APMA 3080 Notes
Linear Algebra

James Houghton
jhoughton@virginia.edu

Fall 2017

About These Notes

These notes come mostly from my memory of Dr. Jonathan Osborne's Linear Algebra course at TJHSST; however, Ms. Meiqin Li's lectures on Linear Algebra at UVA have helped to refresh my memory. These notes go in an order that I think is appropriate for introducting concepts. There is no guarantee that this covers all the information in a particular Linear Algebra course, and there is no guarantee any particular Linear Algebra course will cover all this material. I also don't intend for these notes to completely replace a Linear Algebra course. The only background you might need to fully understand these notes is a familiarity with vectors, matrices, and how to do basic algebra with the two.

One thing you don't get from written notes is pronunciation of certain words. The following are some of the words that students commonly mispronounce:

1. "Gaussian" is pronounced [ˈɡaʊs.i.ən] ("GOW-see-in").
2. "Echelon" is pronounced [ˈɛʃəlɒn] ("ESH-uh-lawn").
3. "Stochastic" is pronounced [stəˈkæstɪk] ("stuh-CAH-stik").
4. "Hermitian" is either pronounced [hɜː(ɹ)ˈmiʃən] or [hɜː(ɹ)ˈmɪʃən] ("her-MEE-shun" or "her-MIH-shun" respectively).

Please pronounce things correctly, otherwise no one will take you seriously.

I hope these notes are helpful, fun to read, and give you a good intuition for Linear Algebra. There may very well be mistakes in here, and feel free to send me an email or create a pull request if there are any corrections or improvements you would like to see made. You can find the code for these notes at `github.com/jamesthoughton/uva`. You can also follow the development of the notes there too.

It is my pleasure to be providing these notes, and I thank you for giving up some of your time to read them. Now...

Let's begin.

Contents

1 Linear Systems of Equations

1.1 Definitions and Representations

Systems of equations in the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b_0$ are said to be linear because they are linear in each of their variables. This particular system can be rewritten like this:

$$[a_1, a_2, \ldots, a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = b_0 \tag{1.1}$$

This vector/matrix representation is very helpful when dealing with multiple equations of the same variables. Let's look at a system with more equations:

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Here we call the left-most 3×3[1] matrix the coefficient matrix, and we call the x vector the solution vector. This allows for the most abstract representation of any linear system:

$$A\vec{x} = \vec{b} \tag{1.2}$$

We can write this linear system in the augmented-matrix shorthand form:

$$\left[\begin{array}{ccc|c} a_1 & a_2 & a_3 & y_1 \\ b_1 & b_2 & b_3 & y_2 \\ c_1 & c_2 & c_3 & y_3 \end{array}\right]$$

This form will make it easier for us to solve these linear systems of equations by Gauss-Jordan elimination. Any augmented matrix behaves exactly like a normal matrix, and in fact many textbooks choose to forgo the bar.

1.1.1 Partitioned Matrices

Sometimes matrices will be written with other matrices as components. This is normally a sign of a partition matrix. Let's look at the following example:

$$A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}; \quad \vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

A can just as easily be written as a partitioned matrix in the following way:

$$A = \begin{bmatrix} \vec{a} & \vec{b} & \vec{c} \end{bmatrix}$$

Let's say we have another set of matrices:

$$M = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 3 & 3 & 4 & 4 \\ 3 & 3 & 4 & 4 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, C = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}, D = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

[1]Matrix dimensions are always row by column. An $n \times m$ matrix has n rows and m columns.

M can be rewritten as a paritioned matrix in the following way:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

1.2 Row Reduction

1.2.1 Echelon Form

A matrix is said to be in echelon form if the leading element of each row is 1, and there are only 0s to the left of every leading element. Leading elements are the left-most non-zero component of a row. Let's look at some examples of matrices in echelon form:

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

The final two matrices are in the form known as reduced row-echelon form (RREF). This form is the same as echelon form just with an additional constraint: all elements above any leading element must be 0. Every matrix has a *unique* RREF, meaning that no matter how you row reduce, you will always get the same RREF. Different matrices can absolutely have the same RREF.

1.2.2 Performing Row Reduction

Row reduction, more formally known as Gauss-Jordan elimination, is a method to solve linear systems of equations. To perform row reduction, we must go through each of the columns of the coefficient matrix and select a pivot going from the top left going downwards and to the right. For now we will forgo the definition of pivot: it will be clear what they are after going through a few examples of row reduction. These pivots will become our leading elements in our row-echelon matrices. You may choose the pivots in any order you like, but to create a matrix in row-echelon form, they need to go from the top left to the bottom right. If pivots don't line up this way, rows may be flipped until they do.

