Linear Algebra: The Bureau 42 Primer

W. Blaine Dowler

August 8, 2010

Contents

1 Old School Algebra **1**
- 1.1 The Field Axioms . . . 1
  - 1.1.1 Closure Under Addition . . . 2
  - 1.1.2 Closure Under Multiplication . . . 2
  - 1.1.3 Commute Under Addition . . . 3
  - 1.1.4 Commute Under Multiplication . . . 3
  - 1.1.5 Associate Under Addition . . . 3
  - 1.1.6 Associate Under Multiplication . . . 4
  - 1.1.7 The Additive Identity . . . 4
  - 1.1.8 The Multiplicative Identity . . . 5
  - 1.1.9 Additive Inverses . . . 5
  - 1.1.10 Multiplicative Inverses . . . 6
  - 1.1.11 Distributive Property . . . 6
- 1.2 The Exponent Rules . . . 8
  - 1.2.1 $x^a \cdot x^b = x^{a+b}$ . . . 8
  - 1.2.2 $\frac{x^a}{x^b} = x^{a-b}$ . . . 8
  - 1.2.3 $x^0 = 1$ . . . 9
  - 1.2.4 $x^{-b} = \frac{1}{x^b}$ . . . 9
  - 1.2.5 $(x^a)^b = x^{ab}$ . . . 9
  - 1.2.6 $x^{\frac{1}{n}} = \sqrt[n]{x}$ . . . 9
  - 1.2.7 $x^n \cdot y^n = (x \cdot y)^n$ . . . 10
- 1.3 The Limitations Of This Algebra . . . 10
  - 1.3.1 Vectors . . . 10
  - 1.3.2 Rotations . . . 12
- 1.4 Next Lesson . . . 12

2 Linear Algebra: Eliminating Rules **13**
- 2.1 Matrix Algebra . . . 13
- 2.2 The Axioms of Algebra: Which Do We Keep? . . . 14
  - 2.2.1 Addition . . . 14
  - 2.2.2 Multiplication . . . 17
  - 2.2.3 Addition and Multiplication . . . 24
- 2.3 Summary . . . 24

3 All Your Basis Vectors Are Belong To Us **27**
3.1 Vectors: Not Just For Breakfast Anymore . . . 27
3.1.1 Operations With Vectors . . . 28
3.2 Basis Vectors: The Building Blocks . . . 29
3.3 Transposes . . . 30
3.4 Metrics . . . 31
3.5 Inner and Dot Products . . . 32
3.6 One-forms . . . 33
3.7 Norms . . . 33
3.8 Gram-Schmidt . . . 34
3.9 Cross Products . . . 36
3.10 Summary . . . 37

4 Inverting Matrices to Solve Systems **39**
4.1 Systems of Linear Equations the Old and Busted Way . . . 39
4.2 Systems of Linear Equations the New Hotness Way . . . 40

5 Eigenvectors and Eigenvalues **41**

6 More Than Meets the Eye **43**

7 Fun with Idempotents and other Projections **45**

8 Operators **47**

9 Working With Tensors **49**
9.1 Defining Tensors . . . 49

Chapter 1

Old School Algebra

> I didn't come here to tell you how this is going to end. I came here to tell you how it's going to begin.
>
> Neo, *The Matrix*

1.1 The Field Axioms

Although it may come as a shock to most students, there are only 11 arbitrary rules to algebra, called the field axioms. Anything and everything we do in algebra is a direct or indirect consequence of these rules. Sometimes definitions make the connections obscure, but it all follows from these rules. (For example, the exponent rules can be hard to see from these rules, because the definition of the exponent is a shorthand. If we explicitly write x^3 as $x \cdot x \cdot x$ instead, then the rest follows naturally.)

To that end, we begin these lessons by looking at the field axioms in explicit detail, ensuring that their applications are clear and easy to work with.

First, to explain the name: why are they called "field axioms?" The name is descriptive. The algebra we are used to is based on the real number system, meaning any number you could possibly fit on a number line, whether than number is 0, 4, $-\frac{1}{2}$, π, or $\frac{1-\sqrt{5}}{2}$. The numbers do not need to be pretty to work. The real numbers can be added, subtracted, multiplied and divided with each other quite readily.[1] We call an algebra of this type a field, which is where the first word in the name comes from. The second word, axiom, basically means

[1]Well, technically, the adding and multiplying parts are all that are really important. As we are about to learn, there's really no such thing as subtraction or division.

that it's a rule we invented because it makes sense to us. As you will find, these rules are entirely logical, and it's easy to see why they are all true.[2]

Now, for the rules themselves.

1.1.1 Closure Under Addition

Informally: When you add two numbers, your answer is a number.

The only surprising part of this rule is that we actually bother to write it down. We write it down to make math useful: if $2+2 =$elephant, then the whole exercise is pointless, and we might as well go home. Mathematicians need to add a few qualifiers to this one, though. It's called "closure" because we have a closed set of numbers: you can't add two real numbers and get something that isn't a real number. To understand why this is an important rule, think of the odd numbers. 1 and 3 are both odd, but $1 + 3 = 4$ and 4 isn't odd. The odd numbers alone cannot form a complete algebra because they break this rule, which is why we need to have more than the odd numbers in our set.

Semi-formally: If x and y are both real numbers, then $x + y$ is also a real number.

Formally: Let $(F, +, \cdot)$ define an algebraic field, where F is a set of numbers, $+$ is the operation of addition, and $\cdot$ is the operation of multiplication. If x and y are both elements (or members) of the set F, written $x, y \in F$, then $x + y$ is a member of F, written $x + y \in F$, for all x and y, written $\forall x, y$.

1.1.2 Closure Under Multiplication

Informally: When you multiply two numbers, your answer is a number.

Again, this is not a surprising rule, and it is introduced for the same reasons as the corresponding addition rule. This is the rule that says we need to use *all* the real numbers, and not just, for example, the numbers from 0 to 100 to make a complete algebra. If we try to put an upper limit on our numbers, we'll break this rule. For example, 50 and 60 are both numbers between 0 and 100, but $50 \cdot 60 = 3000$ and 3000 is well over 100.

Semi-formally: If x and y are real numbers, then $x \cdot y$ is a real number.

[2] At least, that's the case until our next lesson, at which time we start throwing rules away.

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. If $x, y \in F$, then $x \cdot y \in F \forall x, y$.

1.1.3 Commute Under Addition

Informally: When two numbers are added together, order doesn't matter.

Again, this doesn't surprise people. We know that $1+2$ is the same as $2+1$. If it helps you remember the label, remember that people who carpool on the *commute* to work take turns in the driver's seat. They still get to the same workplace at the end of the day.

Semi-formally: $x + y = y + x$

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then $x + y = y + x \forall x, y \in F$.

1.1.4 Commute Under Multiplication

Informally: When two numbers are multiplied together, order doesn't matter.

Semi-formally: $x \cdot y = y \cdot x$

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then $x \cdot y = y \cdot x \forall x, y \in F$.

1.1.5 Associate Under Addition

Informally: When adding three numbers together, it doesn't matter which pair you add first.

Again, this is something we've known since we first started learning about addition. What it really says is that it doesn't matter how we *associate* the numbers in brackets when we add them.

Semi-formally: $(x + y) + z = x + (y + z)$.

You may notice that the combination $(x + z) + y$ isn't listed. It's the same as the others, but you can't get there with this rule alone. You must also apply the commuting rule, as follows:

$(x + y) + z = x + (y + z) = x + (z + y) = (x + z) + y$

It seems like doing things the long way, but learning to apply the 11 rules in pedantic detail makes things easier, not harder.

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then $(x+y)+z = x + (y + z) \forall x, y, z \in F$.

1.1.6 Associate Under Multiplication

Informally: When multiplying three numbers together, it doesn't matter which pair you multiply first.

Semi-formally: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then $(x \cdot y) \cdot z = x \cdot (y \cdot z) \forall x, y, z \in F$.

1.1.7 The Additive Identity

Informally: The number 0 is special. When you add 0 to any number, you get that number back as the answer: it doesn't change the *identity* of the original number.

Semi-formally: $x + 0 = x$

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then, $x + 0 = x \forall x \in F$.

