Linear Algebra and Geometry

by Louis Meunier

notes.louismeunier.net

Contents

Unit 1: The Basics of Linear Algebra

1.1   Solving Linear Systems

1.1.1   Algebraically vs Geometrically

Consider a set of linear equations, $a_1 x + b_1 y = c_1$ and $a_1 x + b_2 y = c_2$. Solving this should be fairly straightforward:

- Isolate one variable in one of the equations

- Substitute said variable in the second equation

- Solve for second variable

- Plug in solution to second variable in either original equation to find the other variable

Doing this reveals one of two things: either the system is consistent, or inconsistent:

- **Consistent**: a solution to the system exists. This can either be:

    - **A unique solution** (Fig. 1)

    - **Infinite solutions** (Fig. 2)

- **Inconsistent**: no solution to the system exists (Fig. 3)

Being able to recognize when these particular circumstances occurs is essential.

Graphically, the differences between these are also very noticeable (see figures 1, 2, and 3.)

(a) Two equations with a unique solution   (b) Three equations with a unique solution
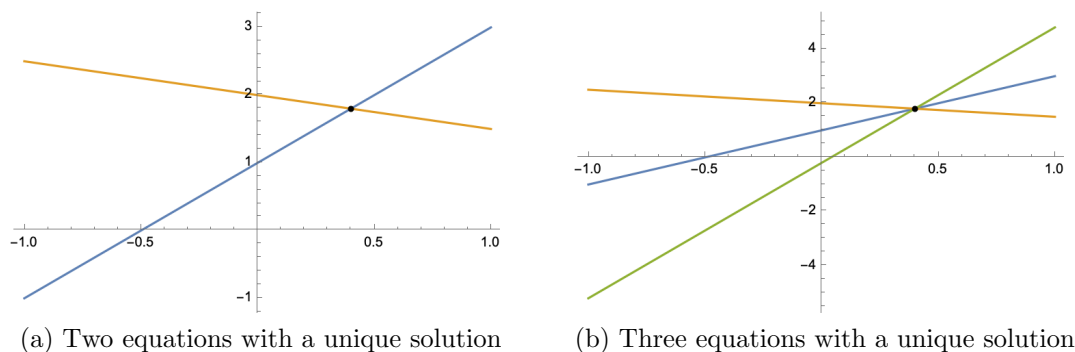
Figure 1: Unique solutions

1.1.2   Gaussian Elimination

Although the procedure described above is fine for smaller equations, it can often become cumbersome. Using the concept of Gaussian Elimination, you can not only more easily solve systems of linear equations, but also determine patterns in said systems more easily.

To begin, create an **augmented matrix** of the system. This involves a **coefficient matrix** on the left, and the augmented part on the right.
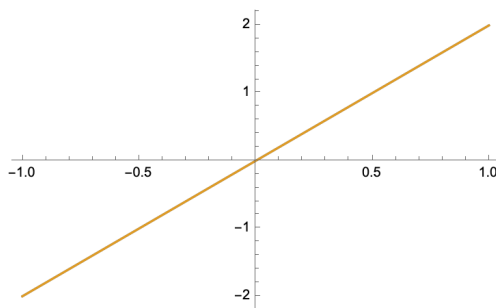
Figure 2: Any number of equations with infinite solutions (always intersecting)

(a) Two equations with no solution   (b) Three equations with no solution

Figure 3: No solutions

The **coefficient matrix** has a row representing each equation, and each column representing the coefficients belonging to each individual variable in the system.

The **augmented matrix** contains each of the corresponding constants that the rows sum up to.

For example, the system

$$
\begin{aligned}
x + 2y - z &= 1 \\
2x - 3y + z &= 4 \\
y + 2z &= 0
\end{aligned}
\tag{1}
$$

becomes

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

Note that even though there is no x in the third equation, its coefficient must still be represented (as a 0).

1.1.3   rref

When working with augmented matrices, there are a number of possible operations that can be used to reach the "end goal": **the row reduced echelon form (rref)**. An rref is achieved when:

- The first non-zero term in each row (aka the **leading 1**) is a 1, and there are zeroes above and below it.

- The leading 1 in any row is to the right of the leading 1 above (*this creates the typical staircase form*).

- Any rows with all zeros in the coefficient matrix are at the bottom.

A (very simple) rref matrix could look like this:

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

To be clear: not all matrices will result in a nice, orderly pattern as this; most won't.

To get to this point at all, you must manipulate the matrix using certain operations:

- Switch rows

- Add a multiple of one row to another

- Multiply a row by a constant

Note that all of these operations work on rows; attempting to manipulate columns will result in a wrong answer.

1.1.4   Determining Solutions from rref

Given any rref matrix, you should be able to tell how many solutions a system has (and, by extension, what they are). The following examples should reveal some patterns in determining the number of solutions for any matrix:

- $$\left[\begin{array}{ccc|c} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 1 & c_3 \end{array}\right]$$

  This system has **one unique solution**: when $x = c_1, y = c_2$ and $z = c_3$. This comes from the fact that each column represents a variable; thus, you can read the first row as $1 * x + 0 * y + 0 * z = c_1$. The same idea applies for the rest of the rows.

- $$\left[\begin{array}{ccc|c} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 0 & c_3 \neq 0 \end{array}\right]$$

This system has **no solutions**: the final row is stating that $0 * x + 0 * y + 0 * z \neq 0$, which is impossible. Thus, no solution to the system exists that would fulfil this requirement.

- $$\left[\begin{array}{ccc|c} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This system has **infinite solutions**: in no row is there any restriction on the value of z. Thus, while $x = c_1$ and $y = c_2$, z can be any real number. When this happens, z is called a **"free variable"**.

In rref matrices of the following form, you can also have variables that are reliant on the value of free variables.

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & c_1 \\ 0 & 1 & 1 & c_2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

In this case, $x = c_1$ and $y = c_2 - z$, where z is free. It should be noted that you could also rewrite this as $z = c_2 - y$, in which case y would be the free variable. Either way of writing it, there is one free variable.

There are many other forms, of varying complexity, that rrefs can have as well. However, following these simple patterns will make it fairly straightforward to determine the solutions in any situation.

1.1.5   Representing Solutions

As above, you can write the solutions to a system as $x = ..., y = ..., z = ...$ However, a more "standard" way, which will also make certain manipulations easier later, is in vector form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Or, if there is a free variable:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ t \end{bmatrix}, t \in \mathbb{R}$$

1.1.6   Rank

The **rank** of a matrix is the number of leading ones in the rref of that matrix. You can similarly say that the rank is equal to the number of columns of that matrix minus the number of free variables (though this way of thinking is more relevant later).

If a matrix has a rank equal to its number of columns, then the system of equations it represents is consistent and has a unique solution.

1.2   Spans

Before proceeding, we need to define a few concepts to analyze solutions and describe vectors and matrices in new ways.

A **linear combination** of the vectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$ is an expression $c_1\vec{v_1} + c_2\vec{v_2} + ... + c_n\vec{v_n}$, where $c_1, c_2, ..., c_n \in \mathbb{R}$. In "simpler" terms, it is a sum of multiples of the vectors involved.

A **span** of vectors is all possible linear combinations of a set of vectors, represented by the notation $span(\vec{v_1}, \vec{v_2}, ..., \vec{v_n})$.

Spans can also be rewritten as a set; $span(\vec{v_1}, \vec{v_2}) = c_1\vec{v_1} + c_2\vec{v_2}, c_1, c_2 \in \mathbb{R}$. Being able to switch between these two forms will be important for later.

Another important concept is being able to determine if a vector exists in a particular span; say, does the vector $\vec{v}$ exist in $span(\vec{x}, \vec{y})$? This is the equivalent of asking whether a particular vector can be reached by summing multiples of a set of other vectors. Based on our "alternative" form of writing a span, we can say:

$$\vec{v} \stackrel{?}{\in} span(\vec{x}, \vec{y}) = c_1\vec{x} + c_2\vec{y}, c_1, c_2 \in \mathbb{R}$$

Thus, we can say $\vec{v} = c_1\vec{x} + c_2\vec{y}$. If and only if c_1 and c_2 exist can we say that $\vec{v}$ is in $span(\vec{x}, \vec{y})$.

1.3   Linear Relations

Piggybacking off the ideas of linear combinations earlier, we can define a **linear relation** of vectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$ as an equation of the form $c_1\vec{v_1} + c_2\vec{v_2} + ... + c_n\vec{v_n} = \vec{0}$.

If you can solve this equation where at least on c_n is $N\ OT$ 0, then this is called a **non-trivial linear relation**. Otherwise, if the only possible solution occurs when all c_n's are 0, then this is a **trivial linear relation**.

From this, we can state yet another definition to describe vectors; *linear (in)dependence*. A set of vectors can be described as **linearly independent** if the only linear relation between them that exists is the trivial one; conversely, a set of vectors are **linearly dependent** if a non-trivial linear relation exists.

Its helpful to think about this logically: if a set of vectors are *dependent*, then one (or more) of the vectors relies on the one (or more) of the others. In more "mathematical" terms, this means that vectors in the set can be created from a sum of different multiples of others vectors in said set. This, on the other hand, is impossible between *independent* vectors.

For example: is the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in $span(\begin{bmatrix} -4 \\ -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix})$? If this were true, then we should be able to say that:

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -4 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, c_1, c_2 \in \mathbb{R}$$

c_1 and c_2 are, clearly, just *scalars* of their respective vectors. As such, we can rewrite:

$$\begin{bmatrix} c_1 \\ c_1 \end{bmatrix} + \begin{bmatrix} -4c_2 \\ -5c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
$$\begin{bmatrix} c_1 - 4c_2 \\ c_1 - 5c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

(2)

It should (hopefully) be clear that this is simply a system of linear equations, where c_1 and c_2 are the unknowns!

$$c_1 - 4c_2 = 1; c_1 - 5c_2 = 2$$

From here, we can simply use Gaussian elimination to see whether or not a solution does exist:

$$\begin{bmatrix} 1 & -4 & | & 1 \\ 1 & -5 & | & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & | & -3 \\ 0 & 1 & | & -1 \end{bmatrix}$$

Therefore, $c_1 = -3$ and $c_2 = -1$; as this system is consistent, then the vector is indeed in the span.

When working with spans of a number of vectors, it is important to be able to rewrite spans with as few vectors as possible; this is called the **basis** of the set. With more notation, we can say that the vectors $v_1, v_2, ..., v_n$ of set V are a *basis* of V if they both *span* V and are *linearly independent*.

The first requirement here should make intuitive sense; if a set of vectors doesn't span V, then of course it can't represent it fully. The second requirement takes care of the "as few vectors as possible" part; if the opposite were true, and the vectors were linearly dependent, then one or more of the vectors could be represented by vectors already existing in the set.

Using the basis, we can then define the **dimension** of a set; the number of vectors in the *basis* of the set.

1.4   Subspaces

A **subspace** of $\mathbb{R}^N$ is a non-empty set of vectors in $\mathbb{R}^N$, that can be described as a span of vectors. These two terms can often be interchanged, but be careful when you do so.

