AN ABSTRACT OF THE THESIS OF

Jason E. Siefken for the degree of Honors Baccalaureate of Science of Mathematics presented on June 4, 2008. Title: The LA Book.

Abstract approved:

Robert Burton

The LA Book is a free-to-students Linear Algebra textbook written at the introductory level and suitable for teaching MTH341 at Oregon State University. It functions both as a traditional, printed book (rendered via LaTeX) and as an online textbook (rendered via HTML), viewable via any modern web-browser. It is built with a philosophy of continuous growth and added content and is expected to be of practical use to students and professors.

Key Words: Mathematics, Linear Algebra, Textbook

Corresponding e-mail address: siefkenj@onid.orst.edu

The LA Book

by

Jason Siefken

A PROJECT

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Mathematics

Presented June 4, 2008
Commencement June 2008

Honors Baccalaureate of Science in Mathematics project of Jason Siefken presented on June 4, 2008.

APPROVED:

Mentor, representing Mathematics

Committee Member, representing Mathematics

Committee Member, representing Mathematics

Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes release of my project to any reader upon request.

Jason E. Siefken, Author

PROJECT OVERVIEW

MOTIVATION

With the advent of the internet, the cost associated with sharing information has shifted from the very real physical cost of printing and shipping to the almost non-existent cost of sending bits of information from one place to another via electrical impulses. Furthermore, with technologies encouraging highly interactive uses such as blogging, social networking, and "Web 2.0" (sites whose content is user-contributed), the difficulty of finding high-quality material and marketing it to a large audience no longer requires an expensive central authority. Occurring in parallel has been a decrease in the cost of local printing and an increase in printing flexibility through print-on-demand technologies.

The question then looms: why are textbook prices so exorbitant? Textbooks are the single largest cost for university students and their families after tuition (cost of living is not included in this breakdown because this cost is always present). A single semester might bring $800 in cost to a single student. The same semester's tuition might be on the order of $1500. This seems like a contradiction to the price optimization promised to us by free markets. Do publishing companies give enough added benefit to each of their books to compensate for not only the higher price of textbooks, but also the lower cost of distribution and production? Any student or faculty member outside of the textbook industry will likely say no.

A revolution in the distribution and development of learning materials to direct and supplement class activities is inevitable and likely at hand. The tools are now available to side-step the textbook industry and offer cheaper (or free) alternative books.

The LA Book is a first step in creating a revolution in textbooks. A revolution has to start from the ground up, and a revolution in learning has to start with the iteration of functional textbooks. To this end, we decided to write *The LA Book*, a free-to-students Linear Algebra textbook. We chose Linear Algebra as a subject because it has a curriculum that is reasonably focused, but at the same time has many choices in emphasis and topic order. Linear Algebra is bridge course from introductory courses to the upper division with a relatively large audience. This provides us with a well-defined pivot point for the insertion of materials that will leverage the way we create the physical context and the scaffolding for a courses built upon a new type of textbook.

METHODOLOGY

The LA Book's plan is to grow, with additional sections, problem sets, and examples to meet any instructor's needs. For this, we call upon the power of the community. A fully-realized version of *The LA Book* will function much like a moderated Wikipedia. That is, students and professors will be able to write sections, problems, and examples and submit them for inclusion to the book. However, unlike Wikipedia, content will be carefully scrutinized by an expert who will verify the mathematical legitimacy of proposed content. Accepted submissions will be added to the book. This will eventually produce a very large textbook, of which subsets may be extracted for use. For example, if an instructor is teaching a Linear Algebra class primarily to math majors, he can choose a subset of the book with more abstract examples that interest math majors. If, on the other hand, an instructor is teaching Linear Algebra to engineers, she can choose a subset of examples that are more concrete and applicable to engineering.

A key component of *The LA Book*'s vision is to function both as a web-based technology and as a classical print-based technology. This means that there should be an easy fall-back behavior from the rich media experience available on the computer to that of static paper. To accomplish this, I have made automatic conversion scripts to change the web-based format to LaTeX code, ready for compilation into a high-quality and printable version.

With the capability to function well as a printable book, *The LA Book* has the advantage of being able to be competitive with conventional books. This will allow it to be used in the classroom immediately and then to introduce the audience of students and professors to the benefits of online textbooks: anywhere availability, instant lookup of definitions without flipping through an index, expandable examples with various levels of detail; animations, full use of color, and many others.

The LA Book in its current form serves as the core of an envisioned book. It has attained a state where it could be used to teach a standard Linear Algebra course. It already has multiple contributing authors and aims not to cover every conceivable Linear Algebra concept so that additional authors have easy and targeted areas to start (Orthogonality, Symmetric Matrices, etc. are sections that beginning authors can tackle right away).

The online version of *The LA Book* is functional as well. It currently lacks the ability for collaborative editing, collapsible examples, and many of the envisioned features, but does feature aesthetically pleasing matrices and equation display as well as bare-bones navigation.

TECHNOLOGY

The text of *The LA Book* has been written in XML (eXtensible Markup Language). XML is a widely used and user-friendly language. It is also the standard format for sharing data over the internet (HTML is a subset of XML). That makes XML an ideal language to store textbook content.

The XML format basically amounts to labeled parentheses. For example, paragraph text is encased in <p> </p> tags. Each section is encased in <section>

</section> tags, and each chapter is encased in <chapter> </chapter> tags. These tags function very similarly to LaTeX's \begin{definition} \end{definition} tags. Consequentially, conversion from XML to LaTeX is easy.

I have implemented a script written in XSLT (A language specifically designed for converting XML documents to other formats) and PHP to automatically convert XML versions of *The LA Book* to LaTeX. Because the ease of conversion between the web-based XML version of *The LA Book* and the printed version is key, some finer details of typesetting a book must be sacrificed. For example, the loss of exact positioning of pictures in the printed version and precise spacing between multi-line formulas is deemed an acceptable trade-off for a dual web and print based textbook.

The LA Book is also converted, on the fly, to HTML (the language a web browser knows how to display) via a different XSLT script. This script uses techniques developed by the *xml-maiden* project (xml-maiden.com) to display mathematical formulae in a web browser while maintaining cross-browser viewability. Using technology from the *xml-maiden* project has the added benefit that future versions of *The LA Book* will be able to utilize advancements in web-based mathematical display ability with no change to the XML content of the book.

CONTENT

The LA Book, in its current form, teaches Linear Algebra from the perspective of systems of linear equations. It spends a great deal of time focusing on linear equations and always tries to tie back operations on matrices to their corresponding operations on systems of linear equations. After a firm basis is built up and students have comfortably transitioned to the world of matrices (always being able to relate matrices back to systems of linear equations), Vector Spaces and Linear Operators are introduced. *The LA Book* then concludes with Eigenvectors and Diagonalization explained from a geometric perspective.

The $\mathcal{LA}$ Book

Jason Siefken
Max Brugger
Robert Burton

Editors

Barbara Siefken
Sharon Sternadel
Jon Wherry

Contents

Chapter 1

Systems of Linear Equations (SLE's)

1.1 Linear Equation

The most basic form of a linear equation is $ax = b$ where x is a variable and a and b are constants. If this equation is real-valued, then it is almost trivial to solve. Presuming $a \neq 0$, dividing both sides by a gives that $x = b/a$.

Example:

> We want to solve the equation $4x = 8$. We can solve this in one step by dividing both sides by 4, showing us that $x = 2$.

This single-variable form seems almost trivial, but that is only because it is quite easy to divide by 4. Later equations will have this form, but we'll see they are slightly more complicated.

1.2 Multi-variable Linear Equations

A linear equation need not be restricted to a single variable. For example, the equation

$$ax + by = c$$

where x and y are variables and a, b, c are constants is a linear equation. This equation represents a straight line in a 2-D graph with slope $-a/b$ and y-intercept c/b.

Notice that solving for y does not give us a simple constant like our previous example. Instead, it gives us the line $y = -(a/b)x+(c/b)$ which has a solution corresponding to every possible value of x.

It is important to notice that this equation is the sum of variables times constants—not one of the variables is squared, or cubed-rooted, etc. This is what makes the equation linear.

1.3 Systems of Linear Equations

With just one 2-D linear equation, solving for one of the variables gives you an infinite number of solutions, however, if you have two equations, we can often find one unique solution. This happens where the two lines intersect (or equivalently, when the two equations have different slopes).

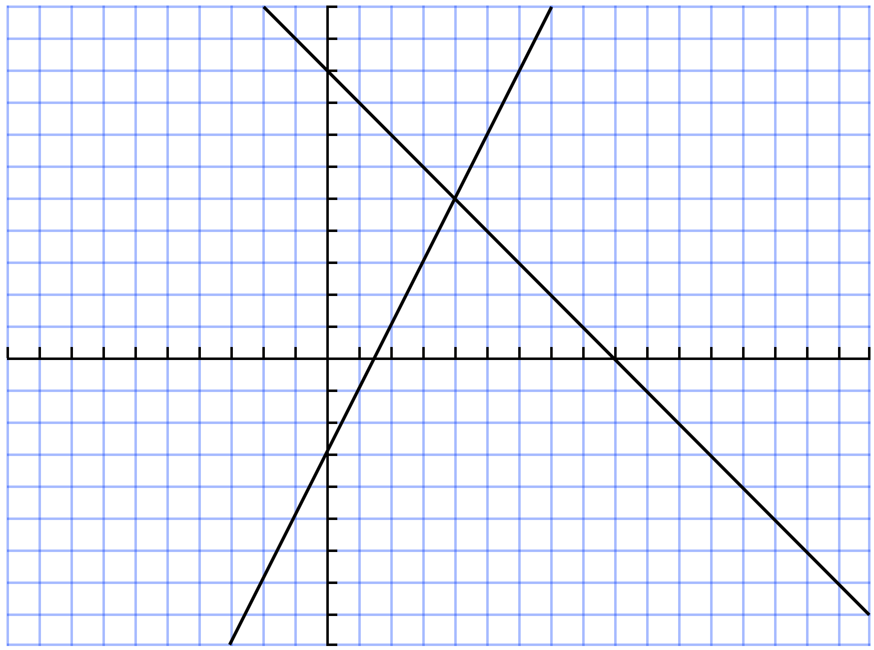
Example:

Consider the two equations

$$2x - y = 3$$

$$x + y = 9.$$

These two equations can be graphed



and we can see they have an intersection point around $(4, 5)$.

When we have a list of equations that are somehow related, we call it a *system*. Because both these equations are linear, we call the pair

$$2x - y = 3$$

$$x + y = 9$$

a *system of linear equations*.

Chapter 2

Solving Systems Algebraically

2.1 Elementary Operations

We have a great example of a system of linear equations:

$$2x - y = 3$$

$$x + y = 9,$$

but how do we solve it? Simple. Using the rules of algebra, we can obtain an exact solution.

Replace the second equation by the sum of equations 1 and 2. This gives us

$$2x - y = 3$$

$$3x = 12.$$

Now, multiply equation 2 by 1/3.

$$2x - y = 3$$

$$x = 4.$$

Next, subtract twice equation 2 from equation 1.

$$-y = -5$$

$$x = 4.$$

Multiply the first equation by -1

$$y = 5$$

$$x = 4.$$

Next, interchange the order of the equations:

$$x = 4$$

$$y = 5.$$

This shows algebraically that the lines intersect at $x = 4$ and $y = 5$, or the point $(x, y) = (4, 5)$ satisfies both equations.

Notice that we used a very small number of very elementary operations to solve our system of equations. We can

1. Interchange the order of equations. (Arrange)
2. Multiply an equation by a constant. (Scale)
3. Add a constant multiple of one equation to another. (Add) (Notice that when we added two equations, we were just adding 1 times an equation to the other.)

2.2 Matrices

Solving our systems of linear equations takes a lot of pencil graphite. However, we seem to be doing some things unnecessarily—our x's, y's, and $=$'s aren't changing and yet we're rewriting them! Let's try re-doing the first example but only writing the numbers that can change. This means

$$2x - y = 3$$

$$x + y = 9$$

will be written as

$$\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix}.$$

Now, as before, let's add the first equation to the second. This is the same thing as adding the first row to the second, which gives us:

$$\begin{bmatrix} 2 & -1 & 3 \\ 3 & 0 & 12 \end{bmatrix}$$

Notice that instead of writing $\begin{bmatrix} 2 & -1 & 3 \\ 3 & & 12 \end{bmatrix}$ we wrote a 0 in the blank's spot so we don't get confused.

Now, multiply equation two by 1/3 (i.e. row two):

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

Subtract twice the second equation from the first (row one $-2*$row two):

$$\begin{bmatrix} 0 & -1 & -5 \\ 1 & 0 & 4 \end{bmatrix}.$$

Next, Multiply the first equation by -1 ($-1*$row one):

$$\begin{bmatrix} 0 & 1 & 5 \\ 1 & 0 & 4 \end{bmatrix}.$$

And finally, interchange the two equations:

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}.$$

If we think back to what symbols we weren't writing (x, y, $+$, and $=$) and put them back in, we get

$$x + 0y = 4 \qquad x = 4$$

$$0x + y = 5 \qquad y = 5,$$

the same result as before.

This notation saved us a lot of work, so let's define how to talk about it.

Definition 2.2.1. *A* matrix *is a two-dimensional grid of numbers, symbols, etc. along with a pair of brackets, parenthesis, square brackets, curly braces, etc.*

Matrices consist of rows (horizontal lines of numbers) and columns (vertical lines of numbers). A matrix's size is written in the form $rows \times columns$. Therefore a 3×5 matrix might look like this:

$$\begin{bmatrix} 3 & 6 & 2 & 8 & 12 \\ 4 & 7 & 16 & 9.4 & 0 \\ .01 & 21 & 5 & 4 & 9 \end{bmatrix}$$

Now, let's consider an arbitrary $m \times n$ matrix with entries a.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{bmatrix}$$

Because there can be many entries in a matrix, we *index* each element. Any element extracted from this matrix looks like $a_{row,column}$. If someone refers to the index of an element, they are referring to the pair of numbers "row,column." A shorthand for writing matrices is to write $A = [a_{ij}]$, where a_{ij} is a generic element from row i and column j.