There is something interesting to note: we can prove that 0 is the *only* number that has this property, but *only* because addition has the commuting rule! Imagine that 0_a and 0_b are two numbers that have this property. We can add them to each other, and prove that they are the same number, like so:

$0_a = 0_a + 0_b = 0_b + 0_a = 0_b$

We need the commuting rule to justify the interchange of the numbers, because the additive identity rule is only formally defined when the 0 appears on the *right* of the other number.

1.1.8 The Multiplicative Identity

Informally: The number 1 is special. If we multiply a number by 1, we get the original number back, without changing the number's identity.

Semi-formally: $x \cdot 1 = x$.

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then, $x \cdot 1 = x \forall x \in F$.

As with the 0, we can prove that there is only one number that has this property, and it is 1.

If you look carefully, you may notice a pattern to these rules. If we didn't already know what $+, \cdot, 0$ and 1 meant, or if I'd used the symbols $\star, \ominus, \vee$ and $\square$ instead, we would be completely unable to tell the difference between addition and multiplication in the formal definitions. Every odd rule has been an addition rule, and the even rules have been the corresponding multiplication rule. This is not a failing of the formal notation; it's actually a strength. If properly interpreted, we see that these rules are *not* distinguishable, and that we have been telling them apart at this stage because of the existing cultural bias to the symbols and the vocabulary, and *not* because of some inherent quality of the math. It is the next pair of rules that allow us to determine exactly what $+, \cdot, 0$ and 1 are in the real numbers.[3]

1.1.9 Additive Inverses

Informally: When you add a number to its negative, you get 0 for an answer.

Semi-formally: $x + (-x) = 0$

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. $x + (-x) = 0 \forall x \in F$.

If these 11 rules of algebra are really going to allow us to do anything and everything we can possibly do in algebra, then restricting them to just addition and multiplication seems limiting. That's only because we think of subtraction and multiplication as distinct operations, and we only think of them that way because that's the way we were taught to think about them back when we were knee high to a St. Bernard. In fact, subtraction is actually a shorthand notation:

[3]The "in the real numbers" bit is not a set of extra, pointlessly included words. When we get further, and start dealing with unusual algebras, we'll use the symbols 0 and 1 to mean entirely different things.

when we write $x - y$, we actually mean $x + (-y)$. This is why order matters in subtraction when it doesn't matter in addition: the sign needs to move with the variable or number after it. So, $x - y \neq y - x$, but $x + (-y) = (-y) + x$.

1.1.10 Multiplicative Inverses

Informally: When you divide a number by itself, you get 1 for an answer, *except* for the number 0, which cannot be a divisor.

This is how we know what the symbols $+, \cdot, 0$ and 1 mean. Multiplication is the operation that has an except under the inverses rule, so we know what + and $\cdot$ mean, and 0 is the number that is the exception, so we know what that means. 1 is now meaningful because we know it is an identity, but that it isn't zero.

Semi-formally: $x \div x = x \cdot \frac{1}{x} = 1$ when you are allowed to divide by x.

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then, $x \cdot \frac{1}{x} = x \cdot x^{-1} = 1 \forall x \in F$ such that $x \neq 0$.

This rule shows an underlying feature of all of these field axioms: they are *reversible*. The reason we can't divide by 0 is that it would *not* be reversible. For example, both $5 \cdot 0$ and $6 \cdot 0$ work out to 0. If we start with zero, we can't divide by zero and know with absolute certainty whether we started with 5, 6, or something else entirely.

Highly attentive readers may notice that we have specific rules to show what $x + 0$ and $x \cdot 1$ work out to, but we haven't defined explicitly that $x \cdot 0 = 0$ for any and all values of x. That's not a rule in itself, but a consequence of other rules. The proof of this fact depends on our final rule.

Highly attentive readers may have also noticed that every rule has either addition *or* multiplication, but not both. The final rule is the rule that has both, and tells us how they mix together.

1.1.11 Distributive Property

Informally: When multiplying a number into brackets, the number outside distributes itself through the brackets.

Semi-formally: $x \cdot (y + z) = x \cdot y + x \cdot z$

Formally: Let $(F, +, \cdot)$ define an algebraic field as before. Then $x \cdot (y + z) = x \cdot y + x \cdot z \forall x, y, z \in F$.

This is the final arbitrary rule of algebra. None of these are particularly surprising in isolation: the surprise usually comes in learning these are the *only* rules we need. They truly are, and we can use them to show a number of things, including $x \cdot 0 = 0$ and the F.O.I.L. "rule" for multiplying binomials.

We start with $x \cdot 0 = 0$. To prove something is true in mathematics, we take one side of the equation, apply known rules, identities, and definitions, and turn it into the other side. Before we do this, we need to take a good look at our second rule, about closure under multiplication. Although it seems trivial, it's actually quite powerful: if multiplying two numbers together produces a number, then we can regard such a product of two numbers as a *single* number with its own inverse. We will do exactly this as a part of our proof, given that $(F, +, \cdot)$ is an algebraic field and $x, a \in F$:

$x \cdot 0 =$	$x \cdot (a + (-a))$	By using the definition of an additive inverse.
$x \cdot 0 =$	$x \cdot a + x \cdot (-a)$	By the distributive property.
$x \cdot 0 =$	$(x \cdot a) + (-(x \cdot a))$	Because, after all, $x \cdot a$ is just a number with an inverse.
$x \cdot 0 =$	0	Because $x \cdot a$ and $-x \cdot a$ are additives inverse of each other.

We can also prove that the F.O.I.L. "rule" is not a rule, but a consequence of these rules, by recognizing (through the "closure under addition" axiom) that $a + b$ is just a number.

$(a + b) \cdot (c + d) =$	$(a + b) \cdot c + (a + b) \cdot d$	Because $a + b$ is just a number to distribute through the right hand brackets.
$(a + b) \cdot (c + d) =$	$c \cdot (a + b) + d \cdot (a + b)$	Because numbers commute when multiplied.
$(a + b) \cdot (c + d) =$	$c \cdot a + c \cdot b + d \cdot a + d \cdot b$	Distributive property.
$(a + b) \cdot (c + d) =$	$a \cdot c + a \cdot d + b \cdot c + b \cdot d$	Because addition and multiplication commute.

This is the exact combination you get through the "first, outside, inside, last" mnemonic you were asked to memorize. This approach takes more steps (partly because the last step was done entirely to arrange things as expected) but it requires less memorization, and understanding this approach means you can start multiplying trinomials and more without batting an eye.

1.2 The Exponent Rules

There are a few rules for dealing with exponents in algebra that are also consequences of these rules, and not new rules unto themselves. These will not only be listed, but will be derived using the axioms, both to demonstrate their power and to reveal *why* the exponent rules are what they are.

The seven exponent "rules" are as follows:

$$\begin{aligned}
x^a \cdot x^b &= x^{a+b} \\
\frac{x^a}{x^b} &= x^{a-b} \\
x^0 &= 1 \\
x^{-b} &= \frac{1}{x^b} \\
(x^a)^b &= x^{ab} \\
x^{\frac{1}{n}} &= \sqrt[n]{x} \\
x^n \cdot y^n &= (x \cdot y)^n
\end{aligned}$$

1.2.1 $x^a \cdot x^b = x^{a+b}$

We start with explicit values for a and b and show how they work. Let's take $a = 3$ and $b = 2$ to begin with. If we then remember the *definition* of an exponent, and write the repeated x values out explicitly, the logic will be apparent:

$$x^3 \cdot x^2 = (x \cdot x \cdot x) \cdot (x \cdot x) = x \cdot x \cdot x \cdot x \cdot x = x^5 = x^{3+2}$$

At no point did the *logic* of the above explicitly depend upon the fact that the exponents were 3 and 2. If we have a xs multiplied together, and then multiply that by b xs, then you have $a + b$ xs left when you are done.

1.2.2 $\frac{x^a}{x^b} = x^{a-b}$

Again, we can write this out explicitly for one example to expose the logic. Again choosing $a = 3$ and $b = 2$, we find

$$\frac{x^3}{x^2} = \frac{x \cdot x \cdot x}{x \cdot x} = \frac{\cancel{x} \cdot \cancel{x} \cdot x}{\cancel{x} \cdot \cancel{x}} = \frac{x}{1} = x = x^1 = x^{3-2}$$

We subtract because the variables cancel.

