MAT 2114 Intro to Linear Algebra

Joe Wells
Virginia Tech

Fall 2023

Contents

Chapter 0

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.

1.1 The Geometry and Algebra of Vectors

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

1.1.1 Definitions and Examples

Definition

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

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

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

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

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

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

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

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

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

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

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

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

Example 1.1.1.1: $\mathbb{R}$ is a vector space.

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

Example 1.1.1.2: 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.

Example 1.1.1.3: 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 $\mathcal{P}^n(x)$, 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: $r(a_n x^n + \cdots + a_0) = (ra_n)x^n + (ra_0)$

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

Example 1.1.1.4: Continues 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 function $f(x) = 0$ is the zero vector in this space, and this space is denoted $C(\mathbb{R})$.

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

1.1.2 Geometric Interpretation of Vector Operations

Now we'll take a geometric interpretation of vectors to help justify the naturality of the operations of vector addition and scalar multiplication. Let $o = (0, 0)$, $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ be some points in the plane. Let $\overrightarrow{op_1}$ be the arrow from o to p_1, and similarly let $\overrightarrow{op_2}$ be the arrow from o to p_2. Furthermore, let $p_3 = p_1 + p_2$ (with addition as described in Example 1.1.1.2). 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.

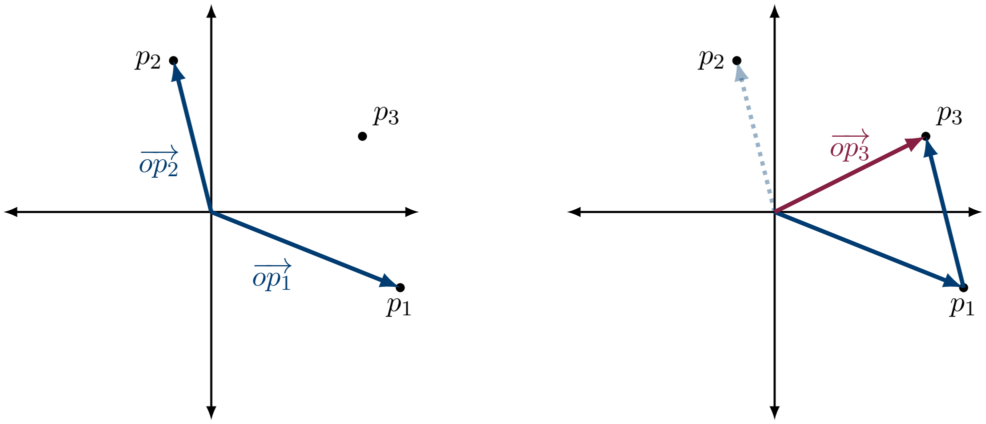


Figure 1.1: The original vectors (left) and "head-to-tail" vector addition (right).

With p_1 as before, consider some real number r. By the scalar multiplication operation described in Example 1.1.1.2, 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)

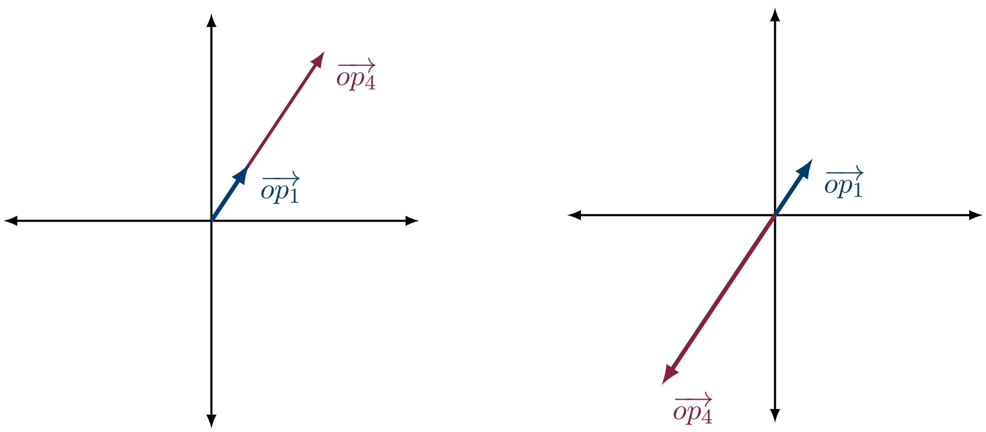


Figure 1.2: The original vector scaled by $r > 0$ (left) and $r < 0$ (right).