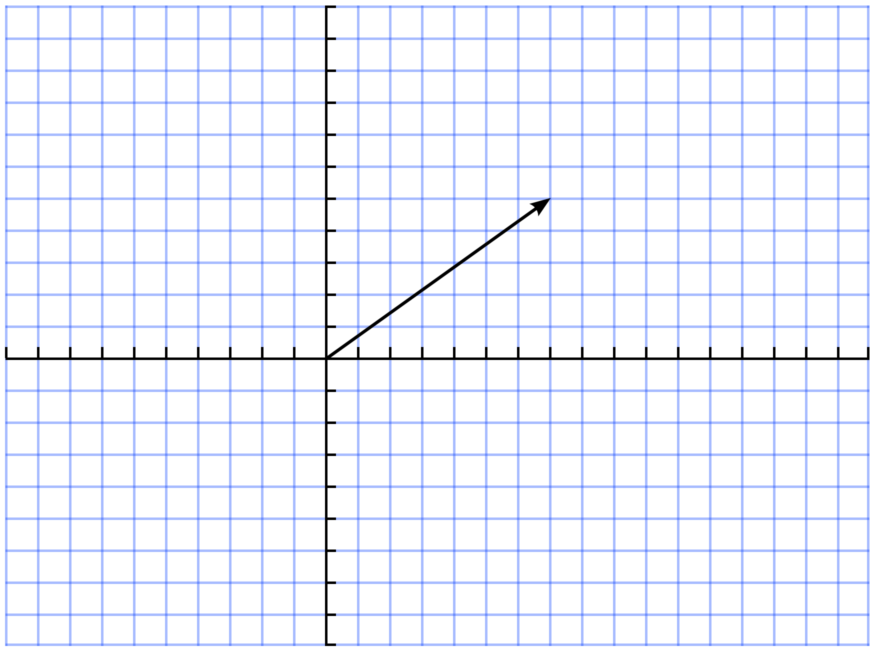
2.3 Vectors

One of the greatest advancements in the history of math was the ability to link algebra and geometry together through a coordinate system. At the core of a coordinate system are, well, coordinates. In this section, we will be looking at a generalized notion of coordinates called *vectors*.

Definition 2.3.1. *A* vector *is a magnitude coupled with a direction.*

You may already be familiar with vectors from a physics or math class. They are not a difficult concept, but they are a very important one. (There would be no video games without them!)

Let's look at a picture of a vector from the origin to the point $(7, 5)$.



We can get the vector's length via the Pythagorean Theorem: Length $= (7^2 + 5^2)^{1/2} \approx 8.602$. We can find the direction with some trigonometry: Angle $= arctan(5/7) \approx 0.775rad$. So our vector from the origin to the point $(7, 5)$ has a magnitude of 8.602 at an elevation of 0.775rad above the x-axis.

However, using the magnitude and direction representation of a vector, produces ugly numbers that tell us nothing additional to "A vector from the origin to $(7, 5)$." In fact, if we assume all vectors come from the origin, then we could represent this vector by just the numbers 7 and 5. This is the way we represent vectors in Linear Algebra.

Definition 2.3.2. *A vector is a list of numbers, i.e. an $n \times 1$ matrix.*

Notice that this definition of vector is equivalent to our earlier one.

In Linear Algebra, a *column vector* is a list of numbers written vertically.

Example:

$$\begin{bmatrix} 7 \\ 5 \end{bmatrix} \text{ or } \begin{bmatrix} 2 \\ 1 \\ 5 \\ 8 \\ 3.2 \end{bmatrix}$$

A *row vector* is a list of numbers written horizontally.

Example:

$$\begin{bmatrix} 7 & 5 \end{bmatrix} \text{ or } \begin{bmatrix} 2 & 3 & 4 & 1 & 9 \end{bmatrix}$$

Each number in a vector is called a *component*.

Traditionally in Linear Algebra, the word *vector* refers to a column vector (this is because mathematicians prefer to multiply on the left). In this book, "vector," unless explicitly stated otherwise, will mean a column vector.

2.4 Vector Operations

There are two important vector operations that make up a linear structure: *scalar multiplication* and *vector addition*.

Adding vectors is done by adding component-wise straight across. (In physics, vector addition is done by laying them head-to-tail and then tracing out the newly-formed vector. Mathematically, this is the same thing as adding component by component straight across.)

Example:

$$\begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 15 \end{bmatrix}.$$

Note that you can only add vectors of the same size or the addition doesn't make sense. (How would you add a line in 2-D to a line in 3-D and get a unique answer?)

Scalar multiplication is our other important operation. A *scalar* is what we normally think of as a "number." For example, the numbers -3.2457 and 0 are scalars. We may write scalars as just the number (like -3.2457) or we may write them in parenthesis (for example, "(-3.2457)"), however, to avoid confusion, we will not write scalars with square brackets []. $\begin{bmatrix} -3.2457 \end{bmatrix}$ refers to a 1×1 matrix, which is slightly different than a scalar.

Multiplying a vector by a scalar is simple—just multiply every component by the scalar.

Example:

$$2 \begin{bmatrix} 5 \\ 6 \\ 15 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 12 \\ 30 \\ 0 \end{bmatrix}$$

Note that we do not use a dot "·" to show multiplication. The dot will be used for a special type of multiplication later on, so use parenthesis if you feel you need to emphasize that things are being multiplied.

Example:

$$(-1/2)(5) \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2.5 \\ -5 \\ 7.5 \\ -10 \end{bmatrix}$$

Traditionally, when multiplying a scalar and a vector, the scalar is always written on the left (and should be written on the left to avoid confusion).

Just like always, in math, it is useful to abstract real objects to symbols. In this way, we give a name as a placeholder for specific vectors. We will use letters with arrows over them to represent vectors. All of our operations still work just the same.

Example:

Suppose $\vec{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \vec{v}_4 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$ then

$$\vec{v}_1 + \vec{v}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} (-1+3) \\ (2+5) \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

just as you thought it would, and

$$2\vec{v}_3 = \begin{bmatrix} 2(3) \\ 2(-4) \end{bmatrix} = \begin{bmatrix} 6 \\ -8 \end{bmatrix}, -10\vec{v}_3 = \begin{bmatrix} -10(3) \\ -10(-4) \end{bmatrix} = \begin{bmatrix} -30 \\ 40 \end{bmatrix}.$$

Scalar multiplication and vector addition are connected by the *distributive law.*

Definition 2.4.1. *If vectors $\vec{v}$ and $\vec{w}$ and scalars c and d obey the* distributive *law, it means that*

$$(c+d)\vec{v} = c\vec{v} + d\vec{v}$$
$$c(\vec{v} + \vec{w}) = c\vec{v} + c\vec{w}$$

When we combine scalar multiplication and vector addition, we arrive at a powerful concept: *linear combinations*.

Definition 2.4.2. *A vector $\vec{w}$ is a* linear combination *of vectors $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$,...,$\vec{v}_n$ if $\vec{w}$ can be written as*

$$\vec{w} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + \cdots + c_n\vec{v}_n$$

where c_1, c_2, c_3,..., c_n are scalars.

Linear combinations are just the sum of scalar multiples of vectors. So, given any vector that isn't the zero vector, we know that there are an infinite number of different linear combinations of that vector we can form.

2.5 Vectors and SLE's

In the previous section we identified a new mathematical object called a vector. Truly, this must be a wondrous thing this vector! For a mathematician, to stumble onto a new mathematical object is akin to someone like Columbus discovering *a new continent*. In the rare moments when we come across something new, it can be dizzying figuring out what to do. I'm *sure* you're just itching to discover this strange new object's properties!

We know how to take any two vectors, multiply each by a different scalar and then add them together to get another vector. It's just a combination of scalar multiplication and then vector addition—two vector properties we've discussed.

This is great, because it looks like we can take any two vectors in a 2-D environment (called 2-space) and construct *any* vector in 2-space as a *linear combination* of the first two vectors. We'll discuss if and when this is true in the next chapter, but you should keep this thought in the back of your head until then.

Definition 2.5.1. *A* linear combination *of the vectors $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$,..., $\vec{v}_n$ is a sum*

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + \cdots + c_n\vec{v}_n$$

where c_1, c_2, c_3,...,c_n are scalars. Note: Any number of c_i could be zero.

Example:

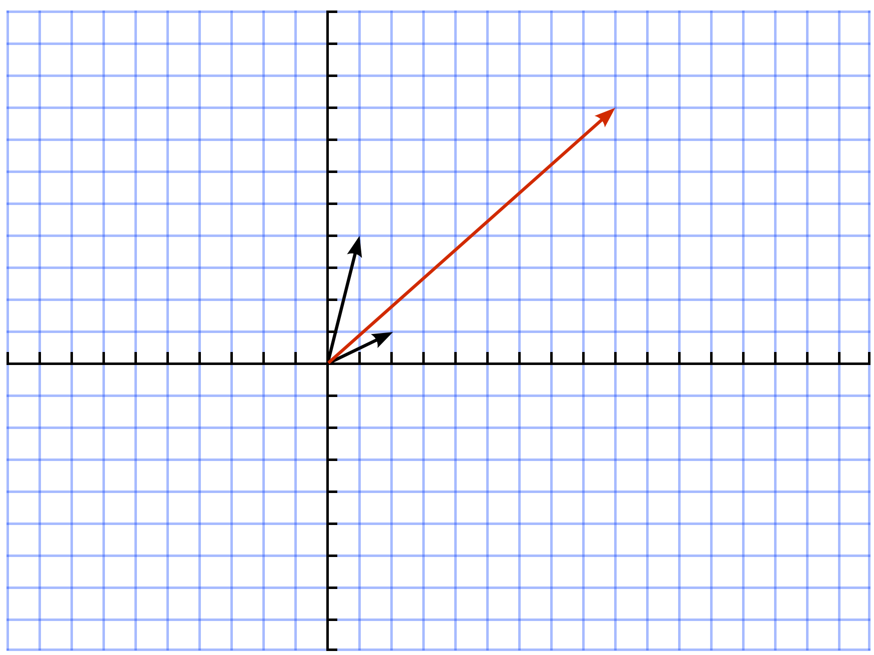
> Let's start by playing a simple game. I'll give you two vectors, $\vec{e}_1$ and $\vec{e}_2$ and a third, $\vec{b}$. All you have to do is tell me what scalars to multiply with $\vec{e}_1$ and $\vec{e}_2$ so that when we add them together we get $\vec{b}$. Sound fun?
>
> This is an easy example, but it should illustrate how these *linear combinations* work.
>
> $$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$$
>
> You probably don't even need to look at the graph, you can just guess that we need to add 9 times $\vec{e}_1$ and 7 times $\vec{e}_2$ to get $\vec{b}$:
>
> $$9\vec{e}_1 + 7\vec{e}_2 = 9\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 7\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 9*1 \\ 7*1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \end{bmatrix} = \vec{b}$$

Now look at this slightly-less-trivial situation.

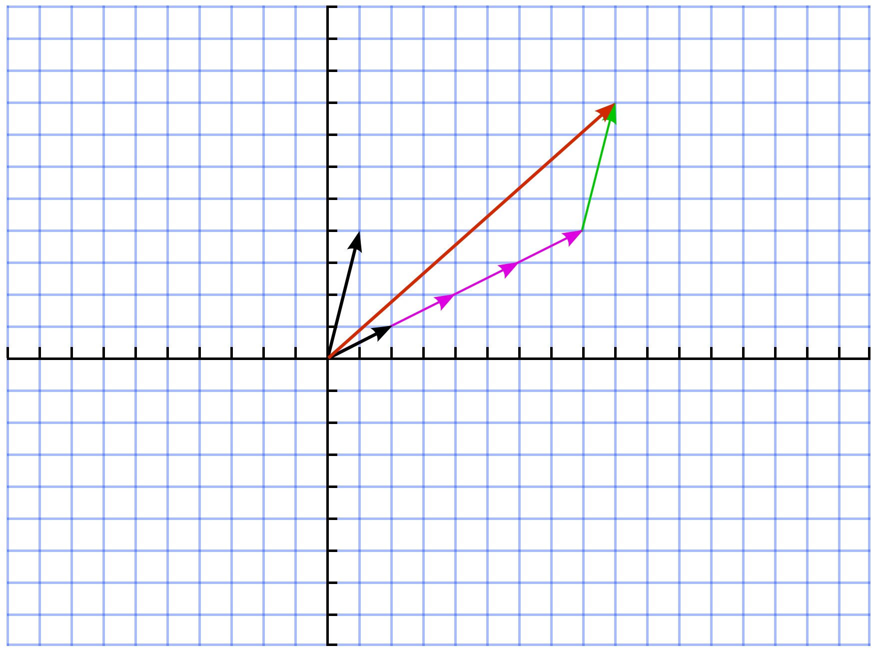
Example:

$$\vec{a}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \vec{a}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \vec{b} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$

Look at the graph of these vectors.



This example is fairly easy because you can probably find the scalars needed by guess and check. Imagine you are sitting down in the graph at the origin and you could only walk in the directions of the two vectors $\vec{a}_1$ and $\vec{a}_2$. Simply count how many steps it takes to get to the vector $\vec{b}$.



Unfortunately, these guess and check processes will only get you so far. Look at the next example, which promises to require scalars that aren't integer-values, but instead decimals or fractions.

Example:

$$\vec{a}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \vec{a}_2 = \begin{bmatrix} -1 \\ 6 \end{bmatrix}, \vec{b} = \begin{bmatrix} 9 \\ -9 \end{bmatrix}$$

You don't want to do it, do you? Even worse than this, consider if you had three vectors in three dimensions, and you had to figure out what three scalars generated a fourth vector, $\vec{b}$? You could attempt some sort of geometrical argument, but it would be difficult and aggravating.

Fortunately, math teachers and math textbooks always want to make things easier for you (what, haven't noticed yet?), so let's look at an easy way to solve any and all of these problems. As a student, Linear Algebra continuously amazed me by its ability to present impossible-looking problems that ultimately have a solution so simple and elegant that even computers and undergraduates can solve them!

Let's look at this example again, but apply some algebra (this is, after all, *Linear Algebra*). We know that we are looking for two scalars. Let's call them x and y. We need x times one vector plus y times the other vector to point at our final vector. In other words, x and y need to satisfy this equation:

$$x\vec{a}_1 + y\vec{a}_2 = x\begin{bmatrix} 3 \\ 4 \end{bmatrix} + y\begin{bmatrix} -1 \\ 6 \end{bmatrix} = \begin{bmatrix} 9 \\ -9 \end{bmatrix}$$

But x and y are just scalars, so let's apply what we know about scalar multiplication:

$$\begin{bmatrix} 3x \\ 4x \end{bmatrix} + \begin{bmatrix} -y \\ 6y \end{bmatrix} = \begin{bmatrix} 9 \\ -9 \end{bmatrix}$$

Adding the two vectors we get,

$$\begin{bmatrix} 3x - y \\ 4x + 6y \end{bmatrix} = \begin{bmatrix} 9 \\ -9 \end{bmatrix},$$

which ought to look awfully familiar. This is, in fact, just a system of linear equations

$$3x - y = 9$$

$$4x + 6y = -9$$

in disguise! And systems of linear equations are so easily solvable that you should weep with joy. I'm not going to weep with joy, because I'm a book. Instead, I'll make this a homework exercise for you to solve. *But that's okay* because all this stuff is easy.