1.2.3 $x^0 = 1$

This isn't even a new rule among the exponent rules! This is a direct consequence of the previous rule, but with $a = b$. In that case, the number in the numerator (x^a) is exactly the same as the denominator (x^a) so we have the following:

$$x^0 = x^{a-a} = \frac{x^a}{x^a} = \frac{\cancel{x^a}}{\cancel{x^a}} = 1$$

Many students instinctively assume $x^0 = 0$. This is because their instincts are coming from the more familiar addition than multiplication. Remember, the exponent of a number is the *count* of the times that number is being multiplied together. If none are being multiplied, the problem reduces to the *multiplicative* identity, not the additive identity. The multiplicative identity is 1.

1.2.4 $x^{-b} = \frac{1}{x^b}$

Again, this is a special case of the preceeding rules. This time, we combine the two rules we just had, starting with the rule for x^{a-b} where $a = 0$. Then we have the following:

$$x^{-b} = x^{0-b} = \frac{x^0}{x^b} = \frac{1}{x^b}$$

1.2.5 $(x^a)^b = x^{ab}$

This we work out explicitly using the first rule.

$$(x^a)^b = \underbrace{x^a \cdot x^a \cdot \ldots x^a}_{b \text{ times}} = x^{\overbrace{a+a+a+\ldots+a}^{b \text{ times}}} = x^{a \cdot b}$$

1.2.6 $x^{\frac{1}{n}} = \sqrt[n]{x}$

This one is a little more esoteric. To figure it out, we need to us the previous rule. What happens to $(x^a)^b = x^{ab}$ when $a = \frac{1}{n}$ and $b = n$?

$$(x^a)^b = (x^{\frac{1}{n}})^n = x^{\frac{1}{n} \cdot n} = x^1 = x$$

This tells us that n instances of $x^{\frac{1}{n}}$, when multiplied together, give us x alone. We already have a number that does that: the nth root of x. For example, if $y = x^{\frac{1}{2}}$, then $y^2 = x$, or $y = \sqrt{x}$. Similarly, if $\left(x^{\frac{1}{3}}\right)^3 = y^3 = x$, then $x^{\frac{1}{3}} = y = \sqrt[3]{x}$.

1.2.7 $x^n \cdot y^n = (x \cdot y)^n$

The final rule is a more direct consequence of the field axioms, particularly the commuting axiom. We will write it out explicitly for $n = 3$, and trust the reader to see that the logic would apply regardless of the exact value of n.

$$x^3 \cdot y^3 = x \cdot x \cdot x \cdot y \cdot y \cdot y = x \cdot x \cdot y \cdot x \cdot y \cdot y = x \cdot y \cdot x \cdot y \cdot x \cdot y = (x \cdot y) \cdot (x \cdot y) \cdot (x \cdot y) = (x \cdot y)^3$$

1.3 The Limitations Of This Algebra

There are two reasons to do math in life.

1. **Passion:** Some people feel a passion for the subject, and find mathematics fun. Those of us in this category will learn linear algebra simply because it's a field of mathematics that is there to be learned.

2. **Application:** For the other 99.9999% of the population, we need to find a practical application that they can apply it to which saves them time and effort in the long run. The rest of this section will establish what some of those applications are.

1.3.1 Vectors

In the late 1800s, physicists had not yet adopted linear algebra. Vectors are a part of linear algebra, and are needed to efficiently represent the physical world.

A sprinter need worry only about his or her speed, because the direction is implied: the sprinter travels from the starting line to the finish line. The finish line in front of the sprinter is close enough that it will never go out of sight, and the sprinter will not get lost.

An intercontinental airline pilot, on the other hand, needs to know a lot more than that. Not only is the destination beyond the line of sight when the plane takes off, but there are very, if any, distinctive landmarks that can be used to guide the flight across an ocean. The pilot must be concerned both with the speed of the aircraft, *and* with the direction of travel. This pilot must compensate for any wind which may blow the plane off course, as well. With the need to maintain limited amounts of excess fuel for the trip, the flight path must be right the first time. If the wind changes, the pilot must rotate the aircraft to adjust the direction of the craft.

How do we represent both speed and direction of travel? If we only use the real numbers with the algebras described above, then we can't. We would be limited to those numbers that we can put on a number line, but a velocity (which is both speed and direction) needs two numbers to represent itself.

From here, we have two choices. We can do what James Clerk Maxwell did (because he didn't know he had alternatives) when he compiled the equations that describe electricity and magnetism in physics: use more than one variable, and come up with distinct sets of equations for every single one of them. Nowadays, we write $\vec{\bigtriangledown} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ for one of Maxwell's equations. Poor Maxwell wrote them out before he was aware of vectors, so he had to write all three of the following out longhand:

$$\begin{aligned}
\frac{\mathrm{dF}}{\mathrm{dy}} - \frac{\mathrm{dE}}{\mathrm{dz}} &= -\frac{\partial A}{\partial t} \\
\frac{\mathrm{dD}}{\mathrm{dz}} - \frac{\mathrm{dF}}{\mathrm{dx}} &= -\frac{\partial B}{\partial t} \\
\frac{\mathrm{dE}}{\mathrm{dx}} - \frac{\mathrm{dD}}{\mathrm{dY}} &= -\frac{\partial C}{\partial t}
\end{aligned}$$

Vectors save in the writing, at least, as we can see from this.[4] A vector is actually an ordered combination of numbers, with each number representing a different piece of the entire picture. For example, a three component vector can have one number that represents the east and west directions, one that represents the north and south directions, and one that represents the up and down directions. If you are on the surface of the Earth and not standing at one of the Poles, this becomes an effective vector. There are other ways to represent

[4] Attentive readers with backgrounds in physics or engineering may now start to see why $\vec{B}$ became the symbol for the magnetic field: without vectors, Maxwell had a different variable for every directional component in every quantity. When linear algebra was introduced to physicists several decades later, shortly after a poor physicist recreated the entire framework from scratch over the course of years, they needed to find a way to use the new tools without making the new representations impenetrable to those familiar with the old notations. By complete and utter coincidence, E was the y component of the electric field, M was the y component of the magnetization vector, and P was the y component of the polarization vector. Three sensible variable assignments were the best that could be managed, so the y component of *every* vector was taken as the de facto standard symbol for the vector quantity.

vectors in three dimensional space, such as vectors with one length and two angles,[5] but all require three numbers in some sort of combination.

1.3.2 Rotations

Now that we see the need for vectors, we need to figure out what we can do to or with them. We know one application we need: our intercontinental pilot needs to figure out how to rotate his direction vector before he runs out of fuel!

Rotations can be represented with linear algebra, but not with vectors. Sure, a vector could be used to represent the direction vector after it has been rotated, but doing it that way means manually adjusting every single component of every vector by hand every time. We would prefer a more general way to do this, and linear algebra will provide such a way with objects called operators instead of vectors. It would be really nice if we could use a rotation object R to rotate a vector $\vec{x}$ through something as simple as $R \cdot \vec{x}$, so that is what we will strive for.

1.4 Next Lesson

In the next lesson, we will examine the properties that we know of vectors and rotations through day to day life, and figure out what kind of algebra we will need to get the job done.

[5] One angle would represent direction counter-clockwise from, say, east, while the other would represent direction up from the horizon. This is one way to do it, but far from the only one.

Chapter 2

Linear Algebra: Eliminating Rules

> You have a problem with authority, Mr. Anderson. You believe you are special, that somehow the rules do not apply to you. Obviously, you are mistaken.
>
> Rhineheart, *The Matrix*

2.1 Matrix Algebra

Unbeknownst to James Clerk Maxwell when he compiled the equations describing electricity and magnetism, there was already an existing mathematical framework which would simplify his work. Linear algebra described matrices and vectors, which (as we're about to learn) neatly and perfectly do the job we are looking to do.

At a fundamental level, a matrix is nothing more than an organized arrangement of numbers. Most of us are already seen a form of these: when plotting points on a graph, the ordered pairs that represent a point form a matrix. There are rules for adding and multiplying matrices, and these rules conform to what we need for vectors and rotations, among other types of objects.

A matrix can have more than one dimension. For example, the following is a two dimensional (3×3) matrix:

$$\begin{pmatrix} 1 & -3 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It contains a lot more than two numbers, but they arranged in columns and rows. As long as you know both the column and row, you can find any one number in that array. Those are the two dimensions: columns and rows. The terminology can get confusing, because it is inconsistent between math and physics. In physics, a three dimensional vector is a mathematical matrix with one column and three rows: in that case, you need only know the *row* a number can be found in, because they are all in the same column. This has three dimensions in physics, but only one in math. As the thrust of this series is in the mathematical structure more than the physics applications, we will use the mathematical terminology from here on.