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.

Definition

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

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

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

Theorem 1.1.2.1: $\mathbb{R}^n$ 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 1.1.3.1

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

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

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 1.1.3.2

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.

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

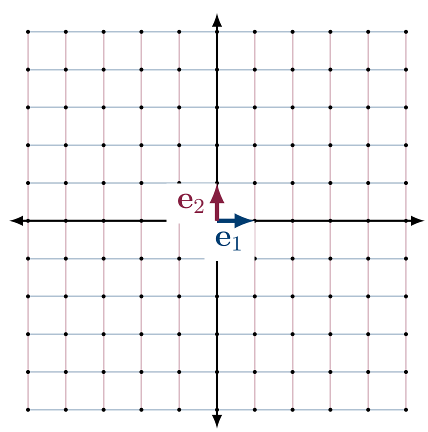
The reader may be wondering precisely *when* a given vector admits a unique linear combination. This is a very important discussion with important implications, and so we will postpone this discussion for a later chapter.

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 $\mathbb{F}^n$. We'll call this a **coordinate grid** (which is nonstandard terminology), and its construction is simple: the lines of the grids should be parallel to the vectors (in standard position) and the intersections of these grid lines correspond to integer linear combinations of vectors.

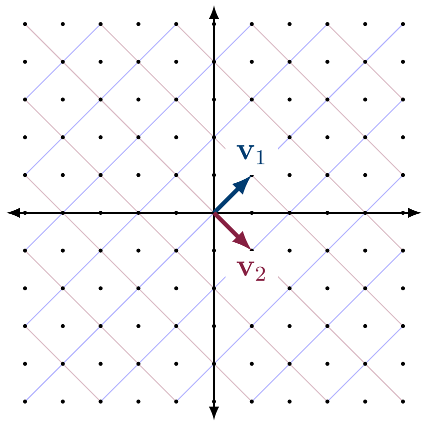
Example 1.1.4.1

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



Example 1.1.4.2

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

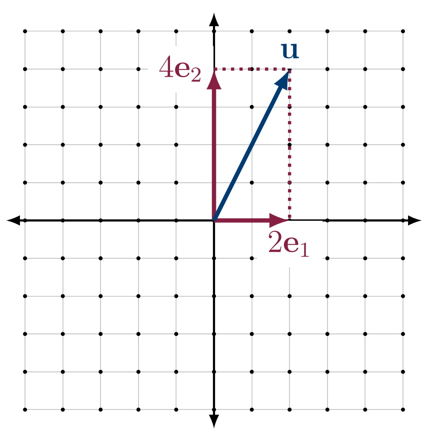


Combined with the geometric intuition about vector addition and scalar multiplication, these coordinate grids provide us with a way to visually identify the linear combination.

Example 1.1.4.3

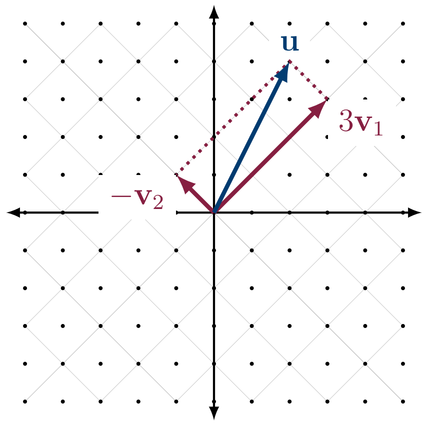
Draw a picture which shows how the vector $\mathbf{u} = [2, 4]^T$ is a linear combination of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

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



Example 1.1.4.4

Show that the vector $\mathbf{u} = [2, 4]^T$ is a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.



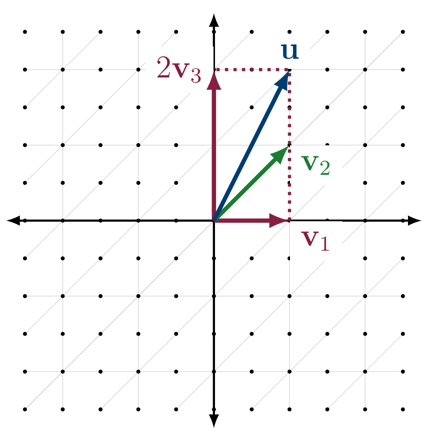
From the above figure it follows that

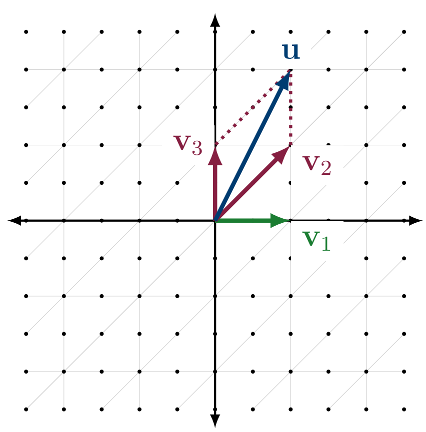
$$\mathbf{u} = 3\mathbf{v}_1 - \mathbf{v}_2.$$