Formally, a subspace, V, must have the following properties:

1. **Closed under scalar multiplication:** if $\vec{u} \in V$, then $k\vec{u} \in V$ for any scalar $k \in \mathbb{R}$

2. **Closed under addition:** if $\vec{v}, \vec{w} \in V$, then $\vec{v} + \vec{w} \in V$.

The concepts of bases and dimension similarly apply to subspaces, with the same rules.

1.5 Standard Basis Vectors

The standard basis vectors of $\mathbb{R}^n$ are written as $\vec{e_i}$, with i from 1 to n, where the vector is made up of all 0's and a 1 in the ith row. For example, the standard basis vectors $\mathbb{R}^2$ are $\vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

These standard basis vectors are, hopefully clearly, linearly independent; that is why they are called the "basis" after all. They will be helpful to use later on.

1.6 Matrix-Vector Multiplication

Being able to multiple a matrix by a vector has a large number of applications that are important to understand. We can approach defining matrix multiplication in a few ways.

1.6.1 Column View

Given a matrix A and a vector $\vec{x}$:

$$A\vec{x} = \begin{bmatrix} | & | & & | \\ \vec{c_1} & \vec{c_2} & ... & \vec{c_n} \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} = x_1\vec{c_1} + x_2\vec{c_2} + ... + x_n\vec{c_n}$$

In other words, this is defining the product of a matrix and vector as a linear combination of the columns of the matrix. It should be noted that this works by treating each column of the matrix A as a vector; this can be helpful in better visualizing this operation.

1.6.2 Row View

Given a matrix A and a vector $\vec{x}$:

$$A\vec{x} = \begin{bmatrix} - & \vec{r_1} & - \\ - & \vec{r_2} & - \\ & ... & \\ - & \vec{r_m} & - \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{r_1} \cdot \vec{x} \\ \vec{r_2} \cdot \vec{x} \\ ... \\ \vec{r_m} \cdot \vec{x} \end{bmatrix}$$

Note that "·" represents a *dot product*: this is the operation involving multiplying each element in one vector by the corresponding element in another vector.

Note that, conversely to above, this way of viewing the operation involves treating each row of A as an individual matrix.

Deciding which "view" to use when is, often, very subjective to the particular situation. However, it should be noted that *the result is the same no matter which view you take.*

1.7 Linear Transformations

A **linear transformation** $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a function that:

1. **preserves vector addition**: $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$ for all $\vec{x}, \vec{y}$

2. **preserves scalar addition**: $T(c\vec{x}) = cT(\vec{x})$ for all $\vec{x}$ and $c \in \mathbb{R}$

Linear transformations are often described by a matrix A, which *induces* said linear transformation. In other terms:

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

This matrix A must be an m x n matrix in order for this formula to make sense.

1.7.1   Determining Linear Transformations

In order to figure out the linear transformation T of any vector, you must know the $T(\vec{x})$ of as many linearly independent $\vec{x}$'s as there are columns in the matrix A. If you know, or are able to determine, the T of the standard basis vectors of A, it is far easier to determine the linear transformation of any vector, thanks to the properties of linear transformations:

$$T(\vec{x}) = T(c_1\vec{e_1} + c_2\vec{e_2} + ...) = c_1T(\vec{e_1}) + c_2T(\vec{e_2}) + ...$$

Using this same pattern, we can rewrite the T of any vector $\vec{x}$ as a sum of multiples of the T of the standard basis vectors.

1.7.2   Geometric Linear Transformations

Linear transformations can be visualized as a number of geometrical transformations; representing these in two dimensions makes it easiest to comprehend, but the same concepts are applicable to an arbitrary number of dimensions.

Say $T(\vec{x}) = A\vec{x}$, consider the following transformations on the standard basis vectors $\vec{e_1}$ and $\vec{e_2}$ of $\mathbb{R}^2$ (see figure 4):

Figure 4: $\vec{e_1}, \vec{e_2}$ before any transformation

- $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

$$T(\vec{e_1}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, T(\vec{e_2}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
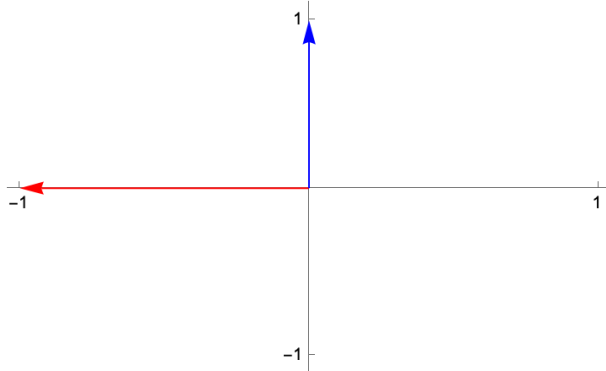
Figure 5: Reflection of $\vec{e_1}, \vec{e_2}$ over the y-axis

- $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$$T(\vec{e_1}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, T(\vec{e_2}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
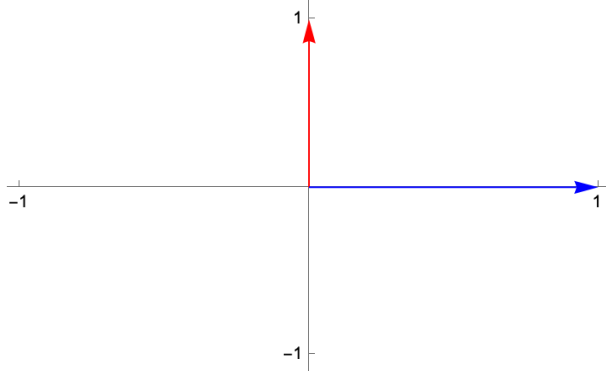
Figure 6: Reflection of $\vec{e_1}, \vec{e_2}$ over $y = x$

- $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$

$$T(\vec{e_1}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, T(\vec{e_2}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
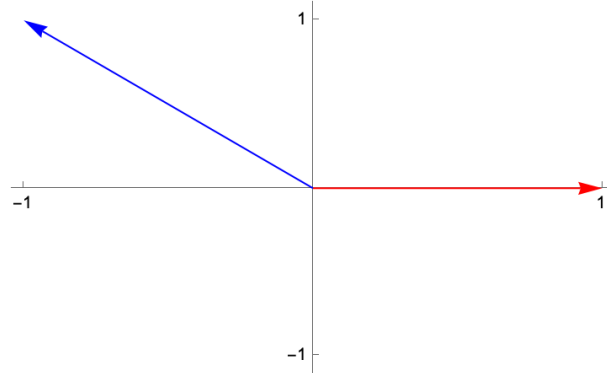
Figure 7: Shear of $\vec{e_1}, \vec{e_2}$ towards left

- $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

$$T(\vec{e_1}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, T(\vec{e_2}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$
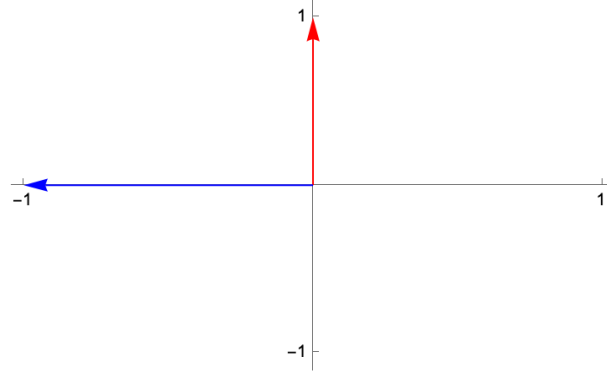
Figure 8: Rotation counter-clockwise of $\vec{e_1}, \vec{e_2}$ about the origin

In these transformations, finding the effect on the standard basis vectors is typically quite simple, while finding the transformation for any random vector can be quite difficult. So, we can instead use the rules of linear transformations to find the effect on any vector in terms of its ratio to the standard basis vectors.

1.7.3 A Quick Look at Projections

If a plane P passes through the origin in $\mathbb{R}^n$, any vector $\vec{x} \in \mathbb{R}^n$ can be written as $\vec{x}^{\parallel} + \vec{x}^{\perp}$, where $\vec{x}^{\parallel}$ is a vector *in the plane P* and $\vec{x}^{\perp}$ is a vector *perpendicular* to P. This can be defined as a linear transformation $\text{proj}_P : \mathbb{R}^n \to \mathbb{R}^n = \text{proj}_P(\vec{x}) = \vec{x}^{\parallel}$

Unit 2: Transformations & Matrix Properties

2.1 Images and Kernels

For a matrix A, we can define the following (useful) subsets:

- **kernel:** aka null space, written $\ker(A)$, is the set of vectors $\vec{x}$ where $A\vec{x} = \vec{0}$.

- **image:** written $\text{im}(A)$, the set of vectors $\vec{y}$ where $\vec{y} = A\vec{x}$. This is more commonly known as the "range" for functions.

You can solve for $\ker(A)$ by simply augmenting A by $\vec{0}$. The image of A is simply the span of the columns of A (which is fairly simple to prove, so it won't be shown here).

2.1.1 Rank-Nullity Theorem

A useful pattern to keep in mind is that the rank of an rref matrix is equal to the dimension of the image, and the number of free columns in the rref matrix is equal to the dimension of the kernel.

This pattern leads to what is called the **Rank-Nullity Theorem**.

2.2 Matrix Algebra

Similarly to processes described earlier for performing algebraic operations between vectors and vectors, as well as vectors and matrices, it is also possible to perform a number of operations between matrices. However, there are some very specific limitations and rules that must be followed that can make said operations rather unintuitive if viewed as purely extensions of the operations performed between numbers, for instance.

- **Matrix times scalar**

  Multiplying a matrix A times a scalar k is very similar to multiplying a vector times a scalar:

$$
\begin{aligned}
kA &= k \begin{bmatrix} A_{1a} & A_{2a} \\ A_{1b} & A_{2b} \end{bmatrix} \\
&= \begin{bmatrix} kA_{1a} & kA_{2a} \\ kA_{1b} & kA_{2b} \end{bmatrix}
\end{aligned}
\tag{3}
$$

- **Matrix plus matrix**

  Adding together two matrices (say A, a n x m, and B, a p x q) requires certain restrictions on the dimensions of the affected matrices, just as in vector addition. In this case, $n = p$ and $m = q$:

$$A + B = \begin{bmatrix} A_{1a} & A_{2a} & A_{3a} \\ A_{1b} & A_{2b} & A_{3b} \end{bmatrix} + \begin{bmatrix} B_{1a} & B_{2a} & B_{3a} \\ B_{1b} & B_{2b} & B_{3b} \end{bmatrix}$$
$$= \begin{bmatrix} A_{1a} + B_{1a} & A_{2a} + B_{2a} & A_{3a} + B_{3a} \\ A_{1b} + B_{1b} & A_{2b} + B_{2b} & A_{3b} + B_{3b} \end{bmatrix} \tag{4}$$

- **Matrix times matrix**

  In order to multiply two matrices (say A, a n x m, and B, a p x q), another set of restrictions apply, that may be a little trickier to interpret. Say you're finding AB; here, $m = p$, and the resulting matrix is an n x q matrix.