2.2 The Axioms of Algebra: Which Do We Keep?

2.2.1 Addition

The axioms related to addition are relatively straightforward, once we define what equality, "addition" and 0 mean for a matrix. The equality part is easy: two matrices are equal if each individual entry in the first matrix is equal to the corresponding entry in the second matrix. For example,

$$\begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix} \neq \begin{pmatrix} 3 & 4 & 1 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix}$$

because the entries in row one, columns two and three do not match. They are the same numbers, but they are not the same numbers in the same places, so they are *not* equal.

The most natural definition of addition for a matrix turns out to be the most convenient one: we add two matrices by adding the corresponding entries. For example,

$$\begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix} + \begin{pmatrix} 2 & 7 & 1 \\ 8 & 2 & 8 \\ 1 & 8 & 2 \end{pmatrix} = \begin{pmatrix} 3+2 & 1+7 & 4+1 \\ 1+8 & 5+2 & 9+8 \\ 2+1 & 6+8 & 5+2 \end{pmatrix} = \begin{pmatrix} 5 & 8 & 5 \\ 9 & 7 & 17 \\ 3 & 14 & 7 \end{pmatrix}$$

The upper left entry in the first matrix is added to the upper left entry in the second matrix to form the upper left entry in the resultant matrix, and so forth. Although this is a natural definition, it does beg the question: what if we need to add two matrices that are different shapes?

The answer to this question is as simple as they come: you can't. Addition among matrices is *only* defined when the two matrices are the same shape, with the same number of rows and the same number of columns.

We need one final definition: what does 0 mean for a matrix? If A is a matrix, can we define a sensible 0 in such a way that $A+0=A$ no matter what we choose for A? Thankfully, we can: the zero matrix is the matrix for which every entry is 0. Similarly, we can define the negative of a matrix A as the matrix $-A$, whose entries are made by taking the negatives of each individual entry of A. With these definitions, we can now test and confirm whether or not the axioms of addition hold.

Closure

As mentioned when the closure property was introduced in lesson one, simply stating that the sum of two numbers was a number wasn't quite detailed enough for an entirely formal treatment. We need to expand this definition to state that the sum of two numbers is the same *type* of number that the two addends[1] were. With all of these tests, we'll use a 2×2 matrix to save space. These arguments can all be extended and generalized to matrices of arbitrary dimensions ($m\times n$, where m and n are any natural numbers[2].).

To verify this, we add two matrices with arbitrary entries:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}$$

Each individual entry in the new matrix, such as $a+e$, is a number of the same type as the original. In other words, if a and e are both real numbers, then $a+e$ is a real number. Therefore, adding these two matrices produces a new 2×2 matrix of real numbers, so the closure property is satisfied.

[1]Terminology reminder: if $a+b=c$, then a and b are called *addends* and c is called the *sum*. Most people recall the term "sum," though "addend" tends to be lost as a result of disuse.

[2]The natural numbers are $1, 2, 3, 4, \ldots$

Commutativity

This one we can verify more directly:

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} &= \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} \\
&= \begin{pmatrix} e+a & f+b \\ g+c & h+d \end{pmatrix} \\
&= \begin{pmatrix} e & f \\ g & h \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\end{aligned}$$

In the middle step, we've noted that the individual entries are sums of real numbers, and we know those commute. As you may already suspect, our definition of addition means we'll be able to inherit many of the properties from the real numbers themselves.

Associativity

This one takes more leg work, but the principle is the same:

$$\begin{aligned}
\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix}\right) + \begin{pmatrix} i & j \\ k & l \end{pmatrix} &= \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} + \begin{pmatrix} i & j \\ k & l \end{pmatrix} \\
&= \begin{pmatrix} (a+e)+i & (b+f)+j \\ (c+g)+k & (d+h)+l \end{pmatrix} \\
&= \begin{pmatrix} a+(e+i) & b+(f+j) \\ c+(g+k) & d+(h+l) \end{pmatrix} \\
&= \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e+i & f+j \\ g+k & h+l \end{pmatrix} \\
&= \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \left(\begin{pmatrix} e & f \\ g & h \end{pmatrix} + \begin{pmatrix} i & j \\ k & l \end{pmatrix}\right)
\end{aligned}$$

Again, we find that our definition of addition inherits this property from the real number system.

Identity

The identity is one of the easiest to prove, given the definitions of the 0 matrix and addition. We essentially prove that the property holds just by written things down.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a+0 & b+0 \\ c+0 & d+0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Inverses

The additive inverse (negative) of a matrix is also a naturally defined object.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix} = \begin{pmatrix} a+(-a) & b+(-b) \\ c+(-c) & d+(-d) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Thus, we know we can retain every property of addition for matrices. This is quite convenient. Before we can test the same for multiplication, we need to define what multiplication is.

2.2.2 Multiplication

One might expect that our definition of matrix multiplication would be comparable to that for addition: we must have matrices that are identical in size, and then we multiply the entries as usual. This is simple, it's natural, and it's easy to work with.

It's also entirely useless for real world applications.

Imagine a spear of length l lying out on a road, parallel to the road itself. We'll call the blunt end the origin of our coordinate system, and align the x-axis in the direction the spear is pointing, with the y axis pointing straight up. The coordinates of the end of the spear are then given by the coordinate pair $(x, y) = (l, 0)$. If we want to rotate this spear through an angle θ towards the positive y-axis,[3] then we need to somehow transform our vector from $(l, 0)$ into $(l\cos\theta, l\sin\theta)$.

This can't be done with component by component multiplication, for two reasons relating to the y components.

1. The original y component was 0, and the resultant y component is not. There is no x that satisfies the equation $x \times 0 = l\sin\theta$.

[3] In other words, the spear will make an angle θ with the ground.

2. The new y component contains the variable l. This variable appears only in the x component before the multiplication, so there has to be some level of mixing between the two variables.

The situation compounds if we start with a vertical spear. We want the mathematical object that rotates our spear through an angle θ to be the same *regardless* of the spear's starting position, so that the mathematical object can be applied in all cases efficiently. In this case, we need to transform $(0, l)$ into $(-l \sin\theta, l\cos\theta)$[4] with the same mathematical object that transforms $(l, 0)$ into $(l\cos\theta, l\sin\theta)$.

The useful definition of multiplication is not particularly intuitive, but it is useful. We multiply rows *into* columns, in the following manner:

$$\begin{pmatrix} {\color{red}a} & {\color{green}b} \\ {\color{blue}c} & d \end{pmatrix} \times \begin{pmatrix} {\color{red}e} & {\color{green}f} \\ {\color{blue}g} & h \end{pmatrix} = \begin{pmatrix} {\color{red}a} \times {\color{red}e} + {\color{green}b} \times {\color{blue}g} & {\color{red}a} \times {\color{green}f} + {\color{green}b} \times h \\ {\color{blue}c} \times {\color{red}e} + d \times {\color{blue}g} & {\color{blue}c} \times {\color{green}f} + d \times h \end{pmatrix}$$

There is no mathematical need for the colours above: they are included to help the reader track the source of the entries being multiplied together. Here is a more concrete example:

$$\begin{aligned} \begin{pmatrix} {\color{red}1} & {\color{green}2} \\ {\color{blue}3} & 4 \end{pmatrix} \times \begin{pmatrix} {\color{red}5} & {\color{green}6} \\ {\color{blue}7} & 8 \end{pmatrix} &= \begin{pmatrix} {\color{red}1} \times {\color{red}5} + {\color{green}2} \times {\color{blue}7} & {\color{red}1} \times {\color{green}6} + {\color{green}2} \times 8 \\ {\color{blue}3} \times {\color{red}5} + 4 \times {\color{blue}7} & {\color{blue}3} \times {\color{green}6} + 4 \times 8 \end{pmatrix} \\ &= \begin{pmatrix} 5 + 14 & 6 + 16 \\ 15 + 28 & 18 + 32 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} \end{aligned}$$

Essentially, to find the entry that appears in row r and column c in our resultant product matrix, we multiply the entries in row r for the left matrix into the entries in column c for the right matrix one pair at a time and add the products.