With our first pivot selected, we must replace every other row with a linear combination of itself and the row with our selected pivot. Importantly, however, the chosen pivot cannot be 0, and you cannot have two pivots in the same row or column. This linear combination should set all elements above and below the pivot to 0. After you've eliminated all non-zero elements in the selected pivot's column, move on to the next column until there are none left. It may be best to understand the process if we go through an example (the chosen pivots are underlined):

$$\left[\begin{array}{ccc|c} \underline{1} & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 4 & 6 \end{array}\right] \xrightarrow[\mathrm{R_3 \to R_3 - 3R_1}]{\mathrm{R_2 \to R_2 - 2R_1}} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & \underline{-1} & 0 & 1 \\ 0 & -1 & 1 & 3 \end{array}\right] \xrightarrow[\mathrm{R_3 \to R_3 - R_2}]{\mathrm{R_1 \to R_1 + R_2}}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & \underline{1} & 2 \end{array}\right] \xrightarrow[\mathrm{R_2 \to -R_2}]{\mathrm{R_1 \to R_1 - R_3}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad (1.3)$$

Converting this back to a system of equations, this is what we get:

$$\begin{aligned} 1x_1 + 0x_2 + 0x_3 &= 0 &\implies x_1 &= 0 \\ 0x_1 + 1x_2 + 0x_3 &= -1 &\implies x_2 &= -1 \\ 0x_1 + 0x_2 + 1x_3 &= 2 &\implies x_3 &= 2 \end{aligned}$$

As we can see, each column of the coefficient matrix corresponds to one of the x variables, and the reduced row-echelon representation of the system easily gives us the solution to the system. More abstractly, row reduction tells us if different columns are linear combinations of the other columns (remember this as you later read about *span*). So, for example, if we had a matrix that row reduced to

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

it means that the second column is just 2 times the first column, and the fourth column is the sum of 3 times the first column and 4 times the third column.

A quick note on polynomials Let's say we have an unknown polynomial of known degree n. We can solve for the coefficients of this polynomial if we know at least $n + 1$ points that it contains. Let's say that the i^{th} point will have coordinates (x_i, y_i). For a polynomial in the form $a_3x^3 + a_2x^2 + a_1x + a_0 = y$, we can throw these points into matrices in this fashion:

$$\begin{bmatrix} x_0^3 & x_0^2 & x_0 & 1 \\ x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ x_3^3 & x_3^2 & x_3 & 1 \end{bmatrix} \begin{bmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

By converting this matrix equation into an augmented form and row reducing, you can easily solve for coefficients of the polynomial.

Let's do another example:

$$\left[\begin{array}{ccc|c} \underline{1} & 1 & 1 & 2 \\ 3 & 2 & 4 & 7 \\ 2 & 2 & 2 & 4 \end{array}\right] \xrightarrow[\text{R}_3 \to \text{R}_3 - 2\text{R}_1]{\text{R}_2 \to \text{R}_2 - 3\text{R}_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & \underline{-1} & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{R}_1 \to \text{R}_1 + \text{R}_2}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{R}_2 \to -\text{R}_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{1.4}$$

There is no place to put a third pivot, so we are done.

Converting this back to a system of equations yields the following:

$$\begin{aligned} 1x_1 + 0x_2 + 0x_3 &= 2 \implies x_1 = 2 \\ 0x_1 + 1x_2 - 1x_3 &= 1 \implies x_2 = 1 + x_3 \\ 0x_1 + 0x_2 + 0x_3 &= 0 \end{aligned}$$

Here we have two equations and three unknowns. That means our solution will be one-dimensional, so one of our variables will be a *free* variable, and the other two will be *basic* (defined by an equation). The variables associated with columns lacking pivots are the free variables we need. So if x_3 is our free variable, we can simply define it to be any real number, t. This gives us our solution

$$\vec{x} = \begin{bmatrix} 0 \\ 1+t \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \forall t \in \mathbb{R}.$$