  When multiplying two matrices, you can imagine the first matrix being multiplied by each of the columns of the second matrix, treating each column as a vector. The result of each vector-matrix multiplication becomes the corresponding column of the resulting matrix. Given these steps, the above restrictions should hopefully make more intuitive sense.

  It should also be clear from these steps that **multiplying matrices is not commutative;**

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -2 & 1 \\ 3 & 1 & 0 \end{bmatrix}$$
$$B = \begin{bmatrix} -1 & 3 \\ 2 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \vec{b_1} & \vec{b_2} \end{bmatrix}$$
$$AB = \begin{bmatrix} A\vec{b_1} & A\vec{b_2} \end{bmatrix} \tag{5}$$
$$= \begin{bmatrix} (-1*2 + 2*0 + 1*2) & (3*1 + 0*0 + 1*2) \\ \dots & \dots \\ \dots & \dots \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 5 \\ -5 & 7 \\ -2 & 9 \end{bmatrix}$$

  Doing BA, on the other hand, would result in a 2 x 3 matrix, which, naturally, would not be equal to the product of AB.

- **Identity Matrix**

  While not exactly an algebraic operation, its important to note that the notation I is often used to denote the **identity matrix**, a matrix such that $AI = A$. I is essentially a rref matrix with the same number of leading ones as columns. Example:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- **Solving for a matrix given its product**

  If you are given a matrix A and the product of A with an unknown matrix B, you can find B fairly easily using a similar idea as discussed above; by treating each column of B as a vector.

  $$AB = \begin{bmatrix} -1 & 0 \\ 4 & 5 \\ 0 & 1 \end{bmatrix} B = \begin{bmatrix} -1 & -1 & 0 \\ 9 & -1 & -10 \\ 1 & -1 & -2 \end{bmatrix} \tag{6}$$

  Treating each column of B as a vector ($\vec{b_1}$, etc), we can write the product of each column of B equal to the corresponding column of the product of A and B:

  $$\vec{b_1}A = \begin{bmatrix} -1 \\ 9 \\ 1 \end{bmatrix}, ..., \vec{b_n}A \tag{7}$$

  Finding the vector that multiplies to a matrix to produce a given vector has already been discussed above, so the steps from here (creating an augmented matrix and solving for each component of $\vec{b_n}$) should be clear. However, we can make the work a little easier by simply creating an augmented matrix of A augmented by the product of AB:

  $$\left[\begin{array}{cc|ccc} -1 & 0 & -1 & -1 & 0 \\ 4 & 5 & 9 & -1 & -10 \\ 0 & 1 & 1 & -1 & -2 \end{array}\right] \tag{8}$$

  All row operations to the left hand side of this matrix would be the same no matter the augmented side, so this technique should (hopefully) seem intuitively correct.

2.3 Compositions

Recall that we can describe a linear transformation $T(\vec{x})$ as a matrix A times the vector $\vec{x}$. However, what if we want to perform two (or more!) linear transformations, one after another? Rather than simply performing said linear transformations one after another, we can use the concept of **compositions**.

For instance, say we want T composed with S. This can be written as follows, given $T(\vec{x}) = A\vec{x}$ and $S(\vec{x}) = B\vec{x}$:

$$\begin{aligned} T \circ S &= T(S(\vec{x})) \\ &= T(B\vec{x}) \\ &= A(B\vec{x}) \end{aligned} \tag{9}$$

This composition, just like the functions is composes, is linear. This can be proved using aforementioned definitions, but should also make intuitive sense.

Another common question is how to find the single matrix, say C, that induces the composition. As shown above, a composition is simply a vector multiplied by a number of matrices, so we can find C by simply multiplying the matrices involved (in this case, A and B) using the method described earlier.

Note that, since this is essentially just matrix-matrix multiplication, **order matters**.

2.3.1   Transformation Inverses

Given a particular transformation $\vec{y} = T(\vec{x})$, we know we can describe it as $\vec{y} = A\vec{x}$. However, it is very helpful to be able to describe its inverse, $T^{-1}(\vec{x}) = S(\vec{y}) = \vec{x}$. By a natural extension, we should be able to find a matrix B such that $B\vec{y} = \vec{x}$. Below will describe strategies for finding B for a number of transformations, and later how to generalize said strategies to any matrix A.

First, let's look at a number of common linear transformations, and how to determine their inverses; if possible.

- **Reflections**

$$T(\vec{x}) = ref_L(\vec{x})$$

How do we reflect a vector back over to where it started? Well, as the question almost answers by itself, you simply reflect it again. As such, we can say $T^{-1} = T$.

- **Rotations**

$$T(\vec{x}) = rot(\vec{x})$$

Say our transformation rotates a vector $\vec{x}$ by an angle Θ clockwise. To invert this, we can simply rotate it counterclockwise by Θ, or, equivalently, clockwise by $-\Theta$.

- **Projections**

$$T(\vec{x}) = proj_L(\vec{x})$$

How can we invert a projection? Simply put: we can't! When a vector $\vec{x}$ is projected onto a line L, for instance, we lose information about the original $\vec{x}$, and thus have no way of returning to $\vec{x}$. Another way to interpret this is that $T(\vec{x})$ for any $\vec{x} \perp L$ equals $\vec{0}$, and thus it is impossible to go backwards to find any individual $\vec{x}$.

From these few examples, we can generalize the following: a transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is only invertible if for all *outputs* $\vec{y} \in \mathbb{R}^n$ there exist a $\vec{x} \in \mathbb{R}^m$ such that $T(\vec{x}) = \vec{y}$ and $\vec{x}$ is unique.

We can think about this a little differently by considering A such that $T(\vec{x}) = A\vec{x} = \vec{y}$. Given our definition above, $A\vec{x} = \vec{y}$ must be consistent for any $\vec{y}$ if T is invertible. As such, we can say that $\text{rref}(A)$ must have a leading one in each *row*.

Further, we also require that $A\vec{x} = \vec{y}$ has a unique solution, so $\text{rref}(A)$ must also have a leading one in each *column*.

The only way that this is possible, to begin with, is that the number of columns of A equals to the number of rows of A, and by extension, $\text{rref}(A) = I$. In other words, the matrix must be square.

2.3.2   Inverting a Matrix

As seen earlier in Unit 1, we can say that a transformation $T(\vec{x}) = A\vec{x} = \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) \end{bmatrix} \vec{x}$. To find the inverse of A (A^{-1}), we can say that $A^{-1} = \begin{bmatrix} T^{-1}(\vec{e_1}) & T^{-1}(\vec{e_2}) \end{bmatrix}$.

We can say that $\vec{c_n} = T^{-1}(\vec{e_n})$, so $T(\vec{c_n}) = \vec{e_n}$. Thus, we can say that each nth column of matrix A times some $\vec{c_n}$ equals $\vec{e_n}$. To find each $\vec{c_n}$ (ie, find each column of A^{-1}), we can therefore augment A by the corresponding number of unit vectors; or, in other words, I_n.

For example, to invert $\begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}$, we write:

$$
\begin{aligned}
& \begin{bmatrix} 3 & 2 & | & 1 & 0 \\ 7 & 5 & | & 0 & 1 \end{bmatrix} \\
\rightarrow\ & \begin{bmatrix} 1 & 0 & | & \vec{c_1} & \vec{c_2} \\ 0 & 1 & | & | & | \end{bmatrix} \\
\rightarrow\ & \begin{bmatrix} I & | & A^{-1} \end{bmatrix}
\end{aligned} \tag{10}
$$

2.4   Bases and Coordinates

2.4.1   Introduction

To understand the use of bases in linear algebra, consider the standard basis vectors, $\vec{e_1}, \vec{e_2}, ..., \vec{e_n}$. Any vectors in R^n can thus be rewritten as a combination of these standard vectors ($c_1\vec{e_1} + c_2\vec{e_2} + ... + c_n\vec{e_n}$). However, consider a situation in which, say, "movements" are restricted to only a particular set of vectors (say $\vec{v_1}$ and $\vec{v_2}$) that aren't the standard basis vectors.

While you can write these vectors, or any combinations of these vectors, as a combination of standard basis vectors, it can be easier to call these vectors as a new set of "basis vectors". We define this as $\mathcal{B} = \{\vec{v_1}, \vec{v_2}\}$, ie $\mathcal{B}$ is a basis of $\mathbb{R}^2$.

When writing about vectors in basis $\mathcal{B}$, we use the notation $[y]_\mathcal{B} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ to indicate the coordinates of $\vec{y}$ in the $\mathcal{B}$ basis. This vector can thus, by extension, be represented as a linear transformation of the vectors in the basis $\mathcal{B}$.

2.4.2 Changing Bases

To change a vector from one base to another, we can find a matrix S such that $S[\vec{x}]_{\mathfrak{B}} = \vec{x}$, where S is typically called the "change of basis matrix", in this case, changing base from $\mathfrak{B}$ to the standard basis.

Finding S is fairly straightforward. We know that S times each vector in $\mathfrak{B}$ must equal the corresponding standard basis vector, so we can write $S\vec{v_1} = \vec{e_1}$, $S\vec{v_2} = \vec{e_2}$, and so on. Therefore, S simply equals the matrix of the vectors in $\mathfrak{B}$, ie $S = \begin{bmatrix} | & | & | \\ \vec{v_1} & \vec{v_2} & ... \\ | & | & | \end{bmatrix}$.

To extend this, since we can say that $S[\vec{x}]_{\mathfrak{B}} = \vec{x}$, we can inversely say that $[\vec{x}_{\mathfrak{B}}] = S^{-1}\vec{x}$, so to find the matrix that changes bases in the opposite direction, we simply find the inverse of S. From here, this same rationale can be extended transform between any two bases in the same $\mathbb{R}^n$.

2.4.3 Application to Transformations

As noted earlier, we can define a linear transformation as a matrix A such that $T(\vec{x}) = A\vec{x}$. With the added knowledge of bases, we can often more easily find the matrix of a transformation by carefully picking a basis for the transformation which makes the matrix easier to find.

To demonstrate this, take a transformation T that projects vectors in $\mathbb{R}^3$ onto the plane P defined by $3x_1 + 2x_2 - 5x_3 = 0$. To find a matrix A such that $T(\vec{x}) = A\vec{x}$, we can create a basis $\mathfrak{B}$ of R^3 using vectors that make A easy to compute.

For this particular case, we can define $\mathfrak{B} = \{\vec{v_1}, \vec{v_2}, \vec{v_3}\}$, where:

- $\vec{v_1} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, since $T(\vec{v_1}) = 0$, as it is perpendicular to P

- $\vec{v_2} = \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}$, since $T(\vec{v_2}) = \vec{v_2}$, as it lies along P

- $\vec{v_3} = \begin{bmatrix} 5 \\ 0 \\ 3 \end{bmatrix}$, since $T(\vec{v_3}) = \vec{v_3}$, as it also lies along P

To put it in more clear notation, we say that $[T(\vec{v_1})]_{\mathfrak{B}} = \vec{0}$, $[T(\vec{v_2})]_{\mathfrak{B}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $[T(\vec{v_3})]_{\mathfrak{B}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$; this step should make sense, as each vector denoted is the corresponding combination of vectors in $\mathfrak{B}$. Here, it should be clear that any vectors that that form a base of $\mathbb{R}^3$ could realistically have been used here, but these are clearly easier to work with.