Let's recap: We just discovered an object called a *vector*. And, as it turns out, when we set up equations with these vectors, we see that they can be *translated* into the language of matrices, which *also* describe systems of linear equations. Thus, for every vector equation, we know that there is a corresponding system of linear equations. Neat, huh?

2.6 Elementary Row Operations

Let's think back to solving a system of equations, like

$$2x - y = 3$$

$$x + y = 9.$$

When solving a system like this, there are only three operations we *really* need to use. Those are the same operations for solving systems of equations we remember from elementary algebra:

1. Interchange Order of Equation (*arrange*)
2. Multiply an Equation by a Constant (*scale*)
3. Add a multiple of one equation to another (*add*)

We saw that we not only could use these operations to solve the equation when it looked like algebra, but also when the system looked like a matrix.

As it turns out, operations of these types are called *elementary row operations* and are very important in Linear Algebra, so let us introduce a definition and notation:

Definition 2.6.1. *An* elementary row operation *is one of the following operations applied to a matrix:*

1. *Interchange one row with another (*arrange*)*
2. *Multiply a row by a non-zero constant (*scale*)*
3. *Add a multiple of one row to another (*add*)*

Note that in rule 2, we added the condition that the scalar we multiply by is non-zero. This will become important later on when we use the fact that all elementary row operations are reversible.

Take the matrix

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix}$$

We are going to define a simple notation with arrows and the letter "R" to represent our elementary operations. Let's interchange row 1 and row 2. We write that as "$R_1 \leftrightarrow R_2$," so

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} R_1 \leftrightarrow R_2 \begin{bmatrix} 1 & 2 & 10 \\ 2 & 4 & 6 \\ 9 & 0 & 15 \end{bmatrix}$$

Multiplying row 1 by 1/2 would be written as "$R_1 \to 1/2R_1$."

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} R_1 \to 1/2R_1 \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix}$$

And subtracting twice row 2 from row 3 would be written as "$R_3 \to R_3 - 2R_2$."

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} R_3 \to R_3 - 2R_2 \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 7 & -4 & -5 \end{bmatrix}$$

2.7 Row-Echelon Form

You may have noticed that when we were solving systems of linear equations (and especially when *you* were solving systems of linear equations) using matrix notation, the process each time seemed very similar—alogrithmic even. We (or you) would use the three elementary row operations in strategic ways to achieve certain goals along the way. In this section, we will put a name to those goals and explain why these steps are important enough to name.

Some math teachers are like shepherds, leading you to the tops of math-mountains and showing you the view. Others are like golf caddies, suggesting the club you should use, but leaving the swing, the estimation of the wind speed and the aim all up to you. My Linear Algebra teacher was a caddy and while both styles are fine and dandy, when one gets confused one wants a shepherd, not a caddy. As you might have guessed, on the subject of those enigmatic acronyms REF and RREF, I got confused. Hopefully, you will not suffer the same confusion, and in the places I got confused I will offer special attention, so you aren't stuck with a caddy for a math book. It doesn't matter if the metaphor doesn't work. This is a math book, not a metaphor book.

Definition 2.7.1. *If a matrix is in* pre-row-echelon form *(abbreviated* pre-REF*) it means there are only zeros below the first number in each row and the rows with leading number are sorted so left-most leading entries are hightest in the matrix.*

Example:

$$\begin{bmatrix} 2 & 0 & -1 & 4 \\ 0 & 7 & -2 & 1 \\ 0 & 0 & 3 & 6 \end{bmatrix}, \begin{bmatrix} 2 & -1 & 0 & -4 & 1 & 0 \\ 0 & 3 & 1 & -3 & -1 & 3 \\ 0 & 0 & 2 & 0 & -2 & 1 \\ 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -4 & 0 & 6 & 7/3 \\ 0 & 0 & 2 & 1 \end{bmatrix},$$

and $\begin{bmatrix} 2 & 0 \\ 0 & 16 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ are all in pre-row-echelon form.

Definition 2.7.2. *If a matrix is in* row-echelon form *(abbreviated* REF*) it means that above and below the leading number in each row are entirely zeros and the rows with a leading number are sorted so left-most leading entries are highest in the matrix.*

Example:

$\begin{bmatrix} 2 & 0 & 0 & 4 \\ 0 & 7 & 0 & 1 \\ 0 & 0 & 3 & 6 \end{bmatrix}$, $\begin{bmatrix} -4 & 0 & 0 & 7/3 \\ 0 & 0 & 2 & 1 \end{bmatrix}$, and $\begin{bmatrix} 2 & 0 \\ 0 & 16 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ are all in row-echelon form.

Definition 2.7.3. *A matrix is in* reduced row-echelon form *(abbreviated* RREF*) if it is in row-echelon form and the first non-zero entry in each row is a 1.*

Example:

$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & -6 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 & -7/3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ are all in reduced row-echelon form.

To show where REF and RREF show up in the process of solving a system of linear equations in matrix notation, let's look at an example.

Example:

Solve the system of linear equations:

$$x + y - z = 7$$

$$-2x + y + 0z = -7$$

$$3x + 0y + 2z = 8$$

First, put the system of linear equations in matrix notation and proceed by applying elementary row operations.

$$\begin{bmatrix} 1 & 1 & -1 & 7 \\ -2 & 1 & 0 & -7 \\ 3 & 0 & 2 & 8 \end{bmatrix}$$

$$R_2 \to R_2 + 2R_1, R_3 \to R_3 - 3R_1 \begin{bmatrix} 1 & 1 & -1 & 7 \\ 0 & 3 & -2 & 7 \\ 0 & -3 & 5 & -13 \end{bmatrix}$$

$$R_1 \to 3R_1 - R_2, R_3 \to R_3 + R_2 \begin{bmatrix} 3 & 0 & -1 & 14 \\ 0 & 3 & -2 & 7 \\ 0 & 0 & 3 & -6 \end{bmatrix}.$$

But wait! We're at about the halfway point! All of the rows are *lined up*, in a sense, with the first entry in each row having only zeroes beneath it. (Pay no attention to the coincidence that all of the entries along the diagonal are threes.) Our matrix is already in *pre-row-echelon form*. After this point (in the solving process) we only have to worry about the entries in the upper right hand corner. In this example, it's the -1 and -2 that remain to be zeroed out before the matrix is solved.

Let us continue to solve our matrix equation.

$$\begin{bmatrix} 3 & 0 & -1 & 14 \\ 0 & 3 & -2 & 7 \\ 0 & 0 & 3 & -6 \end{bmatrix} R_1 \to R_1 + 1/3R_3, R_2 \to R_2 + 2/3R_3, R_3 \to 1/3R_3$$

$$\begin{bmatrix} 3 & 0 & 0 & 12 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix} R_1 \to 1/3R_1, R_2 \to 1/3R_2 \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

Congratulations to you if you've paid full attention this whole time, and we have now reached the end of the solving process!

$$x = 4$$
$$y = 1$$
$$z = -2$$

Now that we know what reduced row echelon form means, let us create an algorithm we can follow every time so that there is no clever guesswork. If we are successful, we can program a computer to follow this algorithm and significantly reduce our workload.

The first thing we did when putting this matrix into pre-row-echelon form was to ensure that in the first column, the only non-zero entry was at the very top. Once we accomplished this, we were free to do row operations with all of the rows starting with a zero without ruining the progress we made. After this, we repeated the same procedure in a sub-matrix (one where we removed the first row and the first column of our original matrix and treated that as a new matrix). When we iterated that process till we could go no further, we could use the the first non-zero entry in each row to zero out all the things above it until we were finally in pre-row-echelon form. Once we accomplished that, all we had to do was divide our rows by the first non-zero entry to ensure the first non-zero entry of every row was a one.

Putting this in more precise language, we have:

1. If the first column is not all zeros, rearrange the rows so that the left uppermost entry is non-zero. That is, if our matrix is $A = [a_{i,j}]$, ensure that entry $a_{1,1}$ is non-zero.

2. Use the elementry row operation of adding a multiple of one row to another to zero out the first entry in all rows below the first one. Phrased in symbols, we know the first column in our matrix is given by $\begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{bmatrix}$.
   We want to perform the row operations $R_2 \to R_2 - a_{2,1}/a_{1,1}R_1$, then $R_3 \to R_3 - a_{3,1}/a_{1,1}R_1$, $\cdots$, and finally $R_n \to R_n - a_{n,1}/a_{1,1}R_1$. Now, the only non-zero entry in the first column should be $a_{1,1}$.

3. If our matrix has more than one column and more than one row, consider the submatrix obtained by ignoring the left-most column and the top-most row and go back to step 1. For example, if our original matrix was

$$A = \begin{bmatrix} 1 & 2 & 3 & 5 \\ 0 & 5 & 6 & 7 \\ 0 & 8 & 9 & 10 \end{bmatrix},$$

Our new matrix would be the purple sub-matrix, rather

$$A' = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}.$$

4. Now, in our original matrix, the first non-zero entry of each row should have zeros below it. Use row operations analogous to step 2 to zero out all the entries above the first non-zero entry in row 2. Repeat this in row three, four, etc., until the matrix is in pre-row-echelon form. As an example, our matrix should look like

$$A = \begin{bmatrix} 1 & 2 & 3 & 5 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 9 & 10 \end{bmatrix}.$$

With the second column, we want to perform the row operation $R_1 \to R_1 - 2/5R_2$. This will zero out all entries above $a_{2,2}$, so we would move on to row three and perform the row operations $R_1 \to R_1 - 3/45R_3$ and $R_2 \to R_2 - 6/9R_3$. After this, our matrix is in row-echelon form.

$$A = \begin{bmatrix} 1 & 0 & 0 & 23/15 \\ 0 & 5 & 0 & 1/3 \\ 0 & 0 & 9 & 10 \end{bmatrix}.$$

5. After the matrix is in row-echelon form, divide each row by the first non-zero entry in that row. This will make the first non-zero entry in each row a 1 and consequentally put the matrix into reduced-row-echelon form.

$$A = \begin{bmatrix} 1 & 0 & 0 & 23/15 \\ 0 & 1 & 0 & 1/15 \\ 0 & 0 & 1 & 10/9 \end{bmatrix}.$$

Following this algorithm may result in some very messy fractions, but in the end, any matrix is guaranteed to be in reduced-row-echelon form.

2.8 Equivalence Classes

Equivalence classes are a powerful yet simple concept that allows us to classify and group things.

Definition 2.8.1. *An* equivalence relation*, denoted "$\sim$" is a symmetric, reflexive, and transitive comparison operation. That means, if you are given three things, A, B, and C: $A \sim A$; if $A \sim B$ then $B \sim A$; and if $A \sim B$ and $B \sim C$ then $A \sim C$.*

Definition 2.8.2. *Two matrices are said to be* row equivalent *if their reduced row echelon forms are equal.*

But, is *row equivalence* really an equivalence relation? Let's check: If $rref(A) = rref(B)$ then $rref(B) = rref(A)$, so row equivalence is symmetric. Certainly $rref(A) = rref(A)$, so row equivalence is reflexive. And if $rref(A) = rref(B)$ and $rref(B) = rref(C)$, by the transitivity of $=$'s, $rref(A) = rref(C)$, so row equivalence is really an equivalence relation.

This abstract proof is all well and good, but what does row equivalence really mean? We know that in order to get $rref(A)$, we apply a series of elementary row operations to the matrix A. One great property of elementary row operations is that every single one is reversible (That's why we aren't allowed to multiply a row by zero—we would lose all information about that row and the operation wouldn't be reversible). That means, if we knew all the row operations it took to get $rref(A)$ and we reversed them one by one, we could get back A.

Suppose now that $A \sim B$, i.e. $rref(A) = rref(B)$. If we knew all the row operations it took to get $rref(B)$, we could apply their inverse operations in the proper order and recover B from $rref(A)$! What have we just shown? That if $A \sim B$, then we can transform A into B using only elementary row operations.

The fact that $A \sim B$ means that we can obtain B from A with just elementary row operations is quite neat (in fact, we could have used this fact as the basis for our definition of *row-equivalence*). This also gives us a way to classify matrices of equal size. Take for instance 2×2 matrices. The only possible reduced row echelon forms of 2×2 matrices are:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So, given any 2×2 matrix, we can classify it as row equivalent to either A_1, A_2, A_3, or A_4. Because $\sim$ is an equivalence relation, we know that since none of the A_i are equivalent to each other, a matrix will only be equivalent to one of the A_is. With this knowledge, we can *partition* the space of 2×2 matrices into four separate categories (based upon which of the A_i they are equivalent to). None of our partitions will overlap and every single 2×2 matrix belongs to exactly one partition. Partitioning is a simple, but powerful, concept that will aid us in the coming journey.

Definition 2.8.3. *A* partition *of a set of objects X is a way to disjointly divide X into subsets called* partitions *while ensuring that every object in X belongs to exactly one partition. In other words, a partition divides a group of objects into different categories, but makes sure none of the categories have any overlap.*

Example:

> As we just saw, using row equivalence, we can partition the set of all 2×2 matrices into four, non-overlapping categories.

Example:

> We can partition the real numbers into two sets: the set of numbers greater than zero and the set of numbers less-than-or-equal-to zero. Note that no number belongs to both of these sets and every number belongs to one of these sets.
>
> It is important to point out that if we tried to divide the real numbers into two sets: those greater than zero and those less than zero, this would not be a partition because we haven't specified what category zero belongs to. (Of course, we could easily fix this problem by saying that zero gets its own category.)

Partitions and equivalence relations are closely related. Notice that all equivalence relations give rise to partitions and all partitions give rise to equivalence relations. When we pair an equivalence relation and a partition together (i.e. the equivalence relation gives rise to that particular partition), we call the collection of objects grouped together by the partition an *equivalence class*.

Definition 2.8.4. *An* equivalence class *is the collection of all objects that are similar to each other (i.e. a collection of objects that relate to each other by an equivalence relation).*

Example:

> Taking $\sim$ to be row equivalence, the set of all matrices $A \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ constitutes one equivalence class (and as we will see later is actually the class of all invertible 2×2 matrices).

Because an equivalence class is the collection of all things similar to each other, we can choose to describe an equivalence class as the collection of all things A where $A \sim X$. In this case, X is called a *representative* of the partition/class.

Definition 2.8.5. *A* representative *of an equivalence class is a named element of that class. Once a representative has been established, the particular class may be described as all things similar to that representative.*

Example:

> Taking $\sim$ to be row equivalence, we can describe a single equivalence class as the set of all matrices $A \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. In this case, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the representative element.