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

Example 1.1.4.5

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

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

Section 1.1 Exercises

1.

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

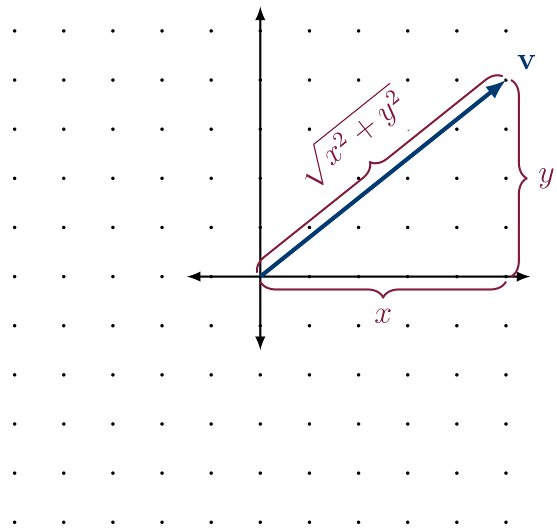
Proof. The proof is entirely straightforward and left as an exercise to the reader. □

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 1.2.0.1

Theorem 1.2.1.1: Properties of Length

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

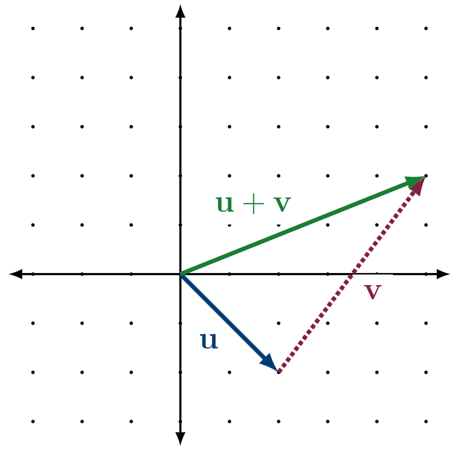
1. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
2. $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.

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

Theorem 1.2.1.2: Triangle Inequality

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

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



Definition: unit vector

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

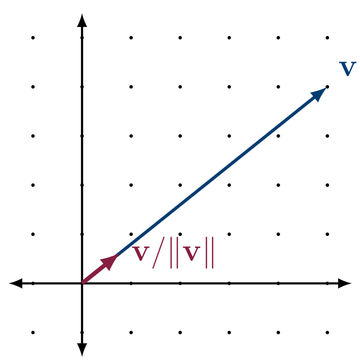
Remark. Every unit vector in $\mathbb{R}^2$ corresponds to a point on the unit circle. Every unit vector in $\mathbb{R}^3$ corresponds to a point on the unit sphere. Generally, every unit vector in $\mathbb{R}^n$ corresponds to a point on the unit $(n-1)$-sphere.

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

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

Definition: normalizing a vector

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



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

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

1.2.2 Distances

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

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

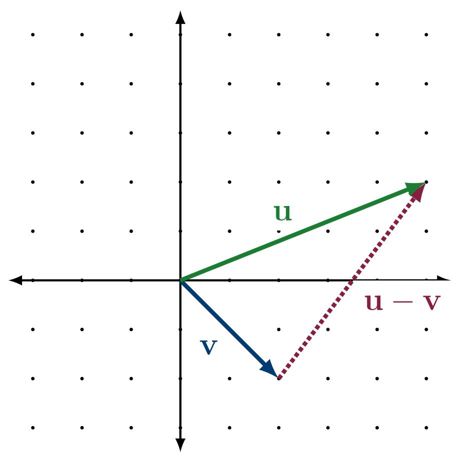
If we identify the point $P(x_1, y_1)$ with the vector $\mathbf{u} = [x_1, y_1]^T$ and the point $Q(x_2, y_2)$ with the vector $\mathbf{v} = [x_2, y_2]^T$, then the right-hand side of the equation is just $\|\mathbf{u} - \mathbf{v}\|$. As such, we can define distances between vectors using the obvious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