From here, we can define B, such that $B[\vec{v_1}]_{\mathfrak{B}} = [T(\vec{v_1})]_{\mathfrak{B}}$, simply enough as $B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

From the previous section, we showed that $S^{-1}\vec{x} = [\vec{x}]_{\mathfrak{B}}$. Extending that to this linear transformation, we can recap our steps above using a number of matrices, as follows:

$$\begin{aligned}
T(\vec{x}) &= A\vec{x} \\
[T(\vec{x})]_{\mathfrak{B}} &= B[\vec{x}]_{\mathfrak{B}} \\
[T(\vec{x})]_{\mathfrak{B}} &= B{\color{red}S^{-1}}\vec{x} \\
T(\vec{x}) &= {\color{red}S}BS^{-1}\vec{x} \\
\implies A &= SBS^{-1}
\end{aligned} \tag{11}$$

The final line above indicates a general process that can be used to find a matrix to represent any linear transformation, and while this process may seem initially more difficult that previously discussed methods, it has many advantages as well.

2.5   Elementary Matrices

Throughout these notes, the idea of "Gaussian elimination" and row reduction has been constantly discussed, with the concept of modifying the rows and columns of a matrix through a number of operations. **Elementary matrices** define matrices that, when multiplied to another matrix, execute these same row operations. These are typically denoted with an E.

2.5.1   Types of Elementary Matrices

Just as there a set number of possible row operations that we can perform on matrices, there are a corresponding number of elementary matrices that we can define to perform the equivalent transformations.

I : Row Interchanges

This type of elementary matrix simply swaps rows of a matrix. For example, for a 2 by 3 matrix:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ p & q & r \end{bmatrix} = \begin{bmatrix} p & q & r \\ a & b & c \end{bmatrix} \tag{12}$$

The elementary matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ swaps the first and second rows, as shown. It should also be clear that there could be *any number of columns in the matrix on the left*, and the elementary matrix would still perform the same operation.

Finding the matrix to exchange rows for any dimension of matrix is fairly straightforward, as the matrix will only involve 0's and 1's in different positions.

If we were to invert this operation, we can simply use the same elementary matrix again. Thus, in this case, $E = E^{-1}$; this can be verified by manually inverting E as well, by using methods described earlier.

II : Row Multiplication

To multiply a particular row (or rows) of a matrix by a constant, we can define another fairly simply elementary matrix that involves essentially all 1's and 0's again, except for the appropriate constant in the appropriate position. For instance, say we want to multiply the second row of the same matrix as above by 3:

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} a & b & c \\ p & q & r \end{bmatrix} = \begin{bmatrix} a & b & c \\ 3p & 3q & 3r \end{bmatrix} \tag{13}$$

To invert this operation, we have to divide the appropriate row by the same constant, or, more appropriately, multiply it by the inverse of the constant. Thus, $E^{-1} = E$ but with the constant inverted (fractional).

III : Adding Multiples of Rows

Adding a certain number of a row (or rows) to another row (or rows) of a matrix is, often, the most complicated to describe with an elementary matrix. However, with an example, it should hopefully be rather clear; so, to add twice the second row to the first row of a matrix, consider the following:

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ p & q & r \end{bmatrix} = \begin{bmatrix} a+2p & b+2q & c+2r \\ p & q & r \end{bmatrix} \tag{14}$$

To invert this operation, we have to subtract the constant times the same row from the row involved. In this particular case, $E^{-1} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$.

It should be noted that, in all these operations, the elementary matrix was invertible; and not only that, the inverted elementary matrix was also an elementary matrix in its own regard. If we want to invert such a matrix, we can either do so with typical methods (augmenting it by I) or by simply finding the logical inverse of the operation that the elementary matrix performs.

It should be noted that you will often see problems with the following format, for some matrix A:

$$E_5 E_4 E_3 E_2 E_1 A \tag{15}$$

While this may seem scary, since it is a whole lot of matrix-matrix multiplication, once you are able to determine what operation each E_n performs, calculating this should be fairly straightforward. Simply find A after E_1's operations on performed on it, then after E_2 is performed on the result of that, etc..

2.6   Distances

Finding the distance between arbitrary figures in a particular space is very important in a number of applications. This section will cover the general trend for finding the distance between points,

lines, and planes in $\mathbb{R}^n$.
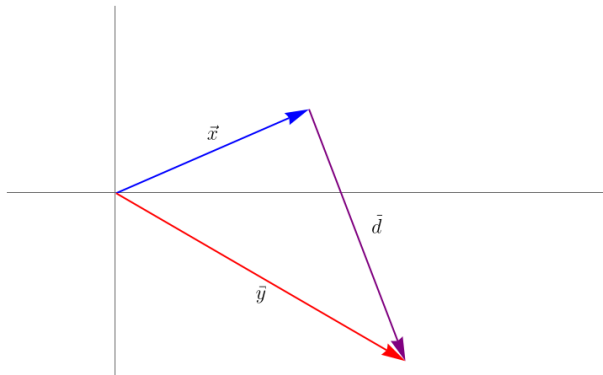
2.6.1 Point to Point

Figure 9: Vector-vector distance

Given two points, we can consider the point as the tip of a vector in the respective base (see figure 9). From here, the distance between the two vectors is clear: it is simply the length of the vector between the two points. If we consider two vectors $\vec{x}$ and $\vec{y}$, then we can write the distance vector $\vec{d}$ as:

$$\vec{x} + \vec{d} = \vec{y} \implies \vec{d} = \vec{y} - \vec{x}$$

From here, we simply find the length of $\vec{d}$, denoted as $\|\vec{d}\|$, using the Pythagorean theorem (and its higher-dimension extensions...):

$$\|\vec{d}\| = \sqrt{d_1^2 + d_2^2 + ... + d_n^2}$$

2.6.2 Point to Line

Finding the distance between a point and a line (specifically, the shortest distance; this is really the only helpful one that should be found, as any point can be considered an infinite distance from a particular point on a line...) requires a little more intuitive and computation (though it can be simplified later). First, to understand how to find this shortest distance, we have to consider the properties of a particular point on the line in question that creates this shortest distance with our given point. Perhaps this is intuitive, but this point is the point on the line that forms a distance vector that is perpendicular to the line.

From the figure, it should be clear that if $\vec{d}$ is drawn anywhere else along the line L, it will be longer than if drawn perpendicular. You can visualize this as $\vec{d'}$ forming the hypotenuse of a triangle, with another side being $\vec{d_\perp}$; the hypotenuse is always the longest side of a triangle, and thus, $\vec{d'}$ would be longer than $\vec{d_\perp}$.

As discussed in earlier sections, the projection of a vector perpendicular onto a particular line (or any space really, but a line in this case) is simply $\vec{0}$. As such, we can determine the vector $\vec{d}$

(and from here, naturally, finding its distance) by writing the linear transformation $T(\vec{x}) = proj_L\vec{x}$, finding the matrix A that induces this transformation, and finding $\vec{d}$.

$$\vec{d} \perp L \implies T(\vec{d}) = \vec{0}, T(\vec{x}) = proj_L(\vec{x})$$
$$T(\vec{x}) = M\vec{x} \therefore M\vec{d} = \vec{0} \tag{16}$$

Note that a very similar method can be used to find the distance between a point and a line that doesn't pass through the origin, by adding the distance from the line to the origin when computing the transformation.

2.6.3   Point to Plane

Finding the distance between a point and a plane is conceptually similar to above, but much harder to visualize. The general methods are as follows, given a plane P, that doesn't go through the origin, and a point Q (ie the tip of a vector $\vec{Q}$):

- Pick any point R on the plane. Using this, we can define $\vec{Q} = \vec{R} + \vec{x}$, where $\vec{R}$ describes the vector with a point at R. Note that the important vector here is $\vec{x}(= \vec{Q} - \vec{R})$.

- Define the distance vector $\vec{d}$ between Q and P, which we know is perpendicular to P.

- **The projection of $\vec{x}$ onto $\vec{d}$ is $\vec{d}$, since they share a point at their tips and both touch the plane.**

- By extension, the projection of $\vec{x}$ onto any vector perpendicular to P is $\vec{d}$, so we can use any such vector (say, $\vec{n}$) for the projection:

$$proj_{\vec{n}}\vec{x} = \vec{d}$$

For a plane of form $ax + by + cz = d$, we can use the shortcut of $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ begin perpendicular to P.

- To find this projection, we can simply use the standard formula $(proj_{\vec{b}}\vec{a} = (\frac{\vec{a}\cdot\vec{b}}{||\vec{b}||^2})\vec{b})$.

2.6.4   Further Applications

Beyond these distances, other further combinations can be found by applying similar ideas.

- **Line-Plane:**   use any point on the line, and use the same method as above for the distance between a point and a plane.

- **Plane-Plane:**   use any point on the plane, and use the same method as above for the distance between a point and a plane.

- **Line-Line:** find a distance vector $\vec{d}$ such that $\vec{d}$ is perpendicular to both lines (this will be the shortest vector between the two). You can either do so by using standard algebra and Gaussian elimination to find a vector that is perpendicular to both lines, or by using the **cross product** of the two lines.

  A cross product of two vectors $\vec{a}, \vec{b}$ is denoted as $\vec{a} \times \vec{b}$, where the result is the vector perpendicular to the two. The formula for the cross product of vectors $\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ in $\mathbb{R}^3$ is:

$$\vec{a} \times \vec{b} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix} \tag{17}$$

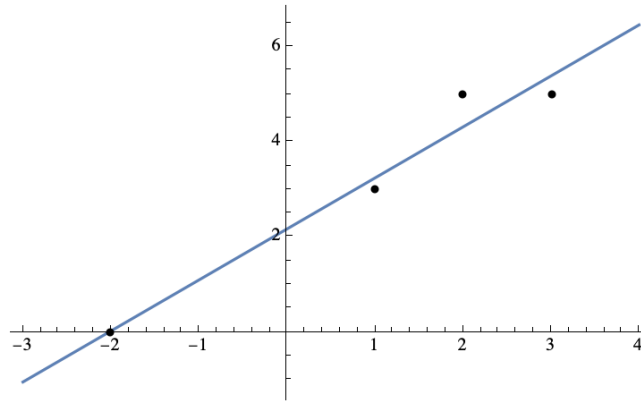# Unit 3: Applications of Matrices & More

3.1   Curve Fitting

Figure 10: Curve fitting example

If you want to see the calculus approach to curve fitting, see here.