2.9 Pivots and Free Variables

In chapter 1, we solved systems of linear equations the old-fashioned way, by writing down a lot of algebraic symbols. But now, with matrix notation and

RREF, we are fully capable of solving all the same problems (and even more complicated ones) with a minimum amount of work.

Consider the following matrix in RREF: $\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & -6 \end{bmatrix}$. We know this is short for the system of equations $1x + 0y + 0z = 4$, $0x + 1y + 0z = 1/2$, and $0x + 0y + 1z = -6$, which easily translates to the answer $x = 4$, $y = 1/2$, and $z = -6$,

Let's look at one more matrix in RREF form and see if we can extract the system of linear equations it stands for: $A = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. By the same process as before, we know A represents the following system of linear equations: $1x+0y+2z = 1$, $0x+1y+3z = 0$, and $0x+0y+0z = 0$. Further simplifying, we get the equations $x = 1 - 2z$ and $y = 0 - 3z$. Any value of z gives a valid solution to this system of linear equations, so this system has infinitely many solutions.

Any time a system of linear equations has an infinite number of solutions, we use a *free variable* when writing down the solution. Values that aren't free (because they are the leading ones in some row of the RREF of a matrix) are called *pivots*.

Definition 2.9.1. Free variables *are variables assigned as place-holders for a whole range of values that result in valid solutions to a system of linear equations.*

Definition 2.9.2. *A* pivot *is the first non-zero entry of a row when a matrix is in pre-row echelon form, row echelon form, or reduced row echelon form.*

Note that if our system of equations is consistent (that is, we don have a row $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$, which would be interpreted as the equation $0 = 1$), the number of free variables plus the number of pivots is the same as the number of equations.

Example:

The matrix $A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ has one pivot and three rows (three equations), so the algebraic solution will have two free variables. The algebraic solution is $x = 3 - y - 3z$. In this solution, y and z are both free variables, showing that there are in infinite number of solutions to the system of linear equations represented by A.

Now that we are familiar with pivots and free variables, we can combine these concepts with vector notation to cleanly write the solution to a system of linear equations.

Recall that the system of linear equations

$$x + 2y + 0z + 3w = 4$$

$$x + 2y + 1z + 5w = 3$$

$$2x + 4y + 4z + 14w = 4$$

can be written in $A\vec{x} = \vec{b}$ form like so:

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 1 & 2 & 1 & 5 \\ 2 & 4 & 4 & 14 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}$$

In order to use row-reduction to solve this system, we concatenate A and $\vec{b}$ together into the following matrix:

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 4 \\ 1 & 2 & 1 & 5 & 3 \\ 2 & 4 & 4 & 14 & 4 \end{bmatrix}.$$

Solving for the RREF, we get the matrix

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 4 \\ 0 & 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

which corresponds to the $A\vec{x} = \vec{b}$ form equation:

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}.$$

This equation is consistent, and there are two pivots, so there are going to be two free variables. Multiplying this equation out, we get

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} y + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} z + \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} w = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}.$$

Solving for our pivots gives us the equation

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} z = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} y - \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} w.$$

Notice that the bottom row in this equation is just zeros, so we can further simplify our notation by leaving it off. Also, because $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x$ is the same as $\begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}$, we can make a further substitution, which gives us the simplest way to write the solution to this system of linear equations:

$$\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \end{bmatrix} y - \begin{bmatrix} 3 \\ 2 \end{bmatrix} w.$$

Notice that this represents the same information as the two equations $x = 4 - 2y - 3w$ and $z = -1 - 0y - 2w$ in a convenient vector form.

Chapter 3

Matrix Operations

3.1 Transpose

Before we start matrix operations, let us introduce the transpose operation. This operation swaps rows and columns, i.e. reflects a matrix along the diagonal. It is represented by a superscript ^{T}.

Definition 3.1.1. *The* transpose *of a matrix A with entries $a_{i,j}$ is the matrix A^T with entries $a_{j,i}$.*

Example:

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

If the matrix is a vector, the transpose simply means changing the vector from a column vector to a row vector or vica versa.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

In a square matrix, it can be hard to see what the transpose does. Watch how the diagonal stays the same while all other numbers are reflected:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

Note that the second matrix was obtained by writing each column of the first as a row of the second.

3.2 Vector Multiplication

In the previous chapter, we defined how to add vectors, which corresponds geometrically to laying a vector head-to-tail and tracing out a new vector. We also defined how to multiply a vector by a scalar. This corresponds to scaling a vector. We refrained from defining what it means to multiply two vectors, but we can go no further until we do.

Vector multiplication is defined as the dot product.

The setup:

For now, we will only be able to multiply a row vector by a column vector (in that order) and the result will be a 1×1 matrix.

The process:

To multiply two vectors, multiply their "corresponding" entries and then add all of those products together and stick it in a 1×1 matrix.

Example:

$$\begin{bmatrix} 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} .5 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 * .5 + 3 * 4 + -1 * 6 \end{bmatrix} = \begin{bmatrix} 7 \end{bmatrix}$$

One might think that it is somewhat restrictive to require all vector multiplication to be between a row vector and a column vector. What if we wanted to multiply two column vectors? Fortunately we have the transpose operation:

$$\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}^T \begin{bmatrix} .5 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} .5 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \end{bmatrix}$$

3.3 Matrix Multiplication

Matrix multiplication is closely tied to vector multiplication. Vector multiplication is a row vector times a column vector. However, any given matrix is just a bunch of row vectors, and/or a bunch of column vectors, depending on your viewpoint.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} \color{red}{1} & \color{red}{2} & \color{red}{3} \\ \color{green}{4} & \color{green}{5} & \color{green}{6} \\ \color{blue}{7} & \color{blue}{8} & \color{blue}{9} \end{bmatrix} = \begin{bmatrix} \color{red}{1} & \color{green}{2} & \color{blue}{3} \\ \color{red}{4} & \color{green}{5} & \color{blue}{6} \\ \color{red}{7} & \color{green}{8} & \color{blue}{9} \end{bmatrix}$$

The multiplication of two matrices is a matrix of vector products. Think of matrix multiplication as dividing up a matrix into row and column vectors and then multiplying just like vectors, sticking everything back in order. Let's multiply a 3×3 matrix with a 3×1 column vector:

$$\begin{bmatrix} \color{red}{1} & \color{red}{2} & \color{red}{3} \\ \color{green}{4} & \color{green}{5} & \color{green}{6} \\ \color{blue}{7} & \color{blue}{8} & \color{blue}{9} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \color{red}{-1+2+3*2} \\ \color{green}{-4+5+6*2} \\ \color{blue}{-7+8+9*2} \end{bmatrix} = \begin{bmatrix} \color{red}{7} \\ \color{green}{13} \\ \color{blue}{19} \end{bmatrix}$$

(Interesting how we multiplied a square matrix by a vector and were left with a vector of the same size.) Let's now do a slightly more complex.

Example:

$$\begin{bmatrix} \color{red}{1} & \color{red}{2} & \color{red}{3} \\ \color{green}{4} & \color{green}{5} & \color{green}{6} \\ \color{blue}{7} & \color{blue}{8} & \color{blue}{9} \end{bmatrix} \begin{bmatrix} \color{magenta}{1} & \color{orange}{0} & \color{olive}{3} \\ \color{magenta}{-2} & \color{orange}{0} & \color{olive}{0} \\ \color{magenta}{2} & \color{orange}{1} & \color{olive}{-2} \end{bmatrix} =?$$

Notice that the second matrix is actually 3 column vectors right next to each other. To multiply the two matrices, we simply multiply a matrix and a column vector three times and stick all the numbers right on top of each other. Let's deal with this one column at a time:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ -2 & 0 & 0 \\ 2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 1(1) - 2(2) + 2(3) & \cdot & \cdot \\ 1(4) - 2(5) + 2(6) & \cdot & \cdot \\ 1(7) - 2(8) + 2(9) & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} 3 & \cdot & \cdot \\ 6 & \cdot & \cdot \\ 9 & \cdot & \cdot \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ -2 & 0 & 0 \\ 2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} \cdot & 0(1) + 0(2) + 1(3) & \cdot \\ \cdot & 0(4) + 0(5) + 1(6) & \cdot \\ \cdot & 0(7) + 0(8) + 1(9) & \cdot \end{bmatrix} = \begin{bmatrix} \cdot & 3 & \cdot \\ \cdot & 6 & \cdot \\ \cdot & 9 & \cdot \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ -2 & 0 & 0 \\ 2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & 3(1) + 0(2) - 2(3) \\ \cdot & \cdot & 3(4) + 0(5) - 2(6) \\ \cdot & \cdot & 3(7) + 0(8) - 2(9) \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & -3 \\ \cdot & \cdot & 0 \\ \cdot & \cdot & 3 \end{bmatrix}$$

Our last step is to smush the columns together to get our final product:

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

Who would have thought math could be so colorful?

Since matrix multiplication is a whole bunch of vector multiplications, what happens when the size of the row vector of the first matrix and the size of the column vector of the second matrix don't match? You get an error. An easy way to check this is to make sure the "inside" dimensions match up on a matrix. For example, suppose you have a 3×4 and a 4×8 matrix. You can multiply these because when you line the dimensions up the inside 4's match: $3 \times 4\ 4 \times 8$. However, if you had a 4×8 and a 3×4 matrix, the inside dimensions do not match: $4 \times 8\ 3 \times 4$ so you cannot multiply these.

Just for fun, let's see what happens when we try to multiply two matrices whose dimensions do not match.

Example:

Let's multiply the following 3×4 and 2×3. We'll only try to calculate the first column, as we'll have plenty enough trouble doing even that.

$$\begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 3 & -2 & -1 \\ -1 & 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 & -4 \\ 3 & 1 & -3 \end{bmatrix} = \begin{bmatrix} 1*2+2*3+0*(?)+5*(?) & \cdot & \cdot \\ 0*2+3*3+(-2)*(?)+(-1)*(?) & \cdot & \cdot \\ (-1)*2+1*3+0*(?)+4*(-3) & \cdot & \cdot \end{bmatrix}$$

The question marks are indications that the rows and columns are not "lining up" per se, so it is impossible to figure out what should be in each entry of the product. This is why the rows of the first matrix must be the same length as the columns of the second matrix for you to be able to multiply them together. But then again you already knew that, didn't you?

3.4 Square Matrices

If two matrices, A and B are square and both have the same dimension, multiplication is defined in either order, AB or BA. However, matrix multiplication is not commutative. That is, in many cases $AB \neq BA$. Let's examine a couple of cases:

Example:

$$\begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 14 \\ 0 & 3 \end{bmatrix}$$
$$\neq$$
$$\begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 0 & 3 \end{bmatrix}$$

Example:

$$\begin{bmatrix} 4 & 1 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 12 \\ 30 & 6 \end{bmatrix}$$
$$\neq$$

$$\begin{bmatrix} 0 & 3 \\ 5 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 18 \\ 20 & 5 \end{bmatrix}$$

Even though matrix multiplication is not commutative, it is associative. That is, in a string of matrix multiplications, parenthesis does not matter. Take matrices A, B, and C.

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

We should also consider what the transpose does to matrix multiplicaiton. The following identity is useful to know:

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

Multiplying out some matrices you can quickly see that this is true.

3.5 Special Matrices I and 0

When working with square matrices, there are two very special matrices. The first is the *zero-matrix*. As its name suggests, this is a matrix written entirely of zeros.

Example:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad [0]$$

are all zero-matrices.

The zero matrix has the special property that when multiplied by another matrix (on either side), the result is the zero matrix.

Example:

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

The zero-matrix is often written with an abbreviation of just “0.” This allows us to write equations like,

$$0A = A0 = 0.$$

Notice how, in the case of square matrices, it doesn't matter whether the “0” in the expression “$0A$” is the zero-matrix or simply the scalar 0.

Identity Matrix

The other very special matrix is the identity matrix. Denoted symbolically as a capital “I,” the identity is so named because a square matrix times the identity equals itself.

Written out, the identity matrix is a matrix with 1's along the diagonal and zeros everywhere else:

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

are all identity matrices.

Let's multiply some matrices to figure out why the identity matrix works.

Example:

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

In the world of matrix multiplication, I behaves just like 1 does in the world of scalar multiplication. That means we may pull out as many identity matrices as we want as a tool to solve whatever problem we may be working on. So, if we are multiplying matrices A and B, we can say,

$$AB = AIB = AIIB = AIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIB = IIAIB$$

3.6 Matrices and SLEs

Consider the system

$$2x - y = 3$$

$$x + y = 9.$$

We can represent this with a square matrix and two vectors:

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

If we multiply the left side out, we see

$$\begin{bmatrix} 2x - y \\ x + y \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

and since we equate vectors by comparing straight across, we see that this is the same as our original system.

However, the real beauty of this representation comes when we swap out the numbers for letters: Let

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = A \qquad \begin{bmatrix} x \\ y \end{bmatrix} = \vec{x} \qquad \begin{bmatrix} 3 \\ 9 \end{bmatrix} = \vec{b}.$$

Now, if we rewrite our original equation, it looks like

$$A\vec{x} = \vec{b},$$

which is the form of the simplest linear equation. (If we could divide by A, we could solve it in a snap.)

In this form, A is called the *coefficient matrix*.

3.7 Elementary Matrices

Matrix multiplication and the identity matrix go hand in hand and are pretty cool. We're almost ready to fully dive into a world of matrices, the only problem is elementary row operations. They don't quite fit nicely into our world of vector/matrix addition, scalar multiplication, and matrix multiplication.

Let's examine what elementary row operations really do and see if we can find an alternative method of representing them.

- Interchanging two rows:

Let's think of a specific interchange, say $R_1 \leftrightarrow R_2$ applied to a 3×3 matrix:

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} R_1 \leftrightarrow R_2 \begin{bmatrix} 1 & 2 & 10 \\ 2 & 4 & 6 \\ 9 & 0 & 15 \end{bmatrix}$$

Thinking about left multiplying A by the identity matrix, we know that I's first row, $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, chooses the first row of A and sticks it into the first row of the result. And we know the second row of I, $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, chooses the second row of A and sticks it into the second row of the result.

What would happen if instead, the first row of I were $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$? It would indeed take the second row of A and stick it in the first row of the result. And if the second row of I were $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, it would take the first row of A and stick it in the second row of the answer.

Let's compute:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 10 \\ 2 & 4 & 6 \\ 9 & 0 & 15 \end{bmatrix}$$

By multiplying our matrix by $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ we have swapped the first and second row. But this matrix is really just the row operation, $R_1 \leftrightarrow R_2$ applied to I:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} R_1 \leftrightarrow R_2 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So, left multiplying by an identity matrix with rows interchanged is the same as applying the elementary row operation of exchanging rows. Neat!