Definition: distance

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

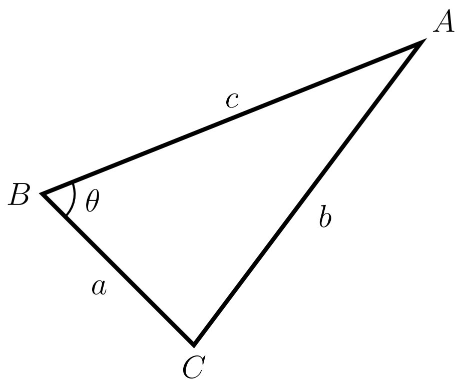
$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

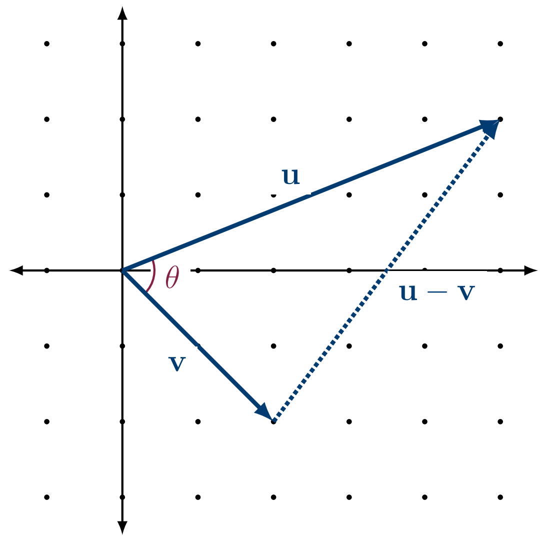
Remark. Visualizing vectors as arrows emanating from the origin, distance, as above, is actually measuring the distance between the heads of the arrows.



1.2.3 Angles

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





Recall that the law of cosines says

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

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

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

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

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

Canceling appropriately and rearranging the equation yields

Definition: angle between vectors

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

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

Example 1.2.3.1

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

From the above, we get that

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

and thus

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

The following follows immediately from the definition.

Corollary 1.2.3.2

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

Section 1.2 Exercises

1.

2.1 Introduction to Linear Systems

In Example 1.1.3.2, we showed that 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. We used ad-hoc methods, but let's reframe this. We are interested in finding any and all values, x, y, z, for which

$$x\mathbf{v}_1 + y\mathbf{v}_2 + z\mathbf{v}_3\mathbf{u}$$
$$\begin{bmatrix} x \\ -x \\ x \end{bmatrix} + \begin{bmatrix} -y \\ y \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} x-y \\ -x+y \\ x+y+z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

By comparing coefficients, the left- and right-hand sides are equal precisely when x, y, z satisfy

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

so if we can figure out how to find these x, y, z-values, then we'll have all linear combinations of $\mathbf{u}$ in terms of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$!

Definition: Linear Equation

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

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

where $a_1, \ldots, a_n$ are real numbers called **coefficients** and b is a real number called the **constant term**. A **solution** of the equation is a vector $[v_1, \ldots, v_n]^T$ so that

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

Example 2.1.0.1

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

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

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

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

This means there are infinitely many possible solutions.

Definition

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

Noticing that

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

we can write the solution set to the previous example as

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

Definition

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

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

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

Example 2.1.0.2

$\sin\left(\frac{\pi}{82364423}\right) x + \sqrt{540.6464}y + z = e^{71}$ is a linear equation, because complicated as they are, $\sin\left(\frac{\pi}{82364423}\right), \sqrt{540.6464}y, e^{71}$ are just real numbers.

Example 2.1.0.3

$x + xy + y = 7$ is *not* a linear equation because of that xy term.

Example 2.1.0.4

$x^2 + 3^y = 8$ is *not* a linear equation because of both the x^2 and 3^y terms.

Definition

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

Example 2.1.0.5

The system

$$\left\{ \begin{array}{rcrcl} 2x & - & y & = & 3 \\ x & + & 3y & = & 5 \end{array} \right.$$

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

Definition

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

Fact. A system of linear equations with real coefficients has either

(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).

Example 2.1.0.6

The system in Example 2.1.0.5 is consistent and the solution is unique.

Example 2.1.0.7

The system

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

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

Example 2.1.0.8

The system

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

has no solutions.