One application (of many) of linear algebra, and specifically, the ability to find distances between different elements in any particular base is in *curve fitting*. The process is a little abstract to describe generally, so first consider the following set of points (and corresponding graph, figure 10):

$$(x, y) = \{(-2, 0), (1, 3), (2, 5), (3, 5)\}$$

It seems as if these points generally follow a linear trend, so we can try to find a line that fits the points as well as possible (*note that a similar method to the following can be used for a polynomial trend, an exponential trend, etc, but with a little more work*). A line can be described by the equation $y = mx + b$, so our goal is to find the values of m and b that minimize the distance between the line and the points. We can write, for each individual point:

$$
\begin{aligned}
y &= mx + b \\
0 &= m(-2) + b \\
3 &= m(1) + b \\
5 &= m(2) + b \\
5 &= m(3) + b
\end{aligned}
\tag{18}
$$

It should (hopefully) be clear that this is a system of equations for the variables m and b. We can, as described earlier, consider this as a matrix equation of $A\vec{x} = \vec{y}$, where A is the matrix of coefficients, $\vec{x}$ is the vector of variables, and $\vec{y}$ is the vector of y values. We can write this as:

$$
\begin{bmatrix} 0 \\ 3 \\ 5 \\ 5 \end{bmatrix} = m \begin{bmatrix} -2 \\ 1 \\ 2 \\ 3 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}
$$

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

$$
\vec{y} = \vec{x} A
$$

(19)

The last line here may seem a little bizarre, but should make more sense using the column-view of a matrix-vector multiplication. We could solve this system of equations using Gaussian elimination as always, but it should be clear that it will be inconsistent. Intuitively, there is clearly no line that goes through all the points, and as such, there is no m and b that makes this equation true. You can also conclude this by looking at the rows of the A matrix, which, if reduced, will not be consistent.

As such, we have to find the m and b that make this equation "as close to true" as possible. We define a new vector $\vec{z}$, where $\vec{z} = A\vec{x}$. As described in an earlier section, we can find the shortest distance between two vectors by finding the projection of one onto the other, and we can then use this said projection to solve the system of equations. This is a fine method, but is very tedious and time-consuming, particularly for systems of multiple hundreds of data points.

For a non-linear trend (say, $y = ax^2 + bx + c$), we can use the same method as above, but with a slightly different A matrix, naturally.

3.2   Orthogonality and the Transpose of a Vector

Given a subspace V of $\mathbb{R}^n$, we can define the *orthogonal complement* of V as the set of all vectors $\vec{u}$ such that $\vec{u} \cdot \vec{v} = 0$ for all $\vec{v} \in V$. This is denoted $V^\perp$. Finding this is fairly simple; for instance, take $V = \{\vec{v_1}, \vec{v_2}\} = \{ \begin{bmatrix} -2 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \}$. To find $V^\perp$, we simply find the set of all vectors perpendicular to $\vec{v_1}$ **AND** $\vec{v_2}$. Let's define $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \vec{x} \in V\perp$, so $\vec{v_1} \cdot \vec{x} = \vec{v_2} \cdot \vec{x} = 0$. Solving from here is fairly straightforward:

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

(20)

Notice: this is the equivalent of multiplying the matrix $\begin{bmatrix} -2 & 2 & 3 & 0 \\ -1 & 1 & 1 & 1 \end{bmatrix}$ by $\vec{x}$. Since we are

trying to find $\vec{x}$ when this product is 0, we can see this instead as finding the kernel of this new matrix (ie $ker(\begin{bmatrix} -2 & 2 & 3 & 0 \\ -1 & 1 & 1 & 1 \end{bmatrix})$).

Now, let us define A as the matrix with columns equal to the vectors in V (ie $A = \begin{bmatrix} \vec{v_1} \vec{v_2} \end{bmatrix}$). As previously discussed, the *image* of a matrix is the span of its columns, so we can say that $im(A) = span(\vec{v_1}, \vec{v_2}) = V$.

To find how this work so far is useful, we need to introduce a new concept, namely the **transpose** of a matrix. Denoted, in our example, A^T, the transpose of a matrix is the matrix with rows equal to the columns of the original matrix:

$$A^T = \begin{bmatrix} -2 & 2 & 3 & 0 \\ -1 & 1 & 1 & 1 \end{bmatrix} \tag{21}$$

Please note that this is not the same as simply "flipping" the matrix on its side. In actuality, you can think about this as the first element of the first row of A becoming the first element of the first column of A^T, and so on.

Also notice that this matrix is equal to the matrix we found earlier when trying to find the perpendicular vectors to $\vec{v_1}$ and $\vec{v_2}$. So, continuing from our earlier logic, we can thus define the vectors perpendicular to V:

$$\vec{x} \in V^{\perp}$$
$$\therefore \vec{x} \in ker(\begin{bmatrix} -2 & 2 & 3 & 0 \\ -1 & 1 & 1 & 1 \end{bmatrix}) = ker(A^T) \tag{22}$$

Putting all of this together, we can develop the following relationship between a matrix's transpose and orthogonal component:

$$V^{\perp} = ker(A^T)$$
$$V = im(A) \implies V^{\perp} = (im(A))^{\perp} \tag{23}$$
$$\therefore (im(A))^{\perp} = ker(A^T)$$

Working through the whole proof of this can be a little tedious, but it is very useful to understand. We can also slightly rewrite this conclusion to another useful form:

$$(im(A))^{\perp} = ker(A^T)$$
$$((im(A))^{\perp})^{\perp} = (ker(A^T))^{\perp}$$
$$im(A) = (ker(A^T))^{\perp} \tag{24}$$
$$im(A^T) = (ker(A^{T^T}))^{\perp}$$
$$im(A^T) = ker(A)^{\perp}$$

3.3   Normal Equations

While we showed a general form to use least squares to calculate the best fit line for an equation earlier, it would have resulted in having to find a projection, which, in higher dimensions, would prove unnecessarily tedious. Here, we build upon previous concepts to show a more efficient method.

Consider a set of data that, following the steps in section 3.1, we have defined as $A\vec{x} = \vec{b}$. Assuming this system is inconsistent (which would thus merit using least squares), our goal is to find a vector $\vec{y}$ that is as close to $\vec{b}$ as possible. We know we can do this using projections, and thus write:

$$A\vec{x} = \vec{y} = proj_{im(A)}(\vec{b}) \tag{25}$$

We can more thoroughly define $\vec{b}$ as well. $\vec{b}$ can be thought of as being the sum of $\vec{y}$ in one direction, and some other vector perpendicular to $\vec{y}$. We've established that $\vec{y} \in im(A)$, so this perpendicular vector is in $A^\perp$. Even more specifically, we can write:

$$\vec{b} = proj_{im(A)}(\vec{b}) + proj_{im(A)^\perp}(\vec{b})$$
$$proj_{im(A)}(\vec{b}) = \vec{b} - proj_{im(A)^\perp}(\vec{b}) \tag{26}$$

Substituting into equation 25, and doing some (admittedly unintuitive) algebra:

$$A\vec{x} = \vec{b} - proj_{im(A)^\perp}(\vec{b})$$
$$A^T(A\vec{x}) = A^T(\vec{b} - proj_{im(A)^\perp}(\vec{b}))$$
$$A^T A\vec{x} = A^T\vec{b} - A^T proj_{im(A)^\perp}(\vec{b}) \tag{27}$$

Recall that $(im(A))^\perp = ker(A^T)$, so we can write:

$$A^T A\vec{x} = A^T\vec{b} - A^T proj_{im(A)^\perp}(\vec{b})$$
$$A^T A\vec{x} = A^T\vec{b} - A^T proj_{ker(A^T)}(\vec{b}) \tag{28}$$

However, in this final line, $\vec{b}$ is being projected onto the kernel of A^T, resulting in a vector in the kernel of A^T, which, when then multiplied by A^T itself, simply results in $\vec{0}$. Thus, we can write our final form:

$$A^T A\vec{x} = A^T\vec{b} \tag{29}$$

In all, the goal is to find $\vec{x}$, which now only requires a couple of matrix operations, rather than a whole mess of projections. This equation is known as the **normal equation**.

3.4   The Determinant

The **determinant** of a matrix is a scalar value that can be used to describe many properties of a matrix, denoted either as $det(A)$ or $|A|$. In particular, it can be used to describe whether a matrix is invertible without needing to actually invert it; by extension, a determinant only exists for a square matrix. The formula for the determinant of a 2×2 matrix can be derived through trying to invert it (by augmenting it with the identity matrix):

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

$$...$$
$$\left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - \frac{bc}{a} & -\frac{1}{a} & 1 \end{array}\right]$$

(30)

Without even finishing the elimination, it should be clear that the term $d - \frac{bc}{a}$ (in red) cannot equal 0 in order for the matrix to be invertible. We can rewrite this term as $ad - bc$; this is the determinant of a 2×2 matrix.

3.4.1   Determinants of Larger Matrices

The determinant of matrices larger than 2×2 could be derived similarly, but the work is tedious. Instead, you can break down the matrix into smaller matrices, and use the determinant of the smaller matrices to find the determinant of the larger matrix. For example, consider the following 3×3 matrix, $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$.

To solve for the determinant of A, we can pick any particular row or column, and break the matrix into smaller matrices based on it. For example, we can pick the first row, and break the matrix into 2×2 matrices:

$$det(A) = a * det(\begin{bmatrix} e & f \\ h & i \end{bmatrix}) - b * det(\begin{bmatrix} d & f \\ g & i \end{bmatrix}) + c * det(\begin{bmatrix} d & e \\ g & h \end{bmatrix})$$

(31)

Note that each "submatrix" used in this case is the matrix of A without the respective row or column that the coefficient of our row in question is in. Also note that we multiplied the second term by -1. In general, we have to multiply by the corresponding sign in a "sign matrix", in which the top left corner is positive, and the sign alternates from there.

We can finish solving this using our determinant formula of 2×2 matrices:

$$det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$$

Theoretically, you could use this formula as a general formula for any 3×3 matrix, but, in practice, it is far easier to simply use the algorithm described above for the particular case.
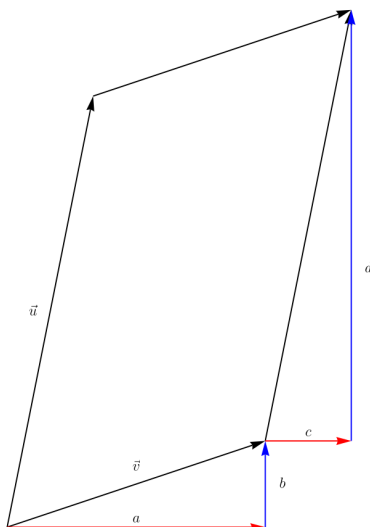
Figure 11: A parallelogram defined by vectors $\vec{v}$ and $\vec{u}$

For determinants of higher order matrices, the same general formula applies, but the resulting "submatrices" have to then be broken down more.