- Multiplying a row by a non-zero constant

This operation is really simple to figure out. If you think about the operation $R_3 \to 2R_3$ and the fact that the third row of I is $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, we could guess that if instead the third row were $\begin{bmatrix} 0 & 0 & 2 \end{bmatrix}$, it might do our row operation for us. Let's see:

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} R_3 \to 2R_3 \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 18 & 0 & 30 \end{bmatrix}$$

is our row operation. And, multiplying by our special matrix gives us:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 9 & 0 & 15 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 18 & 0 & 30 \end{bmatrix}$$

It's the same thing! But our matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is just the operation $R_3 \to 2R_3$ applied to I.

- Adding a non-zero multiple of one row to another

A pattern is emerging. Let's try to continue it by finding a slightly modified identity matrix to perform the operation $R_2 \to -R_1 + R_2$.

This is just as easy to do, but it might be confusing at first. We just take the row we want to modify, in this case R_2, and to make our elementary matrix perform the row operation, we change it from the idenity to: $\begin{bmatrix} -1 & 1 & 0 \end{bmatrix}$.

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 10 \\ 18 & 0 & 30 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ -1 & -2 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

It seems that all our elementary row operations can be done by left-multiplying by a modified version of the identity matrix. For this reason, these modified identiy matrices are called *elementary matrices*.

Definition 3.7.1. *An* elementary matrix *is the identity matrix with one row operation applied to it.*

3.8 Inverses

Using elementary matrices, let's re-examine solving a system of linear equations. We'll start with the system:

$$2x - y = 3$$

$$x + y = 9$$

In matrix form that is

$$\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix}.$$

We know the first thing to do when solving is $R_2 \to R_2 + R_1$, or in terms of an elementary matrix, we get

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ 3 & 0 & 12 \end{bmatrix}.$$

Next we do $R_2 \to 1/3R_2$. Keeping our matrices expanded, that operation is carried out by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 0 & 4 \end{bmatrix}.$$

Now we want to do $R_1 \to R_1 - 2R_2$:

$$\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -5 \\ 1 & 0 & 4 \end{bmatrix}.$$

Next we need to do $R_1 \to -R_1$ and after that $R_1 \leftrightarrow R_2$. That gives us

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}.$$

We've solved our system just like before, but the difference this time is that we have kept track of everything we did with elementary matrices. And, because parenthesis don't matter in matrix multiplication (a.k.a. matrix multiplication is associative), we can collapse

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

down, so we get

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}.$$

This is amazing! There is one matrix that we can multiply our equation matrix by in order to get the solution.

Let's examine a different matrix-form representation of this system of linear equations:

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

If we take our product-of-elementary-matrices matrix times the coefficient matrix of this equation, we get,

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

The product equals the identity matrix! What does this mean? Let's look at normal numbers and normal multiplication.

If we have the number 3 and we find a number a such that $a3 = 1$, then we know $a = 1/3$, that is, a is the inverse of 3. It follows that, if we have a matrix $A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$ and we find another matrix, $B = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix}$, and $BA = I$, then B is the *inverse* of A. Symbolically we represent the inverse of A with a superscript $^{-1}$. Therefore $A^{-1} = B$.

Definition 3.8.1. *Given a matrix A, the* inverse matrix*, A^{-1}, is the unique matrix such that $A^{-1}A = AA^{-1} = I$.*

An inverse basically serves the role of division. Remember our equation $A\vec{x} = \vec{b}$? It looks like we can effectively divide by A by multiplying both sides by A^{-1}. Let's try:

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

At this point, it would be tempting to think that every matrix has an inverse, and that certainly would be nice, however, it is important to know that *not every matrix has an inverse.*

Definition 3.8.2. *A matrix A is* invertible *if there exists an A^{-1} such that $AA^{-1} = A^{-1}A = I$.*

When we first talked about inverses we showed that to find a matrix's inverse, you had to find the elementary matrices that reduced the matrix to the identity. Their product, of course, was the inverse. We can formulate this statement into a theorem:

A matrix A is invertible if and only if it is row equivalent to the identity, that is to say, there is a series of elementary matrices that, when multiplied with A, produces the identity.

Almost trivially, this implies that for a matrix to have even a shining hope of being invertible, it *has* to be square, because there are no elementary row operations that can change a matrix's dimension, and the identity is always square.

An interesting way to classify non-invertible matrices is to realize that a non-invertible matrix has to be row equivalent to something else besides the identity. Let's come up with examples of very simple matrices that are not row equivalent to the identity:

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

Here are some examples. See if you can pick out why they don't row reduce to the identity.

Example:

$$\begin{bmatrix} 2 & 3 & -1 \\ 1 & 1 & 1 \\ 1 & 2 & -2 \end{bmatrix} \quad \begin{bmatrix} 1 & 6 \\ -2 & -12 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 3 \\ 2 & 2 & 4 \\ 1 & 2 & 1 \end{bmatrix}$$

You should be beginnning to see the connection between invertible matrices and whether a system of linear equations has a solution. Before we reveal the awesomeness of the super theorem and bask in its glory, we must introduce another player into this mathematical drama.

3.9 Determinants

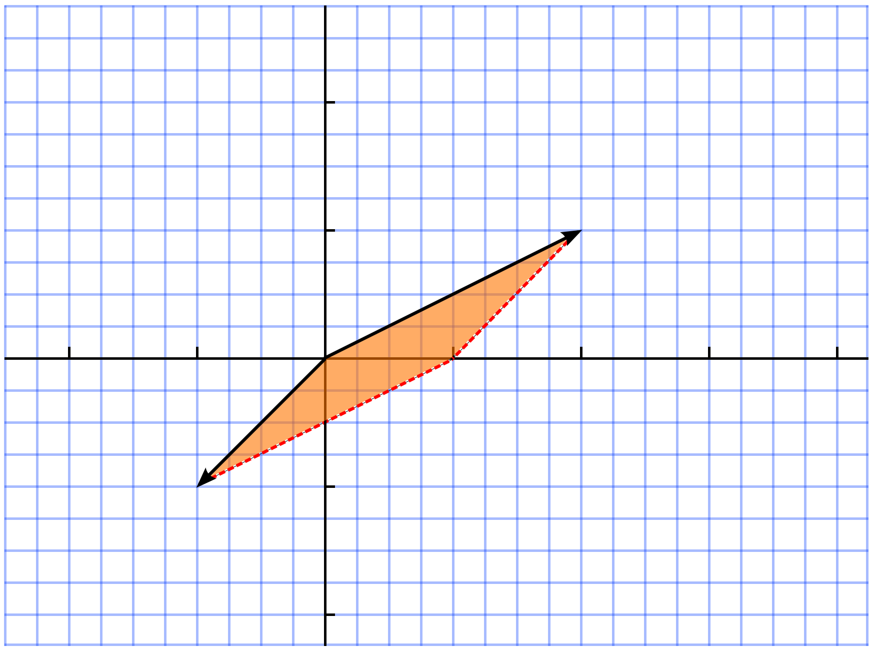
We know that with real numbers, almost every number has an inverse. $3^{-1} = 1/3$, etc. The only real number that doesn't is 0.

With matrices we have seen that there are many invertible ones and many non-invertible ones. It is not so simple as having only one exception. However, the world of vectors and matrices is a world of volumes, areas, and lengths, and as we will soon discover, it is the "magnitude" of a matrix that determines whether it has an inverse.

We know that any matrix is a bunch of vectors squished together. For example,

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

is made up of the vectors $\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. We can trace out the area of a parallelogram with sides $\vec{a}$ and $\vec{b}$.



We find that the area of this parallelogram is 3. This number is called the *determinant* of the matrix.

Definition 3.9.1. *A* determinant *of a square matrix A is the oriented area of the n-dimensional parallelepiped whose sides are give by the column vectors of A.*

An *oriented area* simply means that the sign ($\pm$) is related to which order you list the sides. It is the same concept as the right-hand rule in physics. Even though we can only visualize the determinant of a 1×1, 2×2, or 3×3 matrix, higher dimensional determinants do exist and are not terribly difficult to calculate.

Let us look at the determinant of a non-invertible matrix

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

If we try to calculate the area as base$\times$height, we see that the height$= 0$ and therefore the area is 0. This is looking more and more like real numbers, where all numbers but 0 have an inverse.

3.10 Calculating Determinants

For lower dimensional matrices we can compute the determinant with geometry, but that becomes more and more difficult as the matrix gets larger. Let's examine an alternative way to calculate the determinant.

First, notice that the determinant of I is 1. I always traces out a $1 \times 1 \times ...$ cube, so its volume is $1 * 1 * ... = 1$.

Now let's look at our elementary row operations and elementary matrices.

- Interchanging two rows

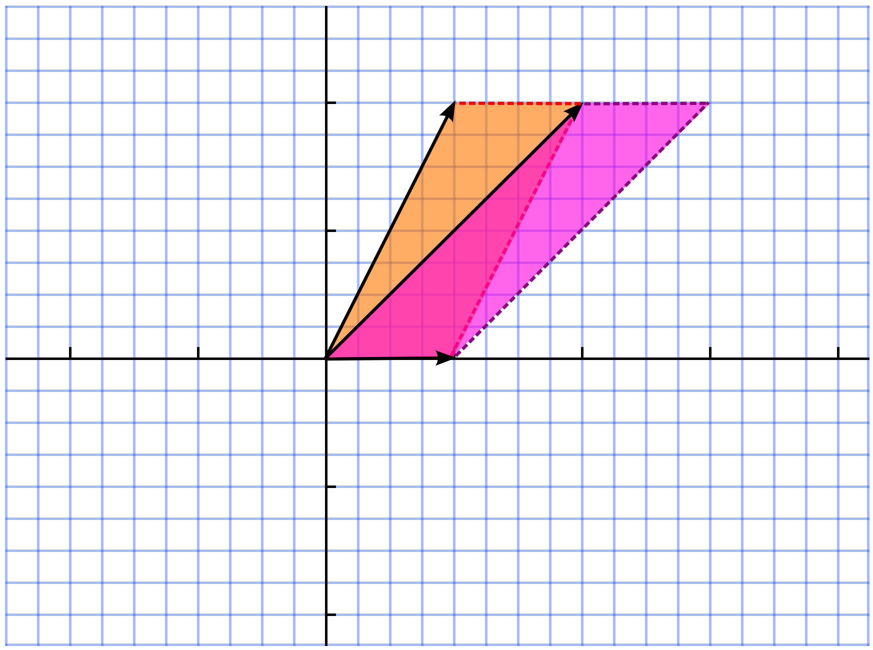
If we interchange two rows, that is the same as interchanging two edges of our parallelogram. This does not change the volume at all. All it changes is the orientation (remember, determinants are an oriented volume). That means that the determinant of a matrix, A, with two rows interchanged is the same as $-1*$determinant(A).

- Multiplying a row by a constant

If we multiply one row by a constant, it is akin to geometrically stretching one component of each vector by that scalar (which could be considered squishing the vector, if the constant is less than 1). This means that the oriented volume changes by a factor of the applied scalar multiple. To put it plainly, the determinant of a matrix, A, with one row multiplied by a scalar, c, is equal to $c*$determinant(A).

- Adding a multiple of one row to another

For this operation, let's look at a picture.



The orange area represents the determinant of the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Doing the row operation $R_1 \to R_1 + R_2$ gives us the matrix $\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$, whose determinant is represented by pink. Notice that since area of a parallelogram is base times height and our row operation did not change height, the determinant of both these matrices is the same. In fact, any row operation of this type does an analogous area-preserving operation on the parallelepiped represented by a matrix in any dimension.

This means that in order to calculate the determinant of any matrix, all we need to do is reduce it to the identity matrix while keeping track of every elementary operation we did along the way. But, this task shouldn't be difficult at all, because adding a multiple of one row to another doesn't change the determinant at all! That mean's we only have to keep track of two of our elementary row operations.

Our procedure for calculating a determinant is to use elementary row operations to put our matrix in row-echelon form. Then we multiply along the diagonal of the matrix and then multiply by all the constants we collected from the row operations we did.

Notation

Because a determinant is an oriented "magnitude," it makes sense to use vertical bars to denote a determinant. The determinant of the matrix

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

can be written

$$det(A) = \begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix}$$

However, we will predominantly use the shorthand "*det*" to represent the determinant operation. So,

$$det(A) = det \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = 3.$$

Example:

Find the determinant of

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

First, let's do the operation $R_2 \to R_2 - R_1$. This doesn't change the determinant and it leaves us with

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

Next, let's do the operation $R_3 \to 1/2R_3$. This gives us

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

and will multiply our determinant by 2. Now, let's do the operation $R_1 \to R_1 - R_3$. This gives us

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

and doesn't change the determinant. Now, let's get rid of the ones in R_3 with the operations $R_3 \to R_3 - R_1$ and $R_3 \to R_3 - R_2$:

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

Those two operations didn't affect the determinant either. And, the only step left in order to get the identity matrix is to rearrange the rows. Let's do the operations $R_3 \leftrightarrow R_2$ and then $R_3 \leftrightarrow R_1$. This multiplies the determinant by -1 twice and gives us the identity matrix.

Now that we have reduced our matrix to the identity matrix, we can compute $det(A) = (2)(-1)(-1) = 2$. Wasn't that painless!

Special Properties

A very important property of determinants is that the determinant of a product of matrices is the product of the determinants of matrices. That is $det(AB) = det(A)det(B)$. This can be seen through elementary matrices and our determinant algorithm.

Firstly, we know to find the determinant of a matrix, we decompose it into elementary matrices and, multiplying by -1 every time we used row operation 1, multiplying by a constant for row operation 2, and ignoring any elementary matrices from row operation 3. Since we know that any square matrix can be decomposed into a string of elementary matrices (with possibly one non-elementary matrix at the very end of the string), we could certainly decompose two matrices into strings of elementary matrices. When we do this, calculating the determinant with our standard algorithm gives the same result as calculating the determinant of two matrices separately and then multiplying them together.

2×2 *Formula*

Even thought it is very quick to calculate a determinant from elementary matrices, the 2×2 case occurs so often that it is worth memorizing the formula:

$$det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

3.11 Super Theorem

Now we are into the meat of Linear Algebra. We have solved systems of linear equations, we have row reduced matrices, we have computed inverse matrices, and we have explored determinants. As it turns out, all these things are related via the *Super Theorem* of Linear Algebra.

As we go on, we will add more and more things to the Super Theorem, but as it stands, this is what we have so far:

Super Theorem: If A is a square, $n \times n$ matrix, the following are equivalent:

1. $A\vec{x} = \vec{b}$ has only one solution for every $\vec{b}$. i.e. $A\vec{x} = \vec{b}$ is consistant.
2. $A\vec{x} = 0$ has only one solution, $\vec{x} = 0$.
3. The reduced row-echelon form of A is the identity matrix.
4. A can be written as a product of elementary matrices.
5. A has an inverse.
6. $det(A) \neq 0$.