Addition with matrices had the restriction that the two matrices must be identical in size. Multiplication also has size restrictions, but it does not have the same restriction, which was a problem when rotating our spear. To define our string of products that we add, the number of *columns* in the matrix on the left must match the number of *rows* in the matrix on the right. So, if had a matrix with two rows and four columns (2×4 matrix) and try to multiply it into itself, we can't. There are four columns in the left matrix, but only two rows in the right matrix: we have nothing to do with the entries in the "extra" columns. However, we *can* multiply a 2×4 matrix into a 4×5 matrix, and we'll get a 2×5 matrix as a result. In this notation, as long as the "middle" dimensions match,

[4]The actual derivation of these results is based on trigonometry and geometry from junior high, middle school, or high school, depending on your local region. It will not be reviewed here.

the "outside" dimensions survive the multiplication and given the dimensions of the resultant matrix. In this case, the 4 in the "middle" dimension is important mainly in telling us how *many* pairs of numbers we need to multiply together and then add to get each entry in the resultant product matrix.

As quick test of practicality, we should check that we can, in fact, build the rotation matrix we need to represent the tricks we are doing with our spear. At this stage, we build it by hand through inspection and brute force: a formal construction method will come in a later lesson. For now, I will just write down the answer and show that it works.

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \times \begin{pmatrix} l \\ 0 \end{pmatrix} = \begin{pmatrix} l\cos\theta \\ l\sin\theta \end{pmatrix}$$

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \times \begin{pmatrix} 0 \\ l \end{pmatrix} = \begin{pmatrix} -l\sin\theta \\ l\cos\theta \end{pmatrix}$$

This does exactly what we want it to do. In fact, if one is familiar with trigonometric identities,[5] one can show that using this to rotate through an angle θ and then rotate through an angle ϕ, it even becomes exactly the same as rotating through the angle $\alpha = \theta + \phi$.

Armed with these definitions, we can now attempt to tackle the axioms defining multiplication.

Closure

For matrices to be closed under multiplication, the product needs to have the same basic characteristics as the matrices that multiply to form it. That includes the dimensions as well as the elements. In short, for closure to apply, we would need to be able to multiply and $m \times n$ matrix by an $m \times n$ matrix and get an $m \times n$ matrix as the answer. There's only one way can do that: m must equal n. In other words, to build a complete algebra including matrix multiplication, we *must* restrict ourselves to square matrices. Now, there are times and situations in which square matrices are not appropriate or applicable for the task at hand. We can *use* rectangular matrices, we just won't be able to apply the closure property to form a complete algebra when we do.

[5]Specifically, one needs to be aware that $\cos(a+b) = \cos a \cos b - \sin a \sin b$ and that $\sin(a+b) = \sin a \cos b + \cos a \sin b$. These shall not be proven here.

Commutativity

If matrices A and B commute under multiplication, then we must have $AB = BA$. Looking at the definition of multiplication, this is restrictive as well. If A is a 2×4 matrix and B is a 4×2 matrix, then the product of the matrices is defined in both arrangements, but AB creates a 2×2 matrix while BA creates a 4×4 matrix. The only way we have a chance at commuting matrices is if both A and B are square matrices with the same dimensions.

Even then, there is no guarantee that matrices commute under multiplication. Let's look at our earlier example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \times \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

However, if we change the order of the matrices, we get an entirely different result:

$$\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \times \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$$

Generally speaking, we cannot guarantee that matrices commute. There are conditions in which they do, but those conditions won't be covered until lesson five.

Associativity

Up to this point, we have been using 2×2 matrices to show specific examples of the way matrices function instead of the general form. Until now, that has been because the informal version has been quicker than the formal version of the proof in question, and it easily extended to the general case. With associativity under multiplication, the formal proof becomes more efficient than a specific example. To get to that proof, we must first define some notation.

Let A be a matrix. It is customary to label regular variables with lowercase letters from the end of the Latin alphabet, such as x, y and z, and to label constants with lowercase letters from the beginning of the Latin alphabet, such as a, b and c. Matrices are generally labeled with uppercase letters from the beginning of the alphabet, such as A, B and C. Elements of a matrix are referred to by the corresponding lowercase letter with subscripts labeling the row and column that element came from, so a particular entry from matrix A would be labeled a_{rc}. For a concrete example, if

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

then $a_{11} = 1$, $a_{12} = 2$, $a_{21} = 3$, and $a_{22} = 4$.

With this notation, then the product of two matrices can be expressed as finite sums. So, if $AB = C$, then

$$c_{rc} = \sum_n a_{rn} b_{nc}$$

In the case of a 2×2 matrix, the sum would allow n to take on the values of 1 or 2. If A is a 3×9 matrix and B is a 9×82 matrix, then n would take on the values of 1, 2, 3, 4, 5, 6, 7, 8 or 9. Now, with this notation established, we can proceed to check matrices for associativity.

If matrices associate under multiplication, then $(AB)\,C = A\,(BC)$. Assuming the multiplications are all defined, then the element in row r and column s on the left hand side is given by

$$[(AB)\,C]_{rs} = \sum_m \left(\left(\sum_n a_{rn} b_{nm} \right) c_{ms} \right) = \sum_{m,n} a_{rn} b_{nm} c_{ms}$$

where we have used the rules of sums from high school mathematics to combine the two sums under a single symbol. Also, the multiplication of AB first is represented by the brackets.

This calculation is the same regardless of the exact values of r and s. So, if this is true for *any* values of r and s, then this must be true for *all* values of r and s. All we need to do is to check and see if we can get an identical result from the right hand side of our original equation, multiplying BC out first.

$$[A\,(BC)]_{rs} = \sum_n \left(a_{rn} \left(\sum_m b_{nm} c_{ms} \right) \right) = \sum_{m,n} a_{rn} b_{nm} c_{ms}$$

This is identical. As the dimensions of the resultant matrix depend on the order matrices are *written* rather than the order they are *multiplied*, we can also guarantee that the dimensions of the result match in both cases. In short, if the multiplications are defined at all, matrix multiplication is associative. This is true even when matrices are rectangular instead of square, as no step in the above algebra depends upon having the same upper limit for r as we have for s.

Identity

A multiplicative identity matrix I must satisfy the equation $AI = A$ for any possible A. In short, we want

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \times \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Again, the exact nature of the I matrix is easier to see if we use the formal summation notation. We want this:

$$a_{rc} = \sum_n a_{rn} i_{nc}$$

There is an easy way to achieve this. Notice that, when $n = c$, the element of A involved in our sum is a_{rc}, which is exactly what we want the sum to be. So, we can build the I matrix such that $I_{rc} = 1$ when $r = c$ and $I_{rc} = 0$ when $r \neq c$. In other words, I must have a 1 in every entry along the diagonal that starts in the upper left corner, and a 0 everywhere else. This does have another nice side effect. If we examine the product IA we find the following:

$$[IA]_{rc} = \sum_n i_{rn} a_{nc} = a_{rc}$$

which tells us that the identity matrix exhibits this behaviour whether it is multiplied in from the right *or* from the left. This is, in fact, the only matrix that commutes under multiplication with *every* other matrix. This will be explored more fully in lesson five.

As a final note, it needs to be pointed out that the identity matrix only behaves as such if both matrices are square, meaning that r and c have the same upper limits. In other words,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

regardless of the exact values of a, b, c and d, but if we move to rectangular matrices, we run into difficulties. For example, in the multiplication

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we lose the last row of our second matrix. If we reverse the order and multiply them the other way around, we get the following:

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 0 \end{pmatrix}$$

which has added an entire column of 0 entries to the result.

In short, we have the same restrictions on the multiplicative identity that we had on the closure property of multiplication: we can satisfy the relevant property *if and only if* we restrict our attention to a particular size of square matrix. We can work with rectangular matrices, but we can't build complete multiplicative algebras with them.