3.4.2   Properties of Determinants

It is helpful to remember a few properties of determinants of different matrices to make certain problems easier. For matrices A and B:

- $det(A) = det(A^T)$

- $det(A) = -det(A)$, if a row of A is exchanged

- $det(kA) = k^n det(A)$, where n is the order of A

- $det(AB) = det(A)det(B)$

- $det(A) + det(B) \neq det(A + B)$

3.5   Applications to Area and Volume

3.5.1   Area

Take a parallelogram who's sides are defined by the length of vectors $\vec{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{u} = \begin{bmatrix} c \\ d \end{bmatrix}$. You can also think of this as the base being defined as $\vec{v}$, then being translated upwards by $\vec{u}$; see figure 11.

We can see this image as the area of a rectangle removed by the area of 2 triangles on either side of the parallelogram:

$$
\begin{aligned}
A &= (a+b)(c+d) - 2(\tfrac{1}{2}(c(a+b))) + 2(\tfrac{1}{2}(b(c+d))) \\
A &= (a+b)(c+d) - ca - cb - bc - bd \\
A &= ac + ad + bc + bd - ca - bc - bc - bd \\
&= ad - bc = \left| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right|
\end{aligned}
\tag{32}
$$

Note that this is simply the determinant of the matrix defined by the vectors $\vec{v}$ and $\vec{u}$ (though you may have to take the absolute value, depending on the context).

3.5.2   Volume

Consider a parallelepiped, whose sides are defined by the vectors $\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\vec{u} = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} g \\ h \\ i \end{bmatrix}$.

The volume of such a figure can be found with:

$$
A = \vec{u} \cdot (\vec{v} \times \vec{w})
$$

To derive this expression, first consider the base, defined by the vectors $\vec{v}$ and $\vec{w}$. The area of this base is the height times the matrix, which, with a little trigonometry, can be shown to equal $||\vec{v}||\,||\vec{w}|| \sin \theta$, with θ representing the angle between the two vectors. This, as a matter of fact, is simply another way to define cross product of $\vec{v}$ and $\vec{w}$, namely $\vec{v} \times \vec{w} = A_{base}$.

Now, to find the volume we next need the height. We can find this by picturing a right triangle with a hypotenuse defined by $\vec{u}$, and the height in question actually being a scalar multiple of the cross product of $\vec{v}$ and $\vec{w}$. With a little more trigonometry, it is clear that this value is simply equal to $||\vec{u}|| \cos \alpha$, with α representing the angle between $\vec{u}$ and $\vec{v}$.

Multiplying the height by the area of the base, we get the volume of the parallelepiped:

$$
\begin{aligned}
A &= ||\vec{u}|| \cos \alpha * (\vec{v} \times \vec{w}) \\
&= \vec{u} \cdot (\vec{v} \times \vec{w})
\end{aligned}
$$

3.6   Adjugates

To define adjugates of a matrix, the idea of a *cofactor* must first be defined. The cofactor of a matrix is the matrix of the determinants of a matrix's submatrices, with the signs of the submatrices alternating. For example, take $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The cofactors of A can be found as follows (*note the similarity between the work here and the method used to find the determinant of a large matrix*):

$$c_{1,1} = det(\begin{bmatrix} d \end{bmatrix}) = d$$
$$c_{1,2} = det(\begin{bmatrix} c \end{bmatrix}) = -c$$
$$c_{2,1} = det(\begin{bmatrix} b \end{bmatrix}) = -b$$
$$c_{2,2} = det(\begin{bmatrix} a \end{bmatrix}) = a$$
$$\begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{bmatrix} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

Notice the similarity between this and the formula for computing the inverse of A:

$$A^{-1} = \frac{1}{det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{33}$$

The matrix of cofactors we derived is simply the transpose of the matrix in this equation, which is called the **adjugate** of A, denoted $\mathrm{adj}(A)$. We can thus say the following:

$$A^{-1} = \frac{1}{det(A)} adj(A) \tag{34}$$

Note that this formula is, clearly, undefined if $det(A) = 0$; this should make sense, since, by definition of the determinant, the determinant of a matrix is 0 if and only if the matrix is singular.

We can also prove this formula for the general case of an $n \times n$ matrix (for those interested...).

Proof. Consider the inverse matrix A^{-1}, with columns $\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}$, ie

$$A^{-1} = \begin{bmatrix} | & | & & | \\ \vec{x_1} & \vec{x_2} & \cdots & \vec{x_n} \\ | & | & & | \end{bmatrix}$$

We can describe the jth row of A^{-1} by $\vec{x_j} = A^{-1}\vec{e_j}$, which we can rewrite $A\vec{x_j} = A(A^{-1}\vec{e_j}) = I_n\vec{e_j} = \vec{e_j}$. Using Cramer's Rule (explained below), we can find the ith entry of $\vec{x_j}$:

$$x_{i,j} = \frac{det(A_i)}{det(A)}$$

where A_i is the matrix A with the ith column replaced by $\vec{e_j}$. Finding the determinant of A_i by expanding upon the ith row, we see that $det(A_i)$ is simply the ith entry of the cofactor matrix of A, ie $det(A_i) = c_{i,j}$. Thus,

$$x_{i,j} = \frac{c_{i,j}}{det(A)}$$

We can now write the jth column of A^{-1} as

$$\vec{x_j} = \frac{1}{det(A)} \begin{bmatrix} c_{1,j} \\ c_{2,j} \\ \vdots \\ c_{n,j} \end{bmatrix}$$

which is simply the jth column of the adjugate matrix of A, adj(A). Thus, we can say that

$$A^{-1} = \begin{bmatrix} | & | & & | \\ \vec{x_1} & \vec{x_2} & \dots & \vec{x_n} \\ | & | & & | \end{bmatrix} = \frac{1}{det(A)} adj(A)$$

■

Also note that we can rewrite this expression as:

$$\begin{aligned} {\color{red} A * det(A) A^{-1}} &= {\color{red} A * det(A)} \frac{1}{det(A)} adj(A) \\ det(A) &= A * adj(A) \end{aligned} \tag{35}$$

3.7   Cramer's Rule

Explained here.

In summary: we can find the solution of the ith variable (ie, x_i) in a system of n equations of n variables of the form $ax_1 + bx_2 + ... + zx_n = c$ with the formula:

$$x_i = \frac{det(A_i)}{det(A)} \tag{36}$$

Where A is the matrix of coefficients, and A_i is the matrix of coefficients with the ith column replaced by the vector of solutions of the respective equations.

3.8   Linear Discrete Dynamical Systems

Breaking down the definition of a linear discrete dynamical system:

- *linear:* output of the system is a linear combination of the inputs (recall linear transformations)

- *discrete:* the system is defined over a non-continuous range (ie, integers, $\mathbb{Z}$)

- *dynamical:* the vectors in the system change with time

These types of systems are often used to describe changes in several populations which depend on one another (whether inversely or directly).

For example, consider a system of two populations, x and y, where x is a predator of y. Say we can model the change in each population as follows:

$$\begin{cases} x(t+1) = 0.5x(t) + 0.2y(t) \\ y(t+1) = -0.3x(t) + 1.5y(t) \end{cases} \tag{37}$$

We can interpret this system as follows:

- The population of x in a particular year $(t+1)$ depends on its population in the previous year (t) and the population of y in the previous year (t); in particular, the population of x decreases by 50% and increases by 20% for each y in the previous year.

- The population of y in a particular year $(t+1)$ depends on its population in the previous year (t) and the population of x in the previous year (t); in particular, the population of y decreases by 30% and increases by 150% for each y in the previous year.

This changes in population could be due to predation (clearly, as x is a predator of y), reproduction/survival rates, or any other factor.

This system can be represented as a matrix times a vector; specifically, a matrix of the coefficients of the system, and a vector of the populations in the previous year. If we define $\vec{P}(t)$ as the population vector in year $t+1$, we can write the system as:

$$\vec{P}(t+1) = \begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.2 \\ -0.3 & 1.5 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = A\vec{P}(t) \tag{38}$$

*This matrix is often called a **Leslie matrix**, when used in population dynamics.*

This form of equation, while helpful when $P(t)$ is known for a particular t, but does not make it clear how P changes over time. However, seeing as how this equation is recursively defined, we can say the following:

$$\begin{aligned} \vec{P}(t) &= A\vec{P}(t-1) \\ &= A(A\vec{P}(t-2)) \\ &= A^2\vec{P}(t-2) \\ &= A^3\vec{P}(t-3) \\ &= \dots \\ &= A^t\vec{P}(0) \end{aligned} \tag{39}$$

This has somewhat generalized our equation, but we still need the initial population in order

to find the population at any given time. The following section will show how we can modify this equation accordingly...

3.9   Eigenstuff

3.9.1   Application to Linear Discrete Dynamical Systems

An **eigenvector** of a matrix A is a vector $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \lambda\vec{v}$, where λ is a scalar. In other words, the eigenvector is a vector that is simply scaled by some scalar λ when multiplied by the matrix A. This scalar is the corresponding **eigenvalue** of the eigenvector.

Reconsider the previous section, where we said that $\vec{P}(t) = A^t\vec{P}(0)$. Lets say we are given two linearly independent eigenvectors of A *(which, as we'll say later, is the max number of eigenvectors A can have since it is 2×2)*, $\vec{v_1}$ and $\vec{v_2}$, and their corresponding eigenvalues λ_1 and λ_2. WE know that:

$$A\vec{v_1} = \lambda_1\vec{v_1}$$
$$A\vec{v_2} = \lambda_2\vec{v_2}$$

We can multiply each of these by A again (t times...):

$$\textcolor{red}{A}A\vec{v_1} = \textcolor{red}{A}\lambda_1\vec{v_1}$$
$$A^2\vec{v_1} = \lambda_1 A\vec{v_1}$$
$$= \lambda_1^2\vec{v_1}$$
$$\dots$$
$$A^t\vec{v_1} = \lambda_1^t\vec{v_1}; \text{similarly: } A^t\vec{v_2} = \lambda_2^t\vec{v_2}$$

Let's say we are given $\vec{P}(t_0) = \begin{bmatrix} a \\ b \end{bmatrix}$ for some time; we can consider this as the initial population, $\vec{P}(0)$, and write:

$$\vec{P}(t) = A^t\vec{P}(0) = A^t \begin{bmatrix} a \\ b \end{bmatrix}$$

Since $\vec{v_1}, \vec{v_2}$ are two linearly independent vectors (which is generally true of eigenvectors, but will be explained more later...) in $\mathbb{R}^2$, we can therefore say:

$$\begin{bmatrix} a \\ b \end{bmatrix} = c_1\vec{v_1} + c_2\vec{v_2}; c_1, c_2 \in \mathbb{R}$$

Substituting this into the previous equation for $\vec{P}(0)$, we get:

$$\vec{P}(t) = A^t(c_1\vec{v_1} + c_2\vec{v_2})$$

We can distribute the A^t:

$$\vec{P}(t) = c_1 A^t\vec{v_1} + c_2 A^t\vec{v_2}$$

However, since $\vec{v_1}, \vec{v_2}$ are eigenvectors of A, we established that $A^t\vec{v_1} = \lambda_1^t\vec{v_1}$ and $A^t\vec{v_2} = \lambda_2^t\vec{v_2}$, so we can substitute these into the previous equation:

$$\vec{P}(t) = c_1\lambda_1^t\vec{v_1} + c_2\lambda_2^t\vec{v_2}$$

This is a far more usable form of the equation, since, if we can find the eigenvalues and eigenvectors of A, we can find the population at any given time, given the population at some time and the corresponding matrix A. This form of equation is called the **closed-form** for $\vec{P}(t)$.