Stating that a list of statements is equivalent means that if you know any one of those statements is true or false, you know that all of the statements are true or false. So, if two statemtents A and B are equivalent, then if A is true it means B is true and if B is true it means that A is true. (Note that this is logically equivalent to saying if B is false A is false and if A is false, B is false.)

We use a shorthand "$\implies$" symbol as a shorthand for "implies." For example, we may write "If A is true then B is true" as "$A \implies B$." Note that "implies" works much like a logical arrow, indicating where you can get to from which starting point.

Consider now the proposition before us: showing that statements 1, 2, 3, 4, 5, and 6 are equivalent. We will show this in the following way: $1 \implies 2 \implies 3 \implies 4 \implies 5 \implies 1$ and then show $4 \implies 6$ and $6 \implies 3$. Note that if we prove implications in this way, we will be able to follow the arrows and get from any number to any other number. Thus, the statements 1-6 will be equivalent.

Now for the proofs. First, let's prove $1 \implies 2$.

Suppose the equation $A\vec{x} = \vec{b}$ has a unique solution for every $\vec{b}$. That means that for $\vec{b} = \vec{0}$, there is exactly one corresponding $\vec{x}$. We know that $A\vec{0} = \vec{0}$ for any square matrix A, therefore $\vec{x} = \vec{0}$ is a solution to the equation $A\vec{x} = \vec{0}$, which, by our assumption, means it is the only solution. This is precisely statement 2, so if 1 is true, so is 2.

$2 \implies 3$.

We will show that "not $3 \implies$ not 2." This statement is the *contrapositive* of the statement "$2 \implies 3$," and is logically equivalent to it. We are going to show that if A is a square matrix and $rref(A) \neq I$ then $A\vec{x} = 0$ has multiple solutions.

If $rref(A) \neq I$, then $rref(A) = J$ is a diagonal matrix with at least one zero on the diagonal. Assume this zero is at position k. We obtained J via elementary matrices. That is $E_n E_{n-1} \cdots E_2 E_1 A = J$. Inverting these we get $A = E_n^{-1} E_{n-1}^{-1} \cdots E_2^{-1} E_1^{-1} J$. But, J has a zero at position k, k, so if $\vec{x}$ is the vector of all zeros and a 1 at position k, then $J\vec{x} = 0$. Which, in turn, means that $A\vec{x} = E_n^{-1} E_{n-1}^{-1} \cdots E_2^{-1} E_1^{-1} J\vec{x} = 0$, but $\vec{x} \neq 0$, so we have shown that $A\vec{x} = 0$ has multiple solutions.

$3 \implies 4$.

Suppose that $rref(A) = I$. Now, consider the process of row-reduction as multiplying by elementary matrices. Knowing that $rref(A) = I$ means that $E_n E_{n-1} \cdots E_2 E_1 A = I$, where E_i are all elementary matrices. But, all elementary row operations are invertible and therefore elemenatry matrices are invertible, so we may write $A = E_n^{-1} E_{n-1}^{-1} \cdots E_2^{-1} E_1^{-1}$. All the E_n^{-1} are elementary matrices, so we have just show that A can be written as a product of elementary matrices.

$4 \implies 5$. We need to show that if A can be written as a product of elementary matrices, then A has an inverse. If we can write $A = E_n E_{n-1} \cdots E_2 E_1$, then, because every elementary matrix is invertible, we can write $A^{-1} = E_n^{-1} E_{n-1}^{-1} \cdots E_2^{-1} E_1^{-1}$.

$5 \implies 1$. We need to show that if A has an inverse then $A\vec{x} = \vec{b}$ has a unique solution. If A has an inverse, multiply both sides by A^{-1} and we get that $\vec{x} = A^{-1}\vec{b}$ is the unique solution to the equation.

$4 \implies 6$. We need to show that if A can be written as a product of elementary matrices then $det(A) \neq 0$. If $A = E_n E_{n-1} \cdots E_2 E_1$, then $det(A) = det(E_n E_{n-1} \cdots E_2 E_1)$. By the properties of determinants, we know that $det(E_n E_{n-1} \cdots E_2 E_1) = det(E_n)det(E_{n-1}) \cdots det(E_2)det(E_1)$. Since the determinant of an elementary matrix is never zero and the product of non-zero numbers is never zero, we may conclude that $det(A) \neq 0$.

$6 \implies 3$. We need to show that if $det(A) \neq 0$ then A can be written as a product of elementary matrices. Our algorithm for calculating the determinant was to put the matrix into REF form, keeping track of all of our elementary row operations. We then multiplied all the entries on the diagonal and in turn multiplied that product by the factors obtained from our row operations. Since none of our row operations contribute a zero to the determinant, the only place a zero could come from is the diagonal of $ref(A)$. Since $det(A) \neq 0$, $ref(A)$ must have no zeros on the diagonal, so $rref(A) = I$.

Chapter 4

Vector Spaces

4.1 Vector Spaces

We've been working with linear combinations for quite some time. We know that the solution to $A\vec{x} = \vec{b}$ is the coefficients of the column vectors of A so that they point to $\vec{b}$. We also know that sometimes there are combinations of A and $\vec{b}$ so that there is no solution.

Let's clarify with an example.

Example:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

Lets give the column vectors names, $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\vec{w} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$. It is clear that $3\vec{v} = \vec{w}$. Let's look at everything we can get as a linear combination of these vectors: Taking just $\vec{v}$, we can get every point on the line $y = 2x$. Now if we also include $\vec{w}$, we can get all of the points on the line $y = 2x$. That means we don't get anything new, but we already knew that because the vectors were linear combinations of each other. What does this mean for

$\vec{b}$? It means that unless $\vec{b}$ lies on the line $y = 2x$, this system will not have a solution.

In this case, the set of all the points in the line $y = 2x$ or alternatively, all vectors that can be described by $s\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ with some value of s are called the *solution space*.

But, we have not yet defined what a space is, so let's define a *Vector Space*:

Definition 4.1.1. *A* vector space *is a set of vectors, V, with several useful properties:*

1. *You can multiply a vector by a scalar and it is still in the space. (If you multiply any vector pointing along $y = 2x$ by a scalar, it will still lie on the line $y = 2x$.)*

2. *We have a way to add vectors and when we add, the order doesn't matter.*

3. *All linear combinations of any vectors in the space are in the space. (Let's take 2 vectors in our solution space: $\vec{v}$ and $\vec{w}$. We know $\vec{v} + \vec{w} = \vec{w} + \vec{v} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, which is still on the line $y = 2x$.)*

4. *There is a zero vector, $\vec{0}$, and the sum of anything plus the zero vector is itself. ($\vec{v} + \vec{0} = \vec{v}$)*

5. *Every vector has a unique negative vector in the space and their sum is the zero vector. (If we take our vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, we know that $-\vec{v} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$ also lies on the line $y = 2x$, and we know that $\vec{v} - \vec{v} = \vec{0}$.)*

These properties seem very intuitive, and for vectors with real-number values, they are very intuitive. Later on we will work with vectors that are made of things other than real numbers—there we will have to pay closer attention to all of the rules to make sure something is a vector space.

Throughout this entire book, we have been working in the vector space R^n where n is the size of our vectors. However, as we see when solving for the solution space, it may not be all of R^n. We often are dealing with *subspaces.*

Definition 4.1.2. *$W \subset V$ (W contained in the vector space V) is called a* subspace *if W is a vectorspace.*

It may seem trivial to make such a definition, however, whereas we have to prove 5 things each time we declare something a vectorspace, if we already know it is a subset of a vectorspace, we only have to prove 1: closure.

If we show a subset $W \subset V$ of a vector space V is closed, then we know for all $\vec{x}$ in W and for any scalar k, $k\vec{x}$ is in W; thus, property 1 is satisfied. Furthermore, we can add vectors together, so property 2 is satisfied. Closure also implies property 3 is satisfies. If we can multiply by a scalar 0 and remain in the space, then the zero vector must be in W, so property 4 is satisfied. Lastly, we can multiply by the scalar -1 and get inverse vectors, so property 5 is satisfied as long as we show that W is closed.

The Shape of a Vector Space

Let's think a little bit about the shape of a vector space. The solution space to our previous example was a straight line. That is because we had linearly dependent vectors. What would the solution space looked like if the vectors were linearly independent? Let's look at the vectors $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and the new vector $\vec{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We already know $\vec{v}$ defines all the points on the line $y = 2x$. Our new vector $\vec{w}$ defines all the points on the line $y = 0$. So what is our solution space? Is it the union of the line $y = 2x$ and the line $y = 0$?

Let's make sure that the union of $y = 2x$ and $y = 0$ could actually be a vector space: Can we multiply vectors by scalars? Yes. Do we know how to add vectors? Yes. Are linear combinations of vectors also in the space? Let's test: $\vec{v} + \vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, but $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ doesn't lie on either of our lines! This means that our vector space has to be bigger than just the union of these lines. In fact, since the matrix $\begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$ is invertible, we can form any vector $\begin{bmatrix} x \\ y \end{bmatrix}$ from linear combinations of the vectors $\vec{v}$ and $\vec{w}$. This makes means the solution space must be a 2-D plane.

It is interesting how vector spaces that we've looked at so far, a point, a line, a plane, have all been "flat." But, in fact, this is a property of vector spaces. Because of closure, if we can move a tiny bit in one dimension, we can move an any amount we choose in that dimension. Therefore, if we can move a tiny direction away from a point, we can move in an entire line. If we can move a tiny direction away from a line, we can move in a whole plane. If we can move a tiny bit away from a plane, we can move in all of 3-D space!

4.2 Null Space

By now, we are quite comfortable with solving systems of linear equations. We know the solutions to the equation $A\vec{x} = 0$ can have zero, one, or infinitely many solutions. We've also been introduced to the concept of vector space. In this section, we will use the concept of a vector space to more precisely understand what zero, one, or infinitely many solutions to this particular equation means.

We're used to the equations involving $A\vec{x} = 0$. We can solve this equation for a set of vectors V that satisfy this equation. If V consists of one element, it means that there is one solution and A is invertible. If V is has infinitely many elements, then there are an infinite number of solutions to $A\vec{x} = 0$. In any case, the set V is called the *null space*.

Definition 4.2.1. *The* null space *of a a matrix A is the set of vectors that satisfy the equation $A\vec{x} = 0$.*

The null space of any matrix equation, $A\vec{x} = 0$, is a vector space. This fact can be seen through the process of solving matrix equations. First of all, we know that if $A\vec{x} = 0$ consists of only one element, then that element must be $\vec{0}$. And, $\vec{0}$ is a vector space! If there are multiple solutions, then there are some number of free variables. Writing out the solution in vector form, each one of our free variables has a vector coefficient. The solution space consists of all linear combinations of these vector coefficients. Since we are taking all linear combinations of a set of vectors, by definition it is closed. It is also a subspace, so it is a vector space.

Notice that, in essence, the null space describes the set of vectors that multiplying by A sends to zero. If A is the zero matrix, it is clear every vector is in the null space. If A is not the zero matrix, there is some vector (and therefore an entire line) that is not sent to zero. This will become an important concept later on.

4.3 Span

Definition 4.3.1. *The* span *of a set of vectors is the set of all linear combinations of those vectors.*

Span is a concept we have already been using, we are just assigning it a name.

Example:

> The span of the vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is the all points on the line $y = 2x$.

There is a useful theorem linking vector spaces and span:

If you have real-valued vectors, $\vec{a}, \vec{b}, ...$, the span of those vectors is a vector space.

It is clear that a collection n-vectors is a subset of an n-dimensional vector space. By taking the span of them, we are taking all linear combinations. Therefore the span of a set is vectors is closed under linear combination. Since the span of a set of vectors is a closed subset of a vector space, the span of a set of vectors is a vector space.

Minimal Spanning Sets

We know a set of vectors will span a flat vector space, however, we have noticed that the dimension of this vector space is not always equal to the number of vectors that we have. For example, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 6 \end{bmatrix}$ span a line, but $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ span a plane. If we use 6 dimensional vectors, it is possible to span either a 1, 2, 3, 4, 5, or 6 dimensional vector space.

Our task now is to find a way to eliminate redundant vectors from the span to get a minimal set. (Our goal is to eventually find a coordinate system for each vector space that gives us a unique way to represent vectors.)

Consider the span of the vectors

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

We will try to eliminate redundant vectors to find the minimal set that spans.

Let's write the vectors in a matrix:

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

Now, let's put it in row-echelon form.

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 2 \\ 0 & 0 & -2 & -2 & 2 \\ 1 & 2 & 3 & 5 & -1 \\ 1 & 2 & 0 & 3 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We have pivots in the first, third, and fourth row. This means that the first, third, and fourth vectors will span our entire vector space. So we know,

$$span\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 5 \\ 3 \end{bmatrix}\right) = span\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -1 \\ 2 \end{bmatrix}\right).$$

It is clear by inspection that we can write column 2 in terms of column 1. i.e. $\begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and we can write column 5 in terms of column 1 and column 3: $\begin{bmatrix} 2 \\ 2 \\ -1 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ -2 \\ 3 \\ 0 \end{bmatrix}$. Since columns 2 and 5 can be written as linear combinations of columns 1, 3, and 4, it is clear that they are redundant, and therefore the do not add anything to the span.

4.4 Basis and Dimension

Basis is just another word for minimal spanning set, but we will use it slightly differently.

Definition 4.4.1. *If the set of vectors B is a* basis *for the vector space V, it means that B is a minimal spanning set that spans V.*

Because a basis spans our vector space, we know that any vector in our space can be written as a linear combination of basis vectors.

