 1$.

Example 7.2.2

Find the equation of the line passing through the points $P_1(0,1)$, $P_2(2,5)$, and $P_3(1,1)$.

PICTURE HERE

Writing down the equation of a line as $y = mx + b$, we can plug in the point values, and then we'll be left finding the slope and intercept of this line.

$$\begin{cases} y_1 &= mx_1 + b \\ y_2 &= mx_2 + b \\ y_3 &= mx_3 + b \end{cases} \longrightarrow \begin{cases} 1 &= m(0) + b \\ 5 &= m(2) + b \\ 1 &= m(1) + b \end{cases} \longrightarrow \underbrace{\begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} m \\ b \end{bmatrix}}_{\mathbf{x}}$$

Using standard techniques, one sees that there is no solution. Oh no!

In practice, we're usually interested in finding the *best-fit line*, which is an approximate solution to a system $A\mathbf{x} = \mathbf{b}$ (and, as the name suggests, is the *best* approximation).

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.

Definition

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 $\overline{z} = a - bi$.

Complex Addition. The sum two complex numbers z_1 and z_2 is done via by adding the real and imaginary parts separately:

$$z_1 + z_2 = (a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + i(b_1 + b_2).$$

Addition is commutative.

Complex Multiplication . The product of two complex numbers follows the usual distributive law:

$$\begin{aligned} z_1 z_2 &= (a_1 + b_1 i)(a_2 + b_2 i) \\ &= a_1 a_2 + a_1 b_2 i + a_2 b_2 i + b_1 b_2 i^2 \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) i. \end{aligned}$$

Multiplication is commutative.

Complex Division. 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 $1/z\overline{z}$:

$$\frac{z_1}{z_2} = \frac{z_1 \overline{z_2}}{z_2 \overline{z_2}} = \frac{1}{z_2 \overline{z_2}} \left(z_1 \overline{z_2} \right) = \frac{1}{a_2^2 + b_2^2} \left((a_1 a_2 + b_1 b_2) + (-a_1 b_2 + a_2 b_1) i \right)$$

Example C.0.1

Let $z_1 = 2 + i$, $z_2 = 3 + 4i$. Compute

1. $z_1 + z_2$

2. $z_1 z_2$
3. $z_2 \overline{z_2}$
4. $\dfrac{z_1}{z_2}$

1. $z_1 + z_2 = 2 + i + 3 + 4i = 5 + 5i$
2. $z_1 z_2 = (2 + i)(3 + 4i) = 6 + 8i + 3i - 4 = 2 + 11i$
3. $z_2 \overline{z_2} = (3 + 4i)(3 - 4i) = 9 - 12i + 12i + 16 = 9 + 16 = 25.$
4. $\dfrac{z_1}{z_2} = \dfrac{2 + i}{3 + 4i} = \dfrac{(2 + i)(3 - 4i)}{(3 + 4i)(3 - 4i)} = \dfrac{10 - 5i}{9 + 16} = \dfrac{2}{5} - \dfrac{1}{5}i$

C.1 Matrices with complex eigenvalues

It is entirely possible that matrices with real number entries have imaginary eigenvalues. Here is a useful fact to keep in mind:

Fact. If z is a complex eigenvalue (with nonzero imaginary part), then so is $\overline{z}$.

Example C.1.1

Let $A = \begin{bmatrix} 3 & -4 \\ 2 & -1 \end{bmatrix}$. Show that $\lambda = 1 \pm 2i$ are eigenvalues for A.

Begin by computing the characteristic polynomial

$$\text{Char}_A(x) = \det(A - xI) = \det \begin{bmatrix} 3 - x & -4 \\ 2 & -1 - x \end{bmatrix} = (3-x)(-1-x)+8 = x^2-2x-3+8 = x^2-2x+5.$$

The roots of this polynomial can be found using the quadratic formula

$$x = \frac{2 \pm \sqrt{4 - 4(1)(5)}}{2} = \frac{2 \pm \sqrt{-16}}{2} = \frac{2 \pm 4i}{2} = 1 \pm 2i.$$

Example C.1.2: Example C.1.1 Continued

Let $A = \begin{bmatrix} 3 & -4 \\ 2 & -1 \end{bmatrix}$, the matrix from Example C.1.1. Find eigenvectors for A.

In the last example, we saw that A had eigenvalues $1 + 2i$ and $1 - 2i$. Now we find eigenvectors by computing the eigenspaces:

$$E_{1+2i} = \text{Null}(A - (1 + 2i)) \qquad \text{and} \qquad E_{1-2i} = \text{Null}(A - (1 - 2i)).$$

First we do some row reduction

$$A-(1+2i)=\begin{bmatrix}3-(1+2i) & -4\\ 2 & -1-(1+2i)\end{bmatrix}=\begin{bmatrix}2-2i & -4\\ 2 & -2-2i\end{bmatrix}$$
$$=\begin{bmatrix}2(1-i) & -4\\ 2 & -2(1+i)\end{bmatrix}$$
$$\xrightarrow{-(1-i)R_2\mapsto R_2}\begin{bmatrix}2(1-i) & -4\\ -2(1-i) & 2(1+i)(1-i)\end{bmatrix}=\begin{bmatrix}2(1-i) & -4\\ -2(1-i) & 4\end{bmatrix}$$
$$\xrightarrow{R_1+R_2\mapsto R_2}\begin{bmatrix}2(1-i) & -4\\ 0 & 0\end{bmatrix}$$

It follows that

$$E_{1+2i}=\operatorname{Null}(A-(1+2i))=\operatorname{Span}\left(\begin{bmatrix}\frac{4}{2(1-i)}\\ 1\end{bmatrix}\right)=\operatorname{Span}\left(\begin{bmatrix}\frac{4(1+i)}{2(1-i)(1+i)}\\ 1\end{bmatrix}\right)=\operatorname{Span}\left(\begin{bmatrix}1+i\\ 1\end{bmatrix}\right).$$

As well,

$$A-(1-2i)=\begin{bmatrix}3-(1-2i) & -4\\ 2 & -1-(1-2i)\end{bmatrix}=\begin{bmatrix}2+2i & -4\\ 2 & -2+2i\end{bmatrix}$$
$$=\begin{bmatrix}2(1+i) & -4\\ 2 & -2(1-i)\end{bmatrix}$$
$$\xrightarrow{-(1+i)R_2\mapsto R_2}\begin{bmatrix}2(1+i) & -4\\ -2(1+i) & 2(1+i)(1-i)\end{bmatrix}=\begin{bmatrix}2(1+i) & -4\\ -2(1+i) & 4\end{bmatrix}$$
$$\xrightarrow{R_1+R_2\mapsto R_2}\begin{bmatrix}2(1+i) & -4\\ 0 & 0\end{bmatrix}$$

It follows that

$$E_{1-2i}=\operatorname{Null}(A-(1-2i))=\operatorname{Span}\left(\begin{bmatrix}\frac{4}{2(1+i)}\\ 1\end{bmatrix}\right)=\operatorname{Span}\left(\begin{bmatrix}\frac{4(1-i)}{2(1+i)(1-i)}\\ 1\end{bmatrix}\right)=\operatorname{Span}\left(\begin{bmatrix}1-i\\ 1\end{bmatrix}\right).$$

Index