3.9.2   Computing Eigenvalues, Eigenvectors

Recall how we defined an eigenvector $\vec{v}$ of a matrix A:

$$A\vec{v} = \lambda\vec{v}$$

Moving all the terms to the same side, we get:

$$\lambda\vec{v} - A\vec{v} = \vec{0}$$
$$\vec{v}(\lambda - A) = \vec{0}$$

This, however, is not currently possible to solve: A is a matrix and λ is a constant, so we cannot subtract one from the other. However, we can multiply both sides of the equation by I, which will leave A the same (and 0, of course):

$$I * \vec{v}(\lambda - A) = \vec{0} * I$$
$$\vec{v}(\lambda I - A) = \vec{0}$$

Notice that $\lambda I - A$ is a matrix itself now. Since $\vec{v}$ multiplied by this new matrix is equal to $\vec{0}$, then $\vec{v}$ must be in the kernel of $\lambda I - A$. As such, in order for $\vec{v}$ to even have a kernel, then $\lambda I - A$ must be singular (not invertible), and by extension, its determinant must equal 0. Therefore, we can write:

$$\det(\lambda I - A) = 0$$

Solving this formula typically results in a polynomial in terms of λ equal to 0; this is called the **characteristic polynomial** (denoted $c_A(\lambda)$) of A. The roots of this polynomial are the eigenvalues of A. Plugging these values into the original equation for $\vec{v}$, we can find each corresponding eigenvector of A.

Taking the example from the previous section, we can find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 0.5 & 0.2 \\ -0.3 & 1.5 \end{bmatrix}$ as follows (*using our formula for the determinant of a 2×2 matrix*):

$$0 = det(\lambda I - A) = det(\lambda I - \begin{bmatrix} 0.5 & 0.2 \\ -0.3 & 1.5 \end{bmatrix})$$
$$= det(\begin{bmatrix} \lambda - 0.5 & -0.2 \\ 0.3 & \lambda - 1.5 \end{bmatrix})$$
$$= (\lambda - 0.5)(\lambda - 1.5) - 0.3(-0.2)$$
$$= \lambda^2 - 2\lambda + 0.81$$
$$\dots$$
$$\lambda = 0.564, 1.436$$

We could then plug these values into the original formula to find the corresponding eigenvectors.

3.9.3 Algebraic Multiplicity

Recall the characteristic polynomial of matrix A, $c_A(\lambda) = |\lambda I - A|$. Sometimes, this polynomial will have multiple of the same roots, ie, multiple of the same eigenvalue; this will arise if the polynomial is, say, in the form $(\lambda_1 - a)^n$ where $n \geq 2$. When this occurs, we can say that this particular eigenvalue has **algebraic multiplicity** n.

3.9.4 Eigenspaces and Bases

An **eigenspace** is the set of vectors that are eigenvectors of a matrix A with a particular eigenvalue λ, denoted E_λ.

For instance, take the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Solving for the eigenvalues:

$$
\begin{aligned}
0 &= |\lambda I - A| \\
&= |\begin{bmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{bmatrix}| \\
&= (\lambda - 1)^2 = 0 \\
\lambda &= 1
\end{aligned}
$$

Note that the eigenvalue of this matrix has algebraic multiplicity 2. Finding the corresponding eigenvector(s?):

$$
\begin{aligned}
ker(\lambda I - A) &= ker(\begin{bmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{bmatrix}) \\
&= ker(\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}) \\
&= \{\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} t \\ 0 \end{bmatrix}, t \in \mathbb{R}\}
\end{aligned}
$$

As such, we can say that the eigenspace of A with eigenvalue $\lambda = 1$ is this same set, ie, $E_1 = span(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$. In other words, the eigenspace for a matrix A's particular eigenvalue λ has an eigenspace $E_\lambda = ker(\lambda I - A)$.

Note that, in this situation, there exists only one eigenvector, and as such, it is impossible to span R^2 (which this matrix is in) with only eigenvectors. If this were possible, however, then we would have found a **eigenbasis** for R^2, as the eigenvectors would span the entire space. This is what happened in the earlier example: the eigenvectors spanned the entire space, and as such, we found an eigenbasis, and were able to solve the discrete dynamical system by rewriting the

equation in terms of the eigenvectors. It is not possible to use the same method if an eigenbasis does not exist.

Another relevant term to know is the **geometric multiplicity** of a particular eigenvalue, which corresponds to the dimension of its eigenspace (ie, $dim(ker(\lambda I - A))$). In this example, the geometric multiplicity of $\lambda = 1$ is 1, as the eigenspace is a line. Also note the following inequality (which is too difficult to prove in this course):

$$1 \le \text{geom. multiplicity of } \lambda_i \le \text{alg. multiplicity of } \lambda_i \tag{40}$$

3.10   Diagonalization

A diagonalized matrix A (such that $T(\vec{x}) = A\vec{x}$) is one that can be written as $[T(\vec{x})]_{\mathfrak{B}} = B[\vec{x}]_{\mathfrak{B}}$

where B has the form $\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$, where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A. By exten-

sion, **a matrix is diagonalizable if and only if it has a eigenbasis.** Since a matrix can only have an eigenbasis if each of its eigenvectors has an algebraic multiplicity of 1, we can say that a matrix is diagonalizable if and only if each of its eigenvalues has algebraic multiplicity of 1 (note that there are many other ways to state this same property...).

For instance, take some matrix A, with eigenvalues $\lambda_1 = 3$, $\lambda_2 = 2$, and $\lambda_3 = -2$, and corresponding eigenvectors $\vec{v_1} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$, $\vec{v_2} = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}$, and $\vec{v_3} = \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix}$, respectively. We can take $\mathfrak{B} = \{\vec{v_1}, \vec{v_2}, \vec{v_3}\}$, a basis of R^3 spanned by the eigenvectors. We can then solve for the B matrix of this situation:

$$T(\vec{x}) = A\vec{x}$$
$$T(\vec{v_1}) = A * \vec{v_1} = \lambda_1 \vec{v_1} = 3\vec{v_1}$$
$$T(\vec{v_2}) = A * \vec{v_2} = \lambda_2 \vec{v_2} = 2\vec{v_2}$$
$$T(\vec{v_3}) = A * \vec{v_3} = \lambda_3 \vec{v_3} = -2\vec{v_3}$$
$$\therefore B = \left[[T(\vec{v_1})]_{\mathfrak{B}}, [T(\vec{v_2})]_{\mathfrak{B}}, [T(\vec{v_2})]_{\mathfrak{B}} \right]$$
$$= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Note that this matrix B has the eigenvalues of A on its diagonal, and thus, is diagonalizable.

To phrase this in a more general sense, a matrix A is diagonalizable if you can write $A = SBS^{-1}$, where S is an invertible matrix and B is a diagonal matrix. In this form, we can rewrite A^n in a closed form as follows:

$$
\begin{aligned}
A^n &= (SBS^{-1})^n \\
&= SBS^{-1}SBS^{-1}\ldots SBS^{-1} \\
&= SB^nS^{-1}
\end{aligned}
$$

This form is far easier to compute than computing n products of A, since B^n, being a diagonal matrix, is simply:

$$
\begin{aligned}
B^n &= \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}^n \\
&= \begin{bmatrix} \lambda_1^n & 0 & \ldots & 0 \\ 0 & \lambda_2^n & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n^n \end{bmatrix}
\end{aligned}
$$

Note that, since we've written $A^n = SB^nS^{-1}$, we can say that A^n is diagonalizable if A is diagonalizable. It also has the same eigenvectors as A, since S does not change, and it has the same eigenvalues as A but raised to the power of n.

In the context of a linear discrete dynamical system, where $\vec{x}(t) = A^t\vec{x}(0)$, we can derive the general closed form expression in a different way. We want $\vec{x}(0) = c_1\vec{v_1} + \cdots + c_n\vec{v_n}$, where $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ are the eigenvectors of A. As such, we can say $\vec{x}(0) = S \begin{bmatrix} c_1 \\ c_2 \\ c_n \end{bmatrix}_{\mathfrak{B}}$, where $\mathfrak{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$. We can then say:

$$\begin{aligned}
\vec{x}(t) &= A^t \vec{x}(0) \\
&= (SBS^{-1})^t \vec{x}(0) \\
&= \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_n} \end{bmatrix} \begin{bmatrix} \lambda_1^t & & \\ & \ddots & \\ & & \lambda_n^t \end{bmatrix} S^{-1} \vec{x}(0) \\
&= \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_n} \end{bmatrix} \begin{bmatrix} \lambda_1^t & & \\ & \ddots & \\ & & \lambda_n^t \end{bmatrix} S^{-1} S \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\
&= \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_n} \end{bmatrix} \begin{bmatrix} c_1\lambda_1^t & & \\ & \ddots & \\ & & c_n\lambda_n^t \end{bmatrix} \\
&= c_1\lambda_1^t \vec{v_1} + \cdots + c_n\lambda_n^t \vec{v_n}
\end{aligned}$$

This is the same form as derived before for $\vec{x}(t)$.

3.10.1 Properties of Diagonalizable Matrices

- If A is diagonalizable, so is A^n for any $n \in \mathbb{N}$. A^n also has the same eigenvectors as A and the same eigenvalues as A but raised to the power of n.

- If A is diagonalizable and invertible, then A^{-1} is also diagonalizable. Proof:

$$\begin{aligned}
A &= SDS^{-1} \\
A^{-1} &= (SDS^{-1})^{-1} \\
&= S^{-1}D^{-1}S
\end{aligned}$$

Thus, we can write A^{-1} in the form required for it to be diagonalizable. We can also see that A^{-1} has the same eigenvalues as A but raised to the power of -1.

- If A and B are diagonalizable matrices of the same dimension, **their sum is not necessarily diagonalizable**. Counterexample:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$
$$A + B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} (\neq SBS^{-1})$$

- If A and B are diagonalizable matrices of the same dimension, **their product is not necessarily** diagonalizable. Counterexample:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} (\neq SBS^{-1})$$

3.11   Extensions of Eigenvalues, etc

Take a general 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Solving for the eigenvalues of A:

$$\begin{aligned} 0 = |\lambda I - A| &= |\begin{bmatrix} \lambda - a & -b \\ -c & \lambda - d \end{bmatrix}| \\ &= \lambda^2 - (a+d)\lambda + (ad - bc) \end{aligned}$$

Note that the constant value of the characteristic polynomial of A is simply the determinant of A, $ad - bc$.

The term in from of λ, $a + d$, is the sum of the main diagonal of A which is known as the **trace** of A (denoted $tr(A)$).