Inverses

This is another section with details that will be filled in during a later lesson. (In this case, lesson four.) Ultimately, we are looking for a system that satisfies

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \times \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

If we multiply out the left hand side, we find that this results in four different equations:

$$\begin{aligned} ae + bg &= 1 \\ af + bh &= 0 \\ ce + dg &= 0 \\ cf + dh &= 1 \end{aligned}$$

If we apply (and apply and apply) the regular rules of algebra, we find that we can satisfy this relationship if

$$\begin{aligned} e &= \frac{d}{ad - bc} \\ f &= \frac{-b}{ad - bc} \\ g &= \frac{-c}{ad - bc} \\ h &= \frac{a}{ad - bc} \end{aligned}$$

So, it can work, but *only* if $ad - bc \neq 0$. If $ad = bc$, then $ad - bc = 0$ and we would have to divide by 0 to build our inverse matrix, which isn't allowed. A matrix of different dimensions would have a different denominator to deal with, but it would still have a denominator that would not be allowed to equal zero. In short, like the commutative property, it may hold in some cases, but we can't guarantee an inverse for every matrix.

2.2.3 Addition and Multiplication

We have checked each multiplication property and each addition property individually. Now let's see if they still work in combination as we expect.

Distributive Property

Given the matrices A, B and C, assume $B+C$, $A\times(B+C)$, $A\times B$ and $A\times C$ are all defined. Does the distributive property hold? Again, it's easiest to show this using the formal summation notation.

$$\begin{aligned}[A\times(B+C)]_{rs} &= \sum_n a_{rn}\left(b_{ns}+c_{ns}\right)=\sum_n\left(a_{rn}b_{ns}+a_{rn}c_{ns}\right)\\ &= \sum_n a_{rn}b_{ns}+\sum_n a_{rn}c_{ns}=[A\times B+A\times C]_{rs}\end{aligned}$$

So, if the matrices are of the appropriate sizes for the relevant multiplications and additions to be defined at all, then the distributive property holds.

2.3 Summary

If we restrict the size of our matrices to square matrices, we can retain some but not all of our field axioms as listed in lesson one.

With square matrices, we keep the following rules:

1. Closure under addition
2. Associativity under addition
3. Commutativity under addition
4. Additive identity
5. Additive inverses
6. Closure under multiplication
7. Associativity under multiplication
8. Multiplicative identity
9. Distributive property

Without the multiplicative commutativity and inverses, we cannot have an algebraic field. Instead, we have an algebraic *ring*. It is a perfectly valid algebra, but it is missing two properties that we have been trained to use reflexively over our years of public school. We must fight these reflexes, particularly in more complicated situations. For example, with matrices, we cannot use the "F.O.I.L." acronym for squaring a binomial:

$$(A+B)^2 = A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2$$

The simplification of cross terms works if and *only* if $AB = BA$, which is not something we can promise with matrices. When we deal with rotations in more detail,[6] we'll see that we absolutely *must* discard this property in many real life situations.

[6] Lesson eight.

Chapter 3

All Your Basis Vectors Are Belong To Us

> This is your last chance. After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes.
>
> Morpheus, *The Matrix*

3.1 Vectors: Not Just For Breakfast Anymore

The basic unit of measure for quantities that depend both upon magnitude and direction is the vector. In contrast, a scalar has magnitude only, without direction. For example, saying "I am driving at $90\frac{km}{h}$" is a scalar, but "I am driving North at $90\frac{km}{h}$" is a vector, as the direction is included. The direction is meaningful only because there is a standard definition of the word "North." If we did not already know what that meant, then it would be as meaningless as saying "to my right" with no information about the direction the speaker is facing.

The most common coordinate system for real world situations is the Cartesian, or rectangular, coordinate system. In this system, we have three axes, generally labeled x, y and z, which are at right angles to each other in a "right handed" arrangement, meaning that if you open your right hand flat, align your right thumb along the x axis and your straightened fingers along the y axis, then your palm would point in the z direction. The point where the three axes meet is called the origin. These are the coordinate axes covered in high school. So, a point with coordinates (x, y, z) can also define a vector coming from the origin and ending at that point, thus defining both magnitude and direction. That vector $\vec{v}$ is then denoted

$$\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

when expressed in Cartesian coordinates. We can also express this vector by a list of components, where $v_1 = x$, $v_2 = y$ and $v_3 = z$.

Sometimes the standard coordinate systems are not the most convenient ones. For example, if you are describing the position of a horse on a merry go round, then the cylindrical coordinates are more appropriate. We define the z axis as the axis around which the merry-go-round rotates, and then we can simplify our coordinates. Instead of having constantly changing values of x, y and z, we can use the distance from the z axis r, the angle θ relative to some fixed reference direction, and keep the height from the ground z. In this case, r would be constant for any given horse, θ changes uniformly according to the rotation of the merry-go-round, and z changes slightly as the horse bobs up and down. Similarly, there are situations in which a spherical symmetry is the most practical, such as when describing points in the interior of a star.[1]

3.1.1 Operations With Vectors

Vectors can be treated as one column matrices, and inherit all of the algebraic properties therein. So, given

$$\vec{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and

$$\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

we have

$$r\vec{u} + s\vec{v} = \begin{pmatrix} ra + sx \\ rb + sy \\ rc + sz \end{pmatrix}$$

[1] You can expect to encounter the cylindrical and spherical coordinate systems. Do not expect to encounter the bipolar, toroidal, parabolic cylindrical, paraboloidal, elliptic cylindrical, prolate spheroidal or oblate spheroidal coordinates at any time. I had to research them just to write this footnote.

3.2 Basis Vectors: The Building Blocks

The world becomes much easier to deal with if we can express our vectors not only in standard coordinate bases, but as combinations of standard units. These units are called *basis vectors*, and our goal will be to find a set of basis vectors that can be combined to make any possible vector in the vector space. They are also called elementary basis vectors, which is why e has become the standard symbol for them. In three dimensions, it stands to reason that we will need three such vectors to do the job. However, there are conditions that must be met by the three vectors.

The most common choice of basis vectors is also the most intuitive choice:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

With this choice, the above generic vector $\vec{v}$ can be written

$$\vec{v} = x\vec{e}_1 + y\vec{e}_2 + z\vec{e}_3$$

which takes up a lot less vertical space on the page.

This is certainly not the only choice. For example,

$$\vec{e}_a = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \vec{e}_b = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \vec{e}_c = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

is a perfectly valid set of basis vectors. It's cumbersome for two primary reasons:

1. It's a pain to figure out how to represent a given vector as a combination of these three.
2. If you graph the axes defined by these three, you find that they are not at right angles to each other. They still work, but not easily.

To further demonstrate the difficulties in working with this set, examine the vector

$$\vec{w} = \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}$$

Under the normal Cartesian vectors, we can write $\vec{w} = 2\vec{e}_1 + \vec{e}_2 + 5\vec{e}_3$. This is easy to see. However, it's much harder to see the same representation with the new basis: $\vec{w} = 3\vec{e}_a - \frac{3}{2}\vec{e}_b + \frac{1}{2}\vec{e}_c$.

There are some sets that are in worse shape than this one. For example, if we keep the first two vectors in this basis and change the third, we can have:

$$\vec{e}_a = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \vec{e}_b = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \vec{e}_d = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$$

There are vectors which cannot be represented in this basis no matter what we do. That is because they satisfy the following relationship:

$$\vec{e}_a + \vec{e}_b = \vec{e}_d$$

Thus, although we have three different vectors, we only have two *independent* vectors in our basis. There is no possible way to represent the vector $\vec{e}_c$ in the form $\vec{e}_c = r\vec{e}_a + s\vec{e}_b + t\vec{e}_d$. The addition of vector $\vec{e}_d$ does nothing to improve the *span* of the first two vectors. The span of a set of k basis vectors is the amount of n-dimensional space that includes vectors which can be described by any *linear combination* of basis vectors. In other words, the span includes all vectors which satisfy the relation

$$\vec{v} = \sum_{i=1}^{k} r_i \vec{e}_i$$

where the standard summation notation is used. Notice that k and n don't need to be equal to produce a span, but if $k < n$, we cannot span the entire n-dimensional vector space. If $k \geq n$, then we have a chance to span the space, but it's not guaranteed: we can have a set of 56 vectors, but if every one of them has a value of 0 along the z axis, then there's no possible way to represent any vector that has a value other than 0 for that coordinate.

3.3 Transposes

Writing vectors as columns takes up a lot of space on a page. This is why many texts write them as rows, with commas between the components. Though technically improper for reasons we won't see for at least three more sections, it certainly does save space on a page. This is the most obvious application and use of a vector's *transpose*. The transpose $\vec{v}^T$ of vector $\vec{v}$ (or of a matrix) is the version with the same values, but columns and rows have been reversed:

$$\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \vec{v}^T = \begin{pmatrix} x & y & z \end{pmatrix}$$

Similarly, the transpose A^T and the corresponding matrix A would be:

$$A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix}$$

The first row becomes the first column, the second row becomes the second column, and so forth. These will be useful beyond saving page space with vectors: trust me.