Definition

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

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

Example 2.1.0.9

Consider the system

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

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

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

Definition

Given a system of linear equations

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

the corresponding **augmented matrix** is

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

and the corresponding **coefficient matrix** is

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

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

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

Example 2.1.0.10

The "triangular structure" of the system in Example 2.1.0.9 is also apparent in the corresponding augmented and coefficient matrices:

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

Section 2.1 Exercises

1.

2.2 Direct Methods for Solving Linear Systems

2.2.1 Row Operations

Example 2.2.1.1

In Example 2.1.0.5 we saw that the system

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

was consistent and had the unique solution $[x, y]^T = [2, 1]^T$. Consider now the following three systems of equations. The first systems obtained by merely swapping the equations. The second system is obtained by scaling the second equation. The third system is obtained by replacing the second equation with the sum of the first and second equations.

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

All of these have the (unique) solution $[x, y]^T = [2, 1]^T$ (this is left as an exercise for the reader), and so they are all equivalent.

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

Definition: Elementary Row Operations

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

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

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

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

Definition: row equivalence

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

Example 2.2.1.2

Using the systems in Example 2.2.1.1, show that the corresponding augmented matrices are row equivalent.

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

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

2.2.2 (Reduced) Row Echelon Form

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

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

and thus they correspond to the following row equivalent augmented matrices

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

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

Definition: echelon form

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

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

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

Example 2.2.2.1

The following matrices are in row echelon form.

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

Example 2.2.2.2

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

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

Definition: reduced echelon form

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

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

Example 2.2.2.3

The following matrices are in reduced row echelon form.

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

Example 2.2.2.4

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

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

Theorem 2.2.2.5

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

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

Example 2.2.2.6

Using elementary row operations, find a matrix R that is row equivalent to the following 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).

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{array}\right] \xrightarrow{R_2-3R_1 \mapsto R_2} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{array}\right] \quad (2.1)$$

$$\xrightarrow{R_3-2R_1 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{array}\right] \quad (2.2)$$

$$(2.3)$$

4. Ignoring Row 1, find the next nonzero column in this matrix.
   (Ignoring Row 1, the second column is now the next nonzero column.)
5. Among all of the rows below Row 1 with nonzero entries in this column, choose one and move it to Row 2.

$$\xrightarrow{R_2 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{array}\right] \quad (2.4)$$

6. Use elementary row operations to clear all other nonzero entries in this column (below Row 2).
   (Already done.)
7. Repeat this process until the matrix is in row echelon form.
   (Huzzah, the matrix in Equation 2.5 is in row echelon form!)
8. Now scale every row so that the leading term is a 1. The result will be in reduced row echelon form.

$$\xrightarrow{\frac{1}{5}R_3 \mapsto R_3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad (2.5)$$

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

$$\xrightarrow{R_1+R_2\mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 2 & 7 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.6}$$

$$\xrightarrow{R_1-2R_3\mapsto R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.7}$$

$$\xrightarrow{R_2-3R_3\mapsto R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right] \tag{2.8}$$

Remark. It's worth noting that row operations can be performed in any order, and so the order of operations carried out above is not strictly necessary. Indeed, if you are working through this procedure by hand, you may choose to do operations that retain integers and positive entries for as long as possible to reduce arithmetic mistakes.

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

Definition

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

Theorem 2.2.2.7

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

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

Definition: elimination, 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 essentially the same except 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 2.2.3.1

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

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

We set up the augmented matrix and row-reduce

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

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

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

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

The corresponding system is

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

Solving for the leading variables, we get

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

and hence any solution is of the form

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

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

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

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

2.2.4 Rank and Number of Solutions

Example 2.2.4.1

In Example 2.1.0.8, we stated that the system

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

was inconsistent. What do you observe when you row reduce the augmented matrix?

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

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

Proposition 2.2.4.2

Let $A = [a_{ij}]$ be the coefficient matrix for a system and $\mathbf{b} = [b_1, \ldots, b_m]^T$ the vector of constant terms. If the row echelon form of the augmented matrix $[A|\mathbf{b}]$ contains a row where $a_{i1} = a_{i2} = \cdots = a_{im} = 0$ and $b_m \neq 0$, then the system is inconsistent.

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

Definition: rank

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

Example 2.2.4.3: W

at is the rank of the coefficient matrix in Example 2.1.0.8? is 1, and the rank of the coefficient matrix in Example **??** is 2.

1 and 2, respectively.