It is important to note that matrices that are produced from row reduction are said to be *row-equivalent* to their original matrix and any matrix that can be produced by performing any combination of elementary row operations to that original matrix. Row-equivalence is denoted with a tilde ($\sim$), as in the following example:

$$\begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 1 \\ 1 & 0 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Now one last example:

$$\left[\begin{array}{cc|c} \underline{1} & 2 & 3 \\ 2 & 4 & 7 \end{array}\right] \xrightarrow{\mathrm{R}_2 \to \mathrm{R}_2 - 2\mathrm{R}_1} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & \underline{1} \end{array}\right] \xrightarrow{\mathrm{R}_1 \to \mathrm{R}_1 - 3\mathrm{R}_2} \left[\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 0 & 1 \end{array}\right] \tag{1.5}$$

As you can see, we placed a pivot in the augment, so if we were to convert this system back into equation form, we'd be met with the result of $0 = 1$. This is what a system with no solutions looks like in RREF.

A system with no solution is said to be *inconsistent*, and a system with at least one solution is referred to as *consistent*. We can further categorize consistent systems into the following categories: independent and dependent. Systems with exactly one solution are said to be *independent*, whereas systems with infinitely many solutions are said to be *dependent*.

The number of pivots that are produced to reduce a matrix into row-reduced form is also known as the *rank* of that matrix. In our first row reduction example (1.3), we can see that we chose three pivots, therefore the rank of that augmented matrix is 3. In our second example (1.4), we only had two pivots, yielding a rank of 2. If we have a square matrix in $\mathbb{R}^n$ with n pivots, the matrix is said to be of *full rank*.

1.3 Column, Row, and Null Spaces

1.3.1 Span

Span can be thought of as the reachable space[2] any linear combination of a given set of vectors. For example:

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} \right\} = t \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + s \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} \forall t, s \in \mathbb{R}. \tag{1.6}$$

I hope this makes it clear what is meant by span. For a vector to be represented as a linear combination of other vectors, it must be in the span of those other vectors. That means

[2] We will get into explicit definitions for vector spaces in section 4.

for a solvable system of equations, the augment of its matrix must be in the span of the columns of the coefficient matrix, or in other words, as we said before, there cannot be a pivot in the augment.

Now let's get into some definitions. A *spanning set* is a set of vectors that span a certain space. A spanning set can contain vectors that do not carry any new information (when they are parallel to vectors already in the set), and it does not change the dimension of the span. A *basis* is a spanning set of a vector space with the fewest vectors. For any vector space, there must exist a basis containing exactly the same number elements as the dimension of the space. A basis is always a *set* of vectors, not a span of a set. A basis for the span from our first example (1.6) would be

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} \right\}$$

When a basis is asked for, provide only the set of vectors that form the basis, not the span of that set.

Let's look at another span example:

$$\operatorname{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \mathbb{R}^3$$

This particular spanning set is referred to as the standard basis for $\mathbb{R}^3$. These vectors together are sometimes referred to as the e basis (the "easy" basis), and the vectors themselves will sometimes be referred to as the following:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = e_1 = \hat{i}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = e_2 = \hat{j}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = e_3 = \hat{k}$$

1.3.2 Column Space

The *column space* of a matrix is just the span of its columns. We have used "the span of columns" before to help solve problems, but now we are just giving that a name. Importantly, row reduction does change the column space of a matrix, but not its dimension. This means that the number of pivots of a matrix in RREF represents the dimension of the column space, and that the columns with pivots represent the basis elements of that column space. In other words, the columns of any matrix form a spanning set of the column space, but that does not mean they form a basis. Removing the columns lacking pivots produces the spanning set with the minimum number of elements, a basis.

1.3.3 Row Space

The *row space* of a matrix is defined as the span of its rows. The row space of a matrix will become more important later on, but for now it is important to notice that

$$\dim(\operatorname{col} A) = \dim(\operatorname{row} A) = \operatorname{rank} A.$$

It is important to notice that row reduction does not change the row space of a matrix. This is because we are replacing rows with linear combinations of those rows with other rows, and that does not change the span of the rows.