3.4 Metrics

The next critical piece of the definition puzzle is the metric g. The metric of a vector space defines the manner in which the basis vectors interact with each other. This is often omitted from discussion at the elementary level because the metric in Cartesian space is the identity matrix introduced in the previous lesson. As it operates through multiplication, its influence will not be noticed in that context. However, there are real world applications of linear algebra in which the metric is not as simple as the identity matrix, and cannot be ignored.

The first example of an abnormal metric comes from relativity. We are used to dealing with the three spatial dimensions. These behave independently, and they do not change over time, so the components of the metric describing how the basis vectors interact with each other are all 0, while the components that describe how the basis vectors interact with themselves are all 1. When we arrange these values such that the component in row i and column j details the interaction between $\vec{e}_i$ and $\vec{e}_j$ we get the following:

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

This, as mentioned above, is none other than the identity matrix. However, the theory of relativity describes time in addition to space. If time is added as a fourth dimension, with basis vector $\vec{e}_0$, then we get a rather shocking change to our metric:

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We have a -1 in the top left corner. This is the mathematical distinction between space and time coordinates, and will be of vital importance for the (optional) mathematical portions of next year's Bureau 42 summer school course on Einstein's relativity. The three by three identity matrix above serves as the metric for Cartesian space, while the four by four version with the negative sign for time serves as the metric for Minkowski space, which is used for special relativity, though not for general relativity.[2] We'll see how to create metrics in our final lesson.

3.5 Inner and Dot Products

There are two types of products that are particularly useful when working with vectors. The first is the inner product, which takes two vectors as input and produces a single, scalar quantity as output. The notation for an inner product of vectors $\vec{u}$ and $\vec{v}$ is $\langle \vec{u}, \vec{v} \rangle$.

A valid inner product satisfies the following properties:

$$\begin{aligned} \langle \vec{u}, \vec{v} \rangle &= \overline{\langle \vec{v}, \vec{u} \rangle} \\ \langle a\vec{u}, \vec{v} \rangle &= a \langle \vec{u}, \vec{v} \rangle \\ \langle \vec{u} + \vec{z}, \vec{v} \rangle &= \langle \vec{u}, \vec{v} \rangle + \langle \vec{z}, \vec{v} \rangle \end{aligned}$$

where $\overline{\langle \vec{v}, \vec{u} \rangle}$ is the complex conjugate of $\langle \vec{v}, \vec{u} \rangle$.[3] If the vectors exist entirely in spatial (non-time) dimensions, we have the additional condition

$$\langle \vec{u}, \vec{v} \rangle \geq 0$$

The most common form of inner product is the dot product. The dot product of vectors $\vec{u}$ and $\vec{v}$ is defined as $\vec{u} \cdot \vec{v} = \vec{u}^{\dagger} g \vec{v}$, where $\vec{v}^{\dagger}$ is the complex conjugate of $\vec{v}^T$. With specific examples in Cartesian space,

$$\vec{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and

$$\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

[2] The metrics for general relativity get wonky, and vary on a case by case basis.

[3] If you aren't familiar with complex conjugates, don't worry. Complex numbers and imaginary numbers involve the square roots of negative numbers. When you reach a point where that's important, you can make a complex conjugate by replacing every instance of i with $-i$, where i will be meaningful in that context. Unless you are an electrical engineer, in which case you work in a field that had a specific meaning for i before the applicability of complex numbers was recognized, so you'll replace j with $-j$ instead.

the dot product is given by

$$\vec{u}\cdot\vec{v} = \vec{u}^{\dagger} g \vec{v} = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = ax + by + cz$$

This is most useful in its interpretations. If the second vector $\vec{v}$ is 1 unit long, then $\vec{u}\cdot\vec{v}$ is the length of the shadow $\vec{u}$ would cast on the axis that runs through $\vec{v}$, assuming the light source produced parallel light that was perpendicular to $\vec{v}$. It is the amount of $\vec{u}$ that can be represented as a multiple of $\vec{v}$. This will be very important when we discuss the Gram-Schmidt process later, as it means that the dot product between perpendicular vectors is 0. If you study physics, you will be bombarded by dot products.

3.6 One-forms

The one-form is the mathematical object that is naturally represented by a row instead of a column. A vector $\vec{v}$ has a corresponding one-form $\tilde{v}$ defined as:

$$\tilde{v} = \vec{v}^{\dagger} g$$

where g is the corresponding metric tensor.

When working in Cartesian space, $\tilde{v} = \vec{v}^{\dagger}$. However, in other spaces, the one-form becomes a distinct beast from the vector, as the anomalies in basis vector interactions will already be reflected in the vector. For example, in Minkowski space, we can have the pair

$$\vec{v} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \tilde{v} = \begin{pmatrix} -ct & x & y & z \end{pmatrix}$$

This means an alternative definition of the dot product can be written

$$\vec{u}\cdot\vec{v} = \tilde{u}\vec{v}$$

3.7 Norms

We are now fully equipped to define the length of a vector. The length of a vector $\vec{v}$ is given by

$$\|\vec{v}\| = \sqrt{|\vec{v}\cdot\vec{v}|}$$

The absolute value signs are in place for cases such as Minkowski space, which will result in a negative dot product from a vector with 0 spatial components and non-zero time component (among other cases.) In the case of Cartesian space, for vector

$$\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

we have

$$\|\vec{v}\| = \sqrt{x^2 + y^2 + z^2}$$

which is the familiar result taught in high school, if it's taught in high school at all. If you go back to the standard basis vectors for Cartesian space above, we find $\|\vec{e}_1\| = \|\vec{e}_2\| = \|\vec{e}_3\| = 1$. Armed with this definition, readers with an unusual combination of time and patience can verify that $|\vec{u} \cdot \vec{v}| = |\|\vec{v}\|\|\vec{u}\| \cos\theta|$, where θ is the angle between vectors $\vec{u}$ and $\vec{v}$.

3.8 Gram-Schmidt

We are finally equipped to start doing actual stuff with the basis vectors defined at the start of this lesson! Hooray!

As mentioned before, although there are a number[4] of choices for a set of basis vectors for a vector space, some are much more convenient than others.

There are two key features that make a set of basis vectors convenient:

1. Orthogonality - A set of basis vectors $\vec{u}_k$ are more likely to be convenient if $\vec{u}_k \cdot \vec{u}_j$ is 0 when $j \neq k$.
2. Normality - A set of basis vectors $\vec{u}_k$ are more likely to be convenient if $\|\vec{u}_k\| = 1$.

Convenience from this point on depends on the individual situation. For example, when dealing with a typical high school "box on a ramp" problem, it is convenient to deviate from the standard x and y axes and instead define axes that are parallel and perpendicular to the surface of the ramp. So, how does one produce a convenient set of basis vectors, particularly when there's only one or two specific vectors that truly matter?

Let's use the case of the box on a ramp problem. For this one, we can boil things down to a two dimensional world; we can pretend the box never slides

[4] An infinite number, actually.

sideways on the ramp. One convenient basis vector is the one that is parallel to the surface of the ramp. By geometry, if the ramp is elevated by an angle θ, then the vector parallel to the surface $\vec{e}_{\parallel}$ is given by

$$\vec{e}_{\parallel} = \cos\theta \vec{e}_1 + \sin\theta \vec{e}_2$$

where $\vec{e}_1$ points in the direction of a level surface and $\vec{e}_2$ is perpendicular to a level surface. We'll make this our first vector of the new basis, since $\|\vec{e}_{\parallel}\| = 1$ already.