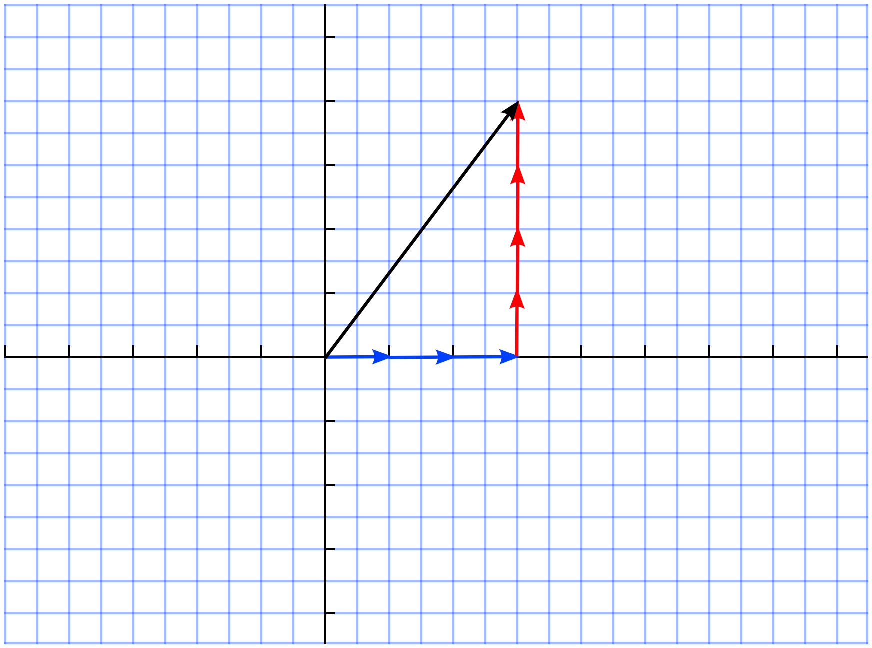
Example:

Consider the set of vectors

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \vec{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Now let's pick a random vector, $\vec{w} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$. What combination of $\vec{e_1}, \vec{e_2}, \vec{e_3}$ gives us $\vec{w}$? The answer is clearly $3\vec{v_1} + 4\vec{v_2} + 5\vec{v_3}$. Or we can consicely write it as $\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ where the first row is the coefficient of v_1 the second row is the coefficient of v_2, etc. In this basis, $\vec{w}$ looks just like itself! And for good reason. It just so happens that this basis is called the *standard basis.*

Let's examine a little more geometrically what it means to describe a vector in terms of basis vectors. Again, consider the standard basis $\vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. To be the vector $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ means to go 3 in the e_1 direction and 4 in the e_2, just like this picture indicates.

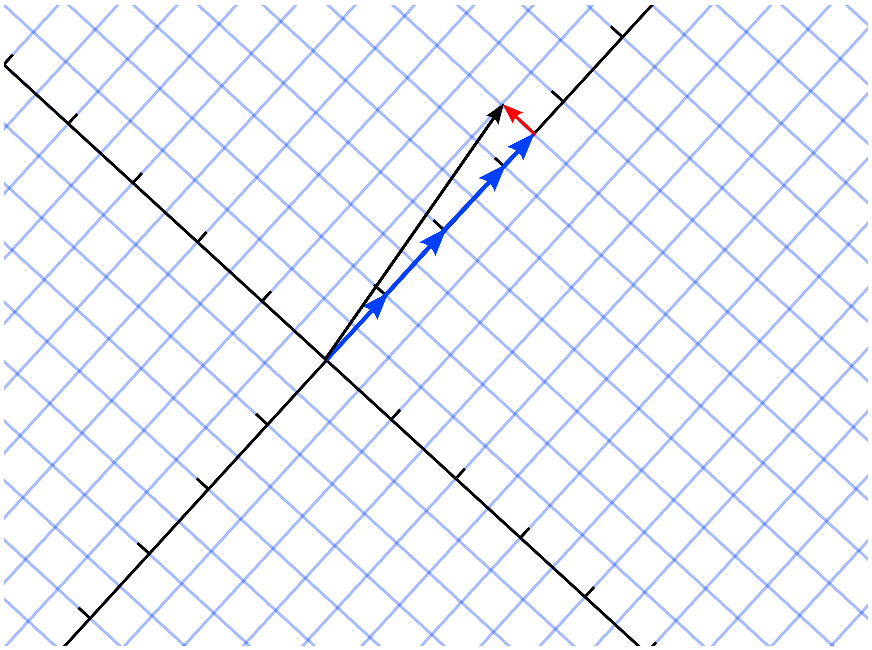


So, the representation of $\vec{v}$ in the standard basis is $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

Now consider a rotated and streached standard basis

$$B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

In order to get to $\vec{v}$ we have to go 3.5 times $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and then .5 times $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.



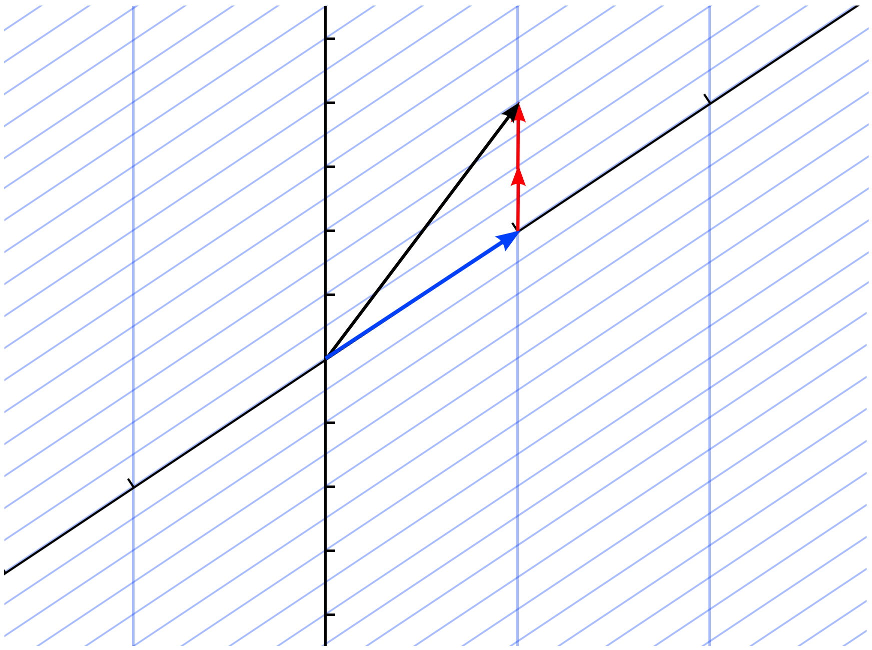
The representation of $\vec{v}$ in the B basis is $\left[\ \vec{v}\ \right]_B = \begin{bmatrix} 3.5 \\ .5 \end{bmatrix}$.

Notice that the two representations we got for $\vec{v}$ are both equally good but involve different numbers because they are written in terms of different bases. We use a subscript letter to indicate what basis we are using so we don't get confused. If we are using the standard basis, we can leave off the letter.

Let's now consider an even stranger basis:

$$C = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

This basis is different from our other bases because its vectors are not perpendicular to each other. But, let this not discourage us, we can still figure out how to get to the vector $\vec{v}$ all the same.



It is clear from the picture that we must go 1 times $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and 2 times $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. That means that $\vec{v}$ represented in terms of the C basis is

$$[\ \vec{v}\ ]_C = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

4.5 Change of Basis

Now that we know what a basis is, it would be nice to use our Linear Algebra methods to figure out how to algebraically compute a change of basis.

In our previous example with $C = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we saw that

$$[\ \vec{v}\ ]_C = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Since the basis vectors in C are written in the standard basis, if we compute

$$C\,[\ \vec{v}\ ]_C = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = [\ \vec{v}\ ]_S,$$

where S is the standard basis.

What we did was quite simple. We just wrote the basis vectors of C in a matrix, and left multiplication by that matrix takes vectors from the C basis to the standard basis. We'll call this matrix $\begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$ a *transition matrix*

Definition 4.5.1. *A* transition matrix *is a matrix that takes vectors in written in one basis and converts them to another basis.*

Since the transition matrix $\begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$ takes vectors written in the C basis and spits them out in the standard basis (S), let's write this matrix as $\left[\ S \leftarrow C\ \right]$.

We now know that if we have a vector written in the C basis, we can use the equation

$$\left[\ S \leftarrow C\ \right] \left[\ \vec{v}\ \right]_C = \left[\ \vec{v}\ \right]_S .$$

What if we wanted to convert things from the standard basis to the C basis? By solving this matrix equation, we see that

$$\left[\ S \leftarrow C\ \right]^{-1} \left[\ \vec{v}\ \right]_S = \left[\ \vec{v}\ \right]_C .$$

So, in fact, $\left[\ S \leftarrow C\ \right]^{-1}$ might be written as $\left[\ C \leftarrow S\ \right]$, the matrix that takes things in the standard basis and converts them to the C basis.

We now have a simple equation for finding a change-of-basis matrix, so we can algebraically compute any change of basis we could ever dream of!

Chapter 5

Linear Operators

5.1 Linear Transformations

So far we have only been viewing the world of matrices as a way to solve systems of linear equations. Now, we are going to approach things from a slightly different angle: *linear tranformations*.

Definition 5.1.1. *A linear tranformation is a function $T : R^n \rightarrow R^m$ taking n-vectors to m-vectors that satisfies the following properies:*

1. $T(a\vec{v}) = aT(\vec{v})$ *for scalar* a.
2. $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$.

This definition seem fairly abstract until you learn that left multiplication by a matrix is a linear transformation that we have been working with the entire time.

If A is a matrix and $\vec{v}$ is a vector, by the properties of matrix multiplication, we know $A(c\vec{v}) = cA\vec{v}$. And, if $\vec{w}$ is also a vector, $A(\vec{v} + \vec{w}) = A(\vec{v}) + A(\vec{w})$.

Let's bring our concepts of linear tranformations and basis vectors together. First off, we know that if we have a basis $B = \vec{v}_1, \vec{v}_2, ..., \vec{v}_n$, we can write any vector $\vec{x}$ as

$$\left[\ \vec{x}\ \right]_B = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

Now imagine that we have some operator T that operates on our vectors. We can do a few computations:

$$T(\vec{x}) = T(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n) =$$

$$c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n)$$

What does this tell us? It means that we can know everything about a linear transformation just by knowing what it does on a basis. If we know where it takes a basis, we can simply find our c_i via the change of basis formula and now we know where the transformation took our vector.

The fact that we can describe a linear transoformation completely just by describing what happens on a basis brings us to a very improtant theorem.

Any linear transformation, T, can be represented as left multiplication by some matrix A.

5.2 Range and Kernel

The *range* and *kernel* of a linear transformation are directly related to two problems we've already been solving: The set of $\vec{x}$ that make $A\vec{x} = \vec{b}$ consistant and the solution to $A\vec{x} = 0$.

Definition 5.2.1. *The* range *of a linear transformation T is the set of all $\vec{b}$ such that $T(\vec{x}) = \vec{b}$ has a solution.*

Another word for the range of T is the *image* of T. Range in plain terms is everywhere that a transformation may send you. The range of a linear transformation is analagous to the solution space to a matrix equation.

Example:

> The range of the zero transformation is the zero vector. This should be fairly straight forward. If our transformation takes everything to zero, the only place we can get is zero, therefore zero is our range.

Kernel, in a way, is the complement of range:

Definition 5.2.2. *The* kernel *of a linear operator T, denoted $kern(T)$ is the set of all vectors $\vec{x}$ such that*

$$T(\vec{x}) = 0.$$

The kernel of a linear transformation is every vector that is sent to zero. The kernel is the same thing to a linear transformation as the null space is to a matrix.

It is clear that range and kernel are complements of each other. If our transformation is the zero transformation then our range is the zero vector and our kernel is the whole space. If our transformation is invertible then our range is the whole space and the kernel consists of only the zero vector.

5.3 Linear Operators

So far, we have been talking about linear transformations from an n-dimensional space to an m-dimensional space, but the really interesting things happen when we are allowed to be recursive, which means a linear transformation that operates on the same space it outputs. We call these linear transfromations *linear operators*.

Definition 5.3.1. *A* linear operator *is a linear transformation $T : R^n \to R^n$ that takes n-dimensional vectors to n-dimensional vectors.*

If T is a linear operator, then it operates on the same space it outputs to. That means it makes sense to talk about $T(T(x)) = T^2(x)$. We use exponents to indicated how many times we have applied the transformation T.

All linear transformations can be represented by left multiplication by a matrix and linear operators are not exception. But, in the case of linear operators, we get a special bonus: our matrices are square. This means we can apply all the knowledge we have about square matrices to linear operators!

5.4 Changing Basis

We know that if we know what a linear operator does on a basis, we know everything about it. This might make us wonder about how to figure out what a transformation is in a different basis.

It's clear that if we have a linear transformation in terms of the C basis and we have vectors in terms of the B basis, we need to find the transition matrices $\left[\ C \leftarrow B\ \right]$ and $\left[\ B \leftarrow C\ \right]$. If we had those, we could simply compute our linear transformation as

$$\left[\ B \leftarrow C\ \right] \left[\ T\ \right]_C \left[\ C \leftarrow B\ \right].$$

From our work with changing basis, we know that $\left[\ C \leftarrow B\ \right] = \left[\ B \leftarrow C\ \right]^{-1}$. If we let Q be our change of basis matrix, we then know

$$\left[\ T\ \right]_B = Q^{-1}TQ.$$

The relationship between T and $Q^{-1}TQ$ is a very special one. In fact, it is an equivelance relation called *similarity*.

Definition 5.4.1. *Two matrices A and B are said to be* similar *if there exists an invertible matrix Q such that*

$$A = Q^{-1}BQ.$$

Two matrices being similar means that the both represent the same transformation but with respect to different bases. This is why it is important to keep that pesky distinction between a matrix and a linear transformation: because many matrices may be representing the same transformation.