Theorem 2.2.4.4: The Rank Theorem

If A is the coefficient matrix of a consistent system of linear equations with n variables, then

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

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

Definition: homogeneous

A system of linear equations, $A\mathbf{x} = \mathbf{b}$ is *homogeneous* if $\mathbf{b} = \mathbf{0}$.

Remark. Homogeneous systems are nice because they ALWAYS have at least one solution (namely the *trivial solution* $\mathbf{x} = \mathbf{0}$).

Theorem 2.2.4.5

If $A\mathbf{x} = \mathbf{0}$ is a homogeneous system of m linear equations and n variables, where $m < n$, then the system has infinitely many solutions.

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

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

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

Example 2.2.4.6

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

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

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

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

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

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

The way this last example differs from Example 2.2.3.1 is that we have exactly as many free variables as we have vectors in the linear combination (instead of also having the extra constant vector added on). This is more ideal because, with the usual vector operations, the collection of all of these solutions is actually a vector space! We will explore this idea a bit further in the next section.

Section 2.2 Exercises

1.

2.3 Spanning Sets and Linear Independence

2.3.1 Span and Spanning Sets

2.3.2 Linear (In)dependence

2.3.3 Using Matrices to Determine Linear (In)dependence

Section 2.3 Exercises

1. Give an example of three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ for which $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linearly dependent set, but $\mathbf{u}$ cannot be written as a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

3.1 Matrix Operations

3.1.1 Matrix Basics

3.1.2 Matrix Operations

3.1.3 Matrix Powers

3.1.4 Transpose

Section 3.1 Exercises

1. Let $A, \ldots, F$ be matrices with the given sizes

matrix	A	B	C	D	E	F
size	2×2	2×3	3×2	2×2	1×2	2×1

   Determine the sizes of each of the following matrices, if possible. If it is not possible, explain why.

   (a) $3D - 2A$
   (b) $D + BC$
   (c) BB^T
   (d) $B^T C^T - (CB)^T$
   (e) $DA - AD$
   (f) $(I_2 - D)^2$

2. Compute $(A - I_2)^2$ where $A = \begin{bmatrix} 1 & -7 \\ 0 & 1 \end{bmatrix}$.

3. Give an example of a nonzero 2×2 matrix A for which $A^2 = O_{2\times 2}$. *Show some work to verify it has this property.*

4. Give an example of a nonzero 3×3 matrix A for which $A^3 = O_{3\times 3}$. *Show some work to verify it has this property.*

5. Let $A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$. Give an example of two 2×2 matrices B and C such that $AB = AC$ but $B \neq C$. *Show some work to verify it has this property.*

6. Give an example of two 2×2 matrices A and B for which $AB \neq BA$. *Show some work to verify it has this property.*

7. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

   (a) Compute A^2, A^3, and A^4.
   (b) What is A^n for some positive integer n?

3.2 Matrix Algebra

3.2.1 Properties of Matrix Multiplication

3.3 The Inverse of a Matrix

3.5 Subspaces, Basis, Dimension, and Rank

3.5.1 Subspaces Associated with Matrices

3.5.2 Dimension and Rank

3.5.3 Coordinates

6.3 Change of Basis

Section 6.3 Exercises

1.

3.6 Introduction to Linear Transformations

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

6.6 The Matrix of a Linear Transformation

Appendix C: Complex Numbers

Matrices with complex entries

4.1 Introduction to Eigenvalues and Eigenvectors

4.2 Determinants

4.2.1 Determinant of a 2×2 Matrix

4.2.2 Determinant of a $n \times n$ Matrix

4.2.3 Properties of Determinants

4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices

4.3.1 Relationship to Matrix Operations

4.4 Similarity and Diagonalization

4.4.1 Properties of similarity and similar matrices

4.4.2 Diagonalization

4.4.3 Computational power of diagonal matrices

3.7 Applications

3.7.1 Markov Chains

4.6 Applications and the Perron–Frobenius Theorem

4.6.1 Applications to Differential Equations

5.1 Orthogonality in $\mathbb{R}^n$

5.1.1 Orthonormality

5.2 Orthogonal Complements and Orthogonal Projections

5.2.1 Orthogonal Complement

5.2.2 Orthogonal Projections

5.3 The Gram–Schmidt Process and the QR Factorization

5.3.1 Gram–Schmidt

5.3.2 QR Factorization

7.3 Least Squares

7.3.1 Least Squares Approximation

Section 7.3.1 Exercises

Index