We have to now create a second vector to serve as the rest of the basis. We could use geometry again to show that

$$\vec{e}_{\perp} = -\sin\theta \vec{e}_1 + \cos\theta \vec{e}_2$$

but in this case, we'll use the Gram-Schmidt process to show the general means to work. We have an answer to compare to from geometry, and knowing the Gram-Schmidt process is useful for cases where the geometry isn't obvious.[5]

The general idea of Gram-Schmidt is this: start with a vector $\vec{u}_1 = \vec{v}_1$ you know you want in your new basis. Pick a second vector $\vec{v}_2$ that is not a multiple of this vector, meaning there is no a for which $\vec{v}_2 = a\vec{u}_1$. We start with $\vec{v}_2$ and take away the bits that "overlap" $\vec{u}_1$ to make basis vector $\vec{u}_2$. This overlapping portion is the dot product of the two: if they are perpendicular, the dot product is zero, which means the dot product extracts the parallel components. Mathematically, with

$$\vec{u}_1 = \vec{e}_{\parallel} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

and

$$\vec{v}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(assuming $\theta \neq 0$, meaning we don't even have a ramp to worry about) then we

[5] For example, the vectors representing different quantum mechanical states need to be generated this way, as it is difficult to use geometry to plot the components of vectors which exist in dimensions formed by quantum possibilities rather than positions.

can find our second basis vector with

$$\begin{aligned}
\vec{u}_2 &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1 \\
\vec{u}_2 &= \vec{v}_2 - \frac{\left\langle \vec{v}_2, \vec{e}_{\|} \right\rangle}{\left\langle \vec{e}_{\|}, \vec{e}_{\|} \right\rangle} \vec{e}_{\|} \\
\vec{u}_2 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{\cos\theta}{1} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \\
\vec{u}_2 &= \begin{pmatrix} 1 - \cos^2\theta \\ -\sin\theta\cos\theta \end{pmatrix} \\
\vec{u}_2 &= \begin{pmatrix} \sin^2\theta \\ -\sin\theta\cos\theta \end{pmatrix} \\
\vec{u}_2 &= -\sin\theta \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}
\end{aligned}$$

This isn't quite the vector we're looking for. That's because we haven't *normalized* it yet; the length of this vector is $\|\vec{u}_2\| = \sin\theta$, not 1. We can divide by $\pm\sin\theta$ to normalize the vector, and leave us with

$$\vec{e}_{\perp} = \frac{\vec{u}_2}{-\sin\theta} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

If we had three dimensional space, we'd need to continue, finding our third basis vector by taking a vector $\vec{v}_3 \neq a\vec{v}_1 + b\vec{v}_2$ for any a and b, and then calculating

$$\vec{u}_3 = \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1 - \frac{\langle \vec{v}_3, \vec{u}_2 \rangle}{\langle \vec{u}_2, \vec{u}_2 \rangle} \vec{u}_2$$

Generally speaking, you can build a set of orthogonal basis vectors $\vec{u}_k$ out of any linearly independent set of vectors $\vec{v}_k$ by calculating

$$\vec{u}_n = \vec{v}_n - \sum_{i<n} \frac{\langle \vec{v}_n, \vec{u}_i \rangle}{\langle \vec{u}_i, \vec{u}_i \rangle} \vec{u}_i$$

3.9 Cross Products

It has been mentioned that there are two useful products with two vectors. The inner product took two vectors and produced a scalar. The cross product starts with two vectors and produces a third vector.

Before defining the cross product, we must first define the Levi-Civita symbol:

$$\epsilon_{ijk} = \begin{cases} 0 & \text{if } i,j,k \text{ are not three distinct values (i.e. 112 instead of 123)} \\ 1 & \text{for even permutations of } 1,2,3 \\ -1 & \text{for odd permutations of } 1,2,3 \end{cases}$$

Even permutations are those in which the digits 1, 2, and 3 have been traded in pairs an even number of times, while odd permutations are those in which the digits 1, 2, and 3 have been traded in pairs an odd number of times. For those unfamiliar with permutations, the 27 values written explicitly are $\epsilon_{111} = \epsilon_{112} = \epsilon_{113} = \epsilon_{121} = \epsilon_{131} = \epsilon_{211} = \epsilon_{311} = \epsilon_{221} = \epsilon_{212} = \epsilon_{122} = \epsilon_{223} = \epsilon_{232} = \epsilon_{322} = \epsilon_{222} = \epsilon_{331} = \epsilon_{313} = \epsilon_{133} = \epsilon_{332} = \epsilon_{323} = \epsilon_{233} = \epsilon_{333} = 0$, $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, and $\epsilon_{132} = \epsilon_{213} = \epsilon_{321} = -1$.

With this established, we can define the cross product of two vectors $\vec{u}$ and $\vec{v}$ which exist in a three dimensional vector space with basis vectors $\vec{e_i}$ as the following sum:

$$\vec{u} \times \vec{v} = \sum_{i,j,k=1}^{n} \epsilon_{ijk} \vec{e_i} u_j v_k$$

With the standard basis vectors, this is equivalent to

$$\vec{u} \times \vec{v} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}$$

Readers with a unusual combination of time and patience can verify that $\|\vec{u} \times \vec{v}\| = |\|\vec{u}\|\|\vec{v}\| \sin\theta|$ where θ is the angle between the vectors $\vec{u}$ and $\vec{v}$. You can also verify that $\vec{u} \times \vec{v}$ is orthogonal to both $\vec{u}$ and $\vec{v}$. This occurs frequently in physics, particularly when dealing with electricity and magnetism. It is also worth noting that $\vec{u} \times \vec{u} = \vec{0}$, and in fact, $\vec{u} \times \vec{v} = \vec{0}$ any time $\vec{u}$ and $\vec{v}$ are parallel.

3.10 Summary

We have seen vectors defined, and have learned about various operations we can do with them, which feel primarily like “stupid math tricks” at the moment. They all have applications, many of which we’ll see during the course of this summer school and some of which will wait for later years.

Chapter 4

Inverting Matrices to Solve Systems

> The matrix is a system, Neo.
>
> Morpheus, *The Matrix*

4.1 Systems of Linear Equations the Old and Busted Way

Many of us remember the classic type of problem that introduces systems of linear equations. These are the problems along the lines of "Sally is twice as old as her sister Alice was three years ago. If their combined ages today add up to 24 years, how old are Alice and Sally now?" Matrices and linear algebra can be used to solve problems of this type.

The first step is to set the problem up. We start with the final question: "how old are Alice and Sally now?" This defines the variables we will need to solve for. The first variable is Alice's age today. Let's call that a. The second is Sally's age today. Let's call that s.

Now that we have our variables, we dissect the first sentence. "Sally is twice as old as her sister Alice was three years ago." The first word is Sally, so we start with the variable s. The next word is "is," a conjugation of the verb "to be." In any problem, conjugations of this verb represent an equals sign. So, far, we have

$$s =$$

which doesn't get us very far. The next words are "twice as old as," indicating that we need to multiply something by 2:

$$s = 2\times ?$$

The final grammatical object is a noun construct, "her sister Alice was three years ago." Well, that means we multiply by $(a-3)$ to represent Alice's age three years ago:

$$s = 2 \times (a-3)$$

To be useful in this context, we use the distributive property to move the 2 through the brackets on the right hand side, and rearrange the variables to get the following:

$$2a - s = 6$$

This is a simple equation. Sadly, it has two unknowns, so we cannot solve it completely. For that, we use the next sentence: "their combined ages today add up to 24 years." We represent this as

$$a + s = 24$$

Now we have what we need to solve it: two equations with two unknowns. We will solve these equations using the technique called *elimination*: add the two equations (or some multiple of them) together to eliminate one of the variables.

$$\begin{aligned} 2a - s &= 6 \\ +(a + s &= 24) \\ 3a \quad &= 30 \\ a &= 10 \end{aligned}$$

Alice is currently 10 years old. By substituting this into $a + s = 24$, we quickly find that Sally is 14 years old.

4.2 Systems of Linear Equations the New Hotness Way

Chapter 5

Eigenvectors and Eigenvalues

Chapter 6

More Than Meets the Eye

Transformations

Chapter 7

Fun with Idempotents and other Projections

Chapter 8

Operators

Chapter 9

Working With Tensors

Free your mind.
Morpheus, *The Matrix*

9.1 Defining Tensors

To this point, we have worked with matrices, which are structured arrangements of numbers. Working solely with numbers, however, is subject to limitations. In many cases, combining multiple matrices is problematic, and the transformations are not consistent with each other. We need something more versatile for other topics, such as the subject of Bureau 42's Summer School 2012.

A tensor is much like a matrix. However, instead of being a structured arrangement of numbers, a matrix is essentially a structured arrangement of *variables*. This is critically important for some purposes: when a matrix is transformed, one may need to reinterpret its entries, but with a tensor, that reinterpretation is unnecessary.