Now, say that the eigenvalues of A are α and β; thus, the $C_A(\lambda)$ can also be written:

$$(\lambda - \alpha)(\lambda - \beta) = 0$$
$$\lambda^2 - (\alpha + \beta)\lambda + \alpha\beta = 0$$

Thus, we can write that $a + d = \alpha + \beta$ and $ad - bc = \alpha\beta$, and thus, the trace of a matrix is the sum of its eigenvalues, and its determinant is the product of its eigenvalues.

This pattern also applies to higher order matrices (but won't be proven here...). In summary, for a matrix A of dimension n:

$$tr(A) = \sum_{i}^{n} \lambda_i$$
$$det(A) = \prod_{i}^{n} \lambda_i$$

Note that eigenvalues of multiplicity ≥ 2, the eigenvalues must included that many times in the sum/product.

3.12   More on Dynamical Systems

3.12.1   Stability

Definition: a dynamical system $\vec{x}(t+1) = A\vec{x}(t)$ is considered *stable* if A is a diagonalizable matrix and all solutions $\vec{x}(t)$ approach $\vec{0}$ as $t \to \infty$, regardless of the initial $\vec{x}0$.

Take a system defined as $\vec{x}(t) = c_1(\frac{1}{2})^t \begin{bmatrix} 3 \\ -1 \end{bmatrix} + c_2(\frac{1}{4})^t \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, where $\vec{x}(0) = c_1 \begin{bmatrix} 3 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Intuitively, it should be clear that as $t \to \infty$, $\vec{x}(t)$ approaches $\vec{0}$, since for both $(\frac{1}{2})^t$ and $(\frac{1}{4})^t$, as t increases, vectors they are being multiplied to approach $\vec{0}$. We can thus say that a linear dynamical system is stable if the eigenvalues of A are all less than 1 in magnitude (ie $|\lambda_i| < 1$). Since the eigenvalues do not depend on the initial condition of the system, we can say that the system is stable.

Consider also the **phase portrait** of this system. A phase portrait is a graph of the system's solutions over time, which essentially treats $\vec{x}(t)$ as a parametric plot of the x and y coordinates of the vector. The phase portrait of the system is shown below:

Summary of Terms

This course relies **heavily** on vocabulary. Below is a brief conglomeration (ignore the oxymoron) of these terms.

4.1   Notation

- $\mathbb{R}$: real numbers

- $\mathbb{Z}$: integers

- $\mathbb{N}$: natural numbers

- $\mathbb{R}^n$: the nth real dimension

- $\{...\}$: set

- $\in$: "is in"

- $\exists$: "exists"

- $\therefore$: "therefore"

- $\vec{u}$: vector named "u"

- $\vec{0}$: "0 vector", vector of any dimension with all 0's

- $\vec{e}_i$: "unit vector", vector of any dimension with all 0's except for the ith element, which is 1

- $f : A \to B$: "a function f takes inputs of dimension A and returns outputs of dimensions B"

- $T \circ S$: "T composed with S"

- $\mathfrak{B}, \mathfrak{C}, ...$: a basis of some $\mathbb{R}^n$

- $[\vec{x}]_{\mathfrak{B}}$ a vector $\vec{x}$ represented in base $\mathfrak{B}$

- E: some elementary matrix

- A^{-1}: the inverse of a matrix A

- $A^{\perp}$: the orthogonal complement of a matrix A

- A^T: the transpose of a matrix A

- $\begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix}$: a diagonal matrix with the 0's omitted

- $tr(A)$: the trace of a matrix A

- $ker(A)$: the kernel of a matrix A

- $im(A)$: the image of a matrix A

- $det(A)$: the determinant of a matrix A

- $|A|$: the determinant of a matrix A

- λ_i: the ith eigenvalue of some matrix

- E_λ: the eigenvector with eigenvalue of λ

- $C_A(\lambda)$: the characteristic polynomial of a matrix A

4.2   Vocabulary

- **rref**:

    - "reduced row echelon form"

    - a matrix (representing a system of linear equations) that has been manipulated to have the following properties:

        * the first non-zero in each row is a *leading 1*, with 0's above and below

        * the leading 1 in any row is to the right and below the leading 1 above it

        * rows with all 0's are at the bottom (*this is purely convention*)

- **rank**: the number of leading 1's in the *rref* of a matrix

- **linear combination**: a sum of multiples of a number of vectors.

    eg, $c_1\vec{v_1} + c_2\vec{v_2}$ would be a linear combination of $\vec{v_1}$ and $\vec{v_2}$.

- **span**: all possible *linear combinations* of some vectors, denoted as $\text{span}(\vec{v_1}, \vec{v_2})$

- **linear relation**: an equation of the form $c_1\vec{v_1}+c_2\vec{v_2}+...+c_n\vec{v_n} = \vec{0}$ for the vectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$.

    - if the only way this relation holds true are when all $c_n = 0$, then this is called a **trivial linear relation**

    - if at least one $c_n \neq 0$, then this is called a **nontrivial linear relation**

- **linear (in)dependence**:

    - a set of vectors are **linearly independent** if the only *linear relation* between them is the *trivial* one

    - a set of vectors are **linearly dependent** if there exists a *linear relation* between them that is *nontrivial*

- **basis**: the minimal set of vectors in a subspace needed to span that subspace. By extension, it can be said that a set of vectors x is the basis of a subspace V if:

    - *x spans V*

    - the vectors in x are *linearly independent*

- **dimension**: the number of vectors in a *basis* of a *span*

- **subspace**: a non-empty set of vectors in $\mathbb{R}^n$ that can be denoted as a span of vectors. A subset V of $\mathbb{R}^n$ can be qualified as a subspace if:

    - the subset is *closed under scalar multiplication*: for any $\vec{u} \in V$, $k\vec{u} \in V$ for any scalar $k \in \mathbb{R}$

    - the subset is *closed under addition*: for any $\vec{u}, \vec{w} \in V$, $\vec{u} + \vec{w} \in V$

- **standard basis**: the set of unit vectors forming a basis of $\mathbb{R}^n$. These are denoted by $\vec{e_i}$, with i ranging from 1 to n.

  For example, in $\mathbb{R}^n$, the standard basis vectors are $\vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- **matrix multiplication**: the process of multiplying a vector $\vec{x}$ by a matrix A. There are two ways of visualizing this process:

  – *column view*:

  $$A\vec{x} = \begin{bmatrix} | & | & & | \\ \vec{c_1} & \vec{c_2} & ... & \vec{c_n} \\ | & | & & | \end{bmatrix} \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} = x_1\vec{c_1} + x_2\vec{c_2} + ... + x_n\vec{c_n}$$

  In this case, each column of A is treated as a vector, and multiplied by each component of $\vec{x}$.

  – *row view*:

  $$A\vec{x} = \begin{bmatrix} - & \vec{r_1} & - \\ - & \vec{r_2} & - \\ & ... & \\ - & \vec{r_m} & - \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{r_1} \cdot \vec{x} \\ \vec{r_2} \cdot \vec{x} \\ ... \\ \vec{r_m} \cdot \vec{x} \end{bmatrix}$$

  In this case, each row of A is treated as a vector, and the *dot product* of $\vec{x}$ and each row becomes the row of the resulting vector.

  Note that in either view, the answers are equivalent, and using one view or the other simply helps make particular operations easier.

- **linear transformations**: a function $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ that has the following two properties:

  – **preserves vector addition**: $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}), \vec{x}, \vec{y} \in \mathbb{R}^n$

  – **preserves scalar multiplication**: $T(c\vec{x}) = cT(\vec{x}), \vec{x} \in \mathbb{R}^n, c \in \mathbb{R}$

- **kernel**: the set of vectors $\vec{x}$ such that $A\vec{x} = \vec{0}$, denoted by $\ker(A)$

- **image**: the set of vectors $\vec{y}$ such that $\vec{y} = A\vec{x}$ for some vector $\vec{x}$, denoted by $\text{im}(A)$; also known as the *range*.

- **rank-nullity theorem**: "the *rank* of an rref matrix is equal to the dimension of the *image*, and the *number of free columns* is equal to the dimension of the *kernel*".

- **identity matrix**: denoted I_n for some integer n, a matrix of n rows and n columns of the

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

- **composition**: a number of transformations performed in a particular order. For instance, given the linear transformations T and S and a vector $\vec{x}$, we can denote performing S and then T on $\vec{x}$ by saying:

$$T \circ S = T(S(\vec{x}))$$

Note that "$\circ$" is read as "T composed with S".

Its important to note that it is *not necessarily true* that $S \circ T = T \circ S$.

- **inverses**: just as the inverse of a function, the inverse of a matrix is a matrix that undoes the operations of a matrix. More information can be found above, but to invert a matrix you can simply augment it by the corresponding I_n.

A matrix, A, is only invertible if:

  - A is square (*number of rows = number of* columns)

  - rref$(A) = I$

This second requirement does in fact incorporate the first, but it is helpful to explicitly state the first as it can make determining if certain matrices are invertible or not very simple.

- **elementary matrix**: a matrix that performs a row operation on a matrix. They are typically denoted E, are always invertible, and their inverses are always elementary matrices as well. Also, by extension of always being invertible, they are, of course, always square.

- **orthogonal component**: denoted $A^\perp$ for some matrix A, this is the set of all vectors perpendicular to A.

- **transpose matrix**: denoted A^T, this is the matrix that is the result of switching the rows and columns of a matrix. For example, the transpose of $\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$ is $\begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$.

- **normal equation**: the equation used to find the least squares solution to a system of linear equations. It is given by:

$$A^T A \vec{x} = A^T \vec{b} \tag{41}$$

- **determinant**: a scalar value that can be used to describe many properties of a matrix A, denoted either as $det(A)$ or $|A|$.

- **singular matrix**: a matrix that is not invertible (i.e. $det(A) = 0$).

- **eigenvectors/eigenvalues**: a vector $\vec{v}$ is an *eigenvector* of a matrix A if:

$$A\vec{v} = \lambda\vec{v} \tag{42}$$

where λ is a scalar, and $\vec{v} \neq \vec{0}$. This λ is an *eigenvalue* for this particular eigenvector.

- **characteristic polynomial**: the polynomial that arises from $|\lambda I - A| = 0$ (when solving for the eigenvalues of A), denoted $C_A(\lambda)$.

- **algebraic multiplicity**: the number of times a particular eigenvalue appears in the characteristic polynomial of a matrix.

- **eigenspace**: the set of all vectors that are eigenvectors for a particular eigenvalue of a matrix.

- **eigenbasis**: for a matrix A in $\mathbb{R}^n$, the set of eigenvectors of A that are linearly independent, ie, are a basis of R^n.

- **geometric multiplicity**: the dimension of the eigenspace of a particular eigenvalue of a matrix.

- **diagonal matrix**: a matrix that has all of its entries outside of the main diagonal equal to zero, ie, $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$. Sometimes the 0's are omitted, for the sake of brevity, such as in $\begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix}$.

- **diagonalizable**: a characteristic of a matrix A that can be written as $A = SDS^{-1}$, where D is a *diagonal matrix* and S is invertible.

- **trace**: the sum of the diagonal entries of a matrix, denoted $tr(A)$.