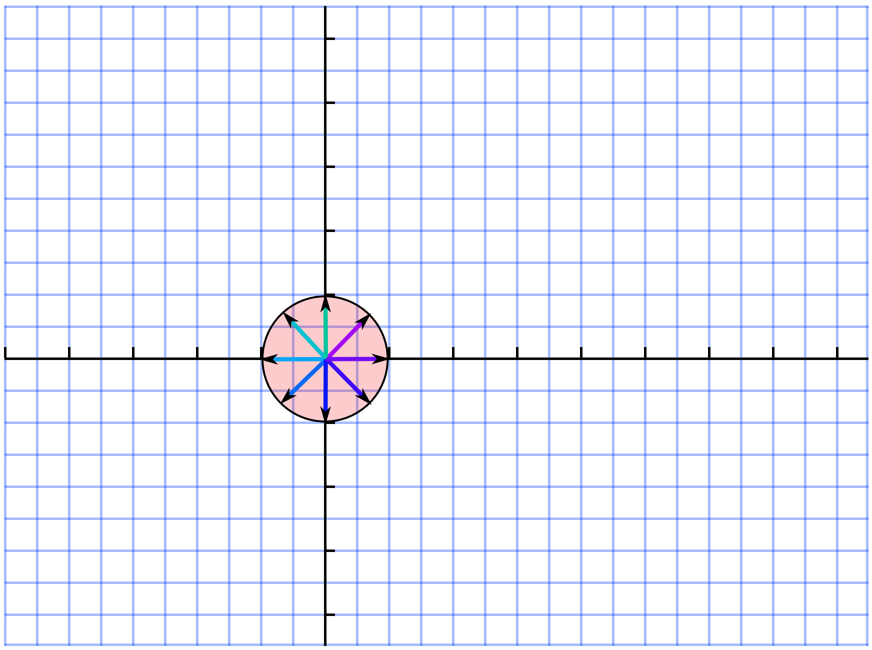
Chapter 6

Eigenvalues and Diagonalization

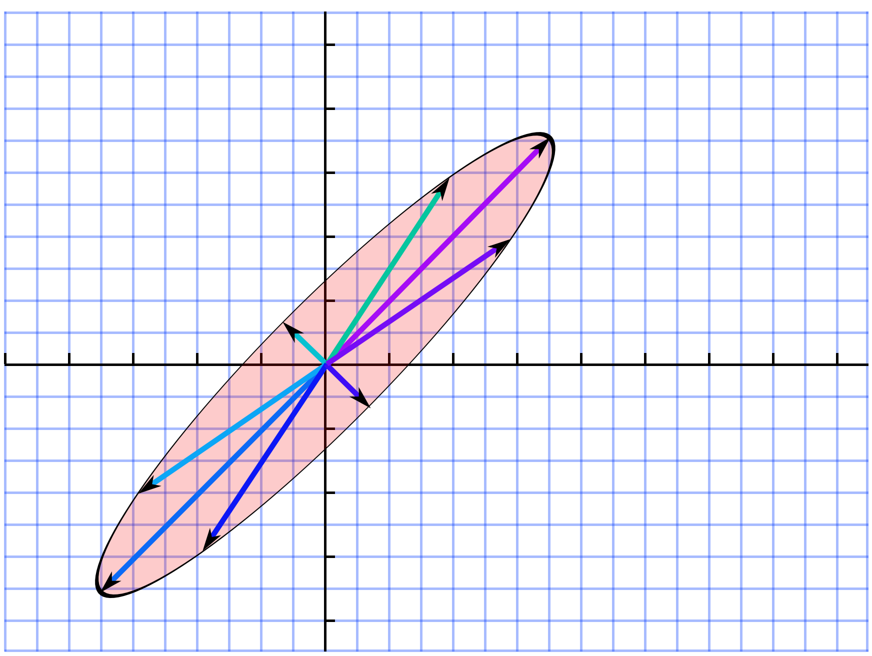
6.1 Eigenvalues and Eigenvectors

Let's consider the linear transformation that is left multiplication by the matrix $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

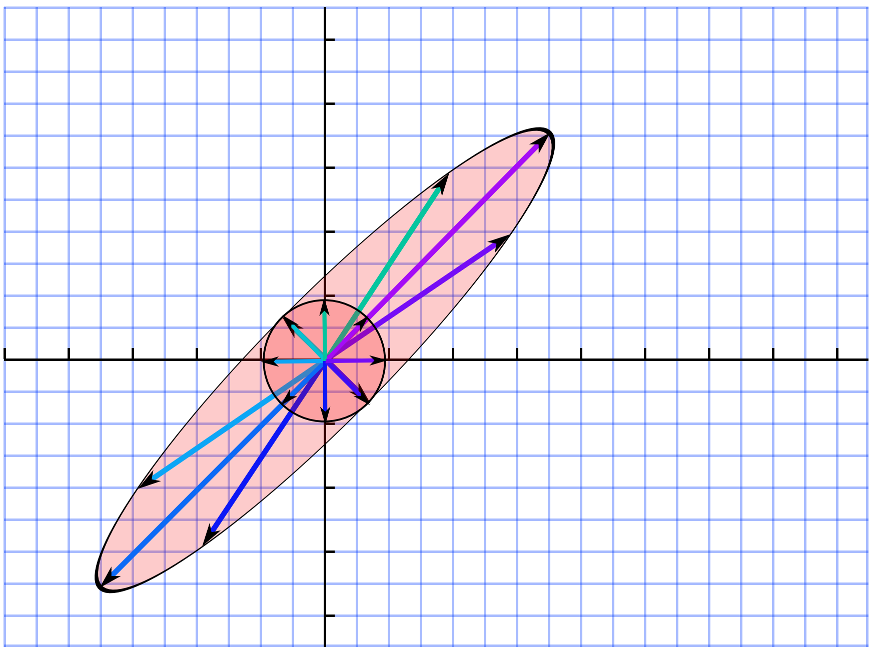
To get a better idea of what this transformation is doing, let's look at what it does to all the vector of length 1. Before the transformation, these vectors point in a circle.



After the transformation, the vectors are pointing in an ellipse.



Most of the vectors changed direction and changed length, but pay special attention to the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. If we overlay the pictures, we see that these vectors didn't change direction, they just changed length. In fact, vectors pointing in these two directions are the only vectors that won't change direction when our linear transformation is applied.



These special vectors that change length but not direction are called *eigenvectors*.

Definition 6.1.1. *An* eigenvector *of a transformation T, is a vector that doesn't change direction when T is applied. That is, if $\vec{v} \neq 0$ is a vector, then*

$$T(\vec{v}) = \lambda\vec{v}$$

where there scalar λ is called the eigenvalue corresponding to the vector v.

We don't include the zero vector as an eigenvector because any linear transformation of the zero vector will be the zero vector.

So, eigenvectors are vectors whose direction is invariant under a linear tranformation and eigenvalues is the amount that each eigenvector is scaled.

6.2 Finding Eigenvectors

Eigenvectors will soon prove to be extremely useful, so let's figure out how they can be computed.

Most of math is finding a way to relate new problems to problems that have already been solved. So, let's play around a bit with the definition of an eigenvector: $\vec{x}$ is an eigenvector of T if $T\vec{x} = \lambda\vec{x}$. Subtracting $\lambda\vec{x}$ from both sides, we see

$$T\vec{x} - \lambda\vec{x} = 0.$$

If we could somehow factor out the $\vec{x}$, this would look just like our homogenious matrix equation that we have been solving for so long. Unfortunately it doesn't make sense to subtract a scalar from a matrix. However, the identity will come to our rescue:

$$T\vec{x} - \lambda\vec{x} = T\vec{x} - \lambda I\vec{x} = (T - \lambda I)\vec{x}$$

which gives us our very important final equation:

$$(T - \lambda I)\vec{x} = 0.$$

If we were given a λ, we could simply find the kernel of $(T - \lambda I)$ and that would be the corresponding eigenvector(s).

Note: we know from finding the null space and kernel of operators the dimension of the null space or kernel maybe be 0, 1, 2, etc., so the dimension of $kern(T - \lambda I)$ may be greater than one in which case there are multiple eigenvectors corresponding to that λ (consider the identity transformation—every vector is an eigenvector corresponding to the eigenvalue of 1) or under special circumstances the dimension may be zero in which case there are no eigenvectors.

Characteristic Polynomial

It's easy to find eigenvectors if we already know the eigenvalue, but what if we don't know the eigenvalue?

Don't worry, we still have hope. Remember our equation $(T - \lambda I)\vec{x} = 0$? We know that this only has non-trivial solutions when $(T - \lambda I)$ is singular. But, if $(T - \lambda I)$ is singular, then $det(T - \lambda I) = 0$, and the determinant is something we can actually calculate.

Definition 6.2.1. *The* characteristic polynomial*, $p(x)$, of a linear operator T is*

$$p(x) = det(T - xI).$$

Let's look at an example to see the proccess of solving.

Example:

> Consider the linear transformation represented by left multiplication by the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$.
>
> We want to find all λ such that $(T - \lambda I)\vec{x} = 0$ has non-trivial solutions. That means that $det(T - \lambda I) = 0$.
>
> Let's compute:
>
> $$det(T - \lambda I) = det\left(\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)$$
>
> $$= det\begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda)$$
>
> Now that we know the characteristc polynomial is $(1-\lambda)(3-\lambda)$, we want to know when $(1-\lambda)(3-\lambda) = 0$, which occur exactly when $\lambda = 3$ and $\lambda = 1$.

The eigenvalues of a transformation are simply the roots of the characteristic polynomial. Once we have the eigenvalues, finding the eigenvectors is easy, so now we can tackle finding the eigenvectors for any linear transformation!

6.3 Non-real Eigenvectors

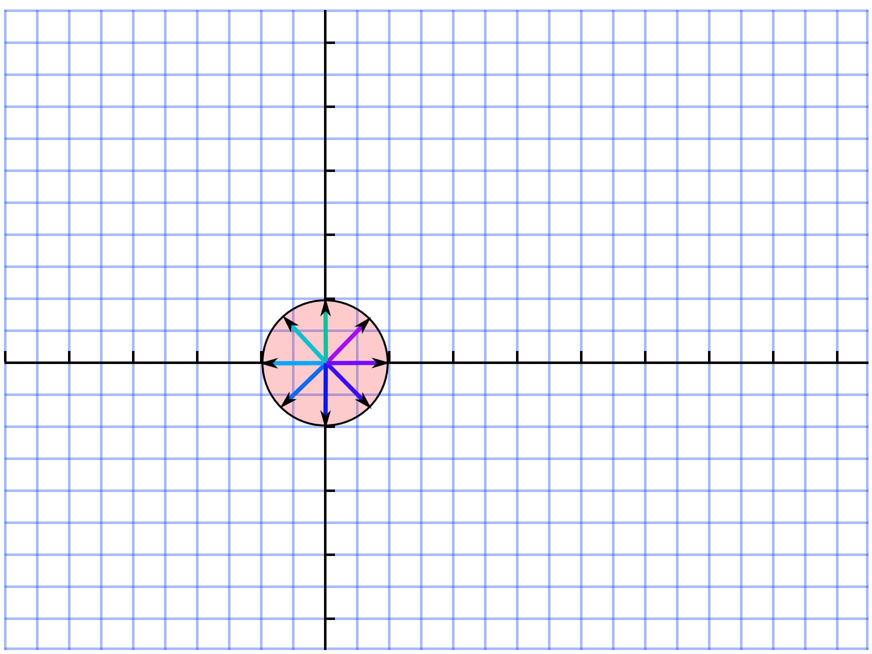
So far, all of our examples have turned out to have real eigenvalues and vectors, but this is not always the case (particularly in physics).

Consider the characteristic polynomial of the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $p(x) = x^2 + 1$. If we were to set this equal to zero, we would see that x must be $\pm i$ where

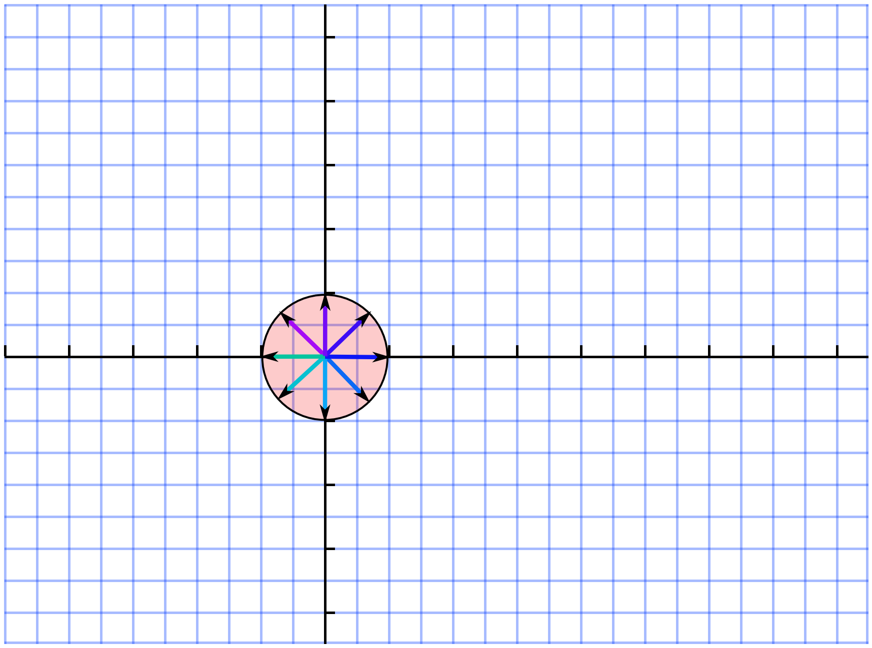
$i = (-1)^{1/2}$ for the equation to have solutions. This means our eigenvalues are $\lambda = \pm i$.

What are our corresponding eigenvectors? If we solve for our null space the same way we have been, only using complex numbers instead of real ones, we see that our eigenvectors are $\begin{bmatrix} i \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -i \\ 1 \end{bmatrix}$. And, sure enough if we plug these vectors into our transformation, we see that they scale by $\pm i$.

To get a better idea of what it really means to have a complex eigenvector, let's figure out what this transformation is actually doing. If we look at a circle of vectors beforehand



and afterwards



we see that our transformation rotates all vectors counterclockwise by $\pi/2$.

From the picture it is clear that all real-valued vectors change direction. It is also clear that if we had a rotation by any amount (other than 0, π, 2π, ...) we won't have any real eigenvectors. It would seem that complex eigenvectors and rotations are closely related, and in fact they are. Describing rotations was part of the reason complex numbers were invented in the first place.

6.4 Diagonalization

We all love diagonal matrices—they are so easy to work with. It would certainly be nice if there were some way to turn ordinary matrices into diagonal ones. Let's ponder and see what we can come up with.

First off, what would it mean to have a basis of eigenvectors? If we have eigenvectors $\vec{v}_1$ and $\vec{v}_2$ of an operator T and our eigenvectors form a basis for our vectorspace, then we can write any vector $\vec{w}$ in our vector space as

$$\vec{w} = a_1\vec{v}_1 + a_2\vec{v}_2.$$

Call our basis of eigenvectors E. Now, let us apply our transformation to $\vec{w}$.

$$T(\vec{w}) = T(a_1\vec{v}_1 + a_2\vec{v}_2) = a_1T(\vec{v}_1) + a_2T(\vec{v}_2),$$

but $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors so

$$a_1T(\vec{v}_1) + a_2T(\vec{v}_2) = a_1\lambda_1\vec{v}_1 + a_2\lambda_2\vec{v}_2.$$

What an amazing occurance! If we consider our vector $\left[\ \vec{w}\ \right]_E = \left[\begin{array}{c} a_1 \\ a_2 \end{array}\right]$ in the eigen basis, then

$$\left[\ T(\vec{w})\ \right]_E = \left[\begin{array}{c} \lambda_1 a_1 \\ \lambda_2 a_2 \end{array}\right]$$

which is the same as left multiplication by the diagonal matrix $\left[\begin{array}{cc} \lambda_1 & 0 \\ 0 & \lambda_2 \end{array}\right]$.

In a way, we have found an ideal basis for a transformation. If we can write our linear transformation in terms of a basis of eigenvectors, then we get a diagonal matrix, which is ideal as far as arithmetic is concerned.