1.3.4 The Homogenenous Equation and the Null Space

The *null space* (or *kernel*[3]) of a matrix is the set of solutions to the *homogeneous equation*:

$$A\vec{x} = 0 \tag{1.7}$$

Let's examine the properties of the homogeneous equation. We can see that there will always be at least one solution: $\vec{x} = \vec{0}$. This solution is referred to as the *trivial solution*. Therefore, $\vec{0}$ will always be in the null space of any matrix. Now, let's examine the dimension of the null space, or *nullity*, of a particular matrix. By the Rank-nullity Theorem, the sum of the rank and nullity of a matrix must equal the number of columns in that matrix. Hopefully this example can make that seem obvious:

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

Clearly here we have a matrix with four columns and of rank 2. We should, therefore, have a null space of dimension 2. Taking $x_3 = s$ and $x_4 = t$, we clearly see that the null space is the following:

$$\vec{x} = \begin{bmatrix} -2s - t \\ -t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \forall s, t \in \mathbb{R}.$$

Just as the theorem predicted, the null space is two-dimensional. As we will discuss later, $\vec{0}$ is always in the column space of a matrix. As with any target vector in the column space of a matrix, there will be an n-dimensional solution where n is the number of free variables in that system. The number of free variables associated with a system is always the difference between the number of columns in the system's associated coefficient matrix and the rank of that associated matrix. Henceforth, we will refer to the number of free variables in a system as the nullity and the number of basic variables as the rank.

1.4 Linear Dependence

Let's take a look at some vectors.

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$$

Now let's throw these into a matrix and row reduce.

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Let's go from left to right looking at span.

$$\vec{u}_2 \notin \text{span}\{\vec{u}_1\}$$
$$\vec{u}_3 \in \text{span}\{\vec{u}_1, \vec{u}_2\}$$

[3]The word "kernel" often refers to matrices in the context of transformations.

Because vectors $\vec{u}_1$ and $\vec{u}_2$ are not in each other's span, they are said to be *linearly independent*. Linear dependence is a term describing a set of vectors, so it is most correct to say the following: given

$$S = \{\vec{u}_1, \vec{u}_2\},$$

S is linearly independent. Adding $\vec{u}_3$ to set S causes it to be *linearly dependent* because $\vec{u}_3 \in \text{span}\{\vec{u}_1, \vec{u}_2\}$. Adding $\vec{0}$ will always make a set of vectors linearly dependent. For now we can prove it by showing that $c\vec{0}$ will always be 0 for all chosen c, so it just adds a dimension to the solution space for any linear system.

2 Matrices as Transformations

In this section, we will interpret matrices as transformations. This should make it easier to explain certain concepts in Linear Algebra.

2.1 Arithmetic Properties of Matrices

We are going to quickly describe the important properties of matrices when performing matrix algebra. The distributive property

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} \tag{2.1}$$

holds. You have the tools to proof this for yourself. Scalar multiplication

$$A(r\vec{x}) = rA\vec{x} \tag{2.2}$$

also holds. You can prove this for yourself. The associative property

$$A(B\vec{x}) = (AB)\vec{x} \tag{2.3}$$

holds as well. You also have the tools to prove this yourself, but this will seem obvious after our discussion of matrices as transformations. Matrices do not commute!

$$AB \neq BA \tag{2.4}$$

This can be understood by examining the resulting dimensions of matrix multiplication. If A is a 4×3 matrix and B is a 3×5 matrix, then AB will be a 4×5 matrix. However, BA won't be a valid product unless we redefine the matrix product (which we will not do), so we can see that the commutative property definitely doesn't hold. The only time $AB = BA$ is when A and B are both square, but it is still not guaranteed. This will be more easily understood after the following discussion.[4]

2.1.1 Linearity

We say that a matrix transformation is a linear transformation because matrix multiplication is a linear operation. Linearity is defined in the following way:

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) \tag{2.5}$$

$$T(r\vec{x}) = rT(\vec{x}) \tag{2.6}$$

[4] There is an operator, the commutator, that calculates how 'close' two matrices are. The commutator of A and B is $[A, B] = AB - BA$. When $[A, B] = 0$, A commutes with B.

Another operation you may already know is linear is scalar multiplication. It is important to realize that any linear transformation T acting on $\vec{0}$ will always yield $\vec{0}$, but those two $\vec{0}$ may belong to different vector spaces. Let's prove this:

$$\begin{aligned} T(\vec{0}) &= T(\vec{0} + \vec{0}) \\ &= T(\vec{0}) + T(\vec{0}) \end{aligned}$$

Subtracting $T(\vec{0})$ from both sides yields

$$\vec{0} = T(\vec{0}),$$

which is what we started with.

2.2 Matrices as Functions

When we look at equation (1.2), we can think about that as $\vec{x}$ being acted open by matrix A to produce $\vec{b}$. If A is an $m \times n$ matrix, then it takes a vector $\vec{x}$ in $\mathbb{R}^n$ and produces a vector $\vec{b}$ in $\mathbb{R}^m$.[5] From the definition of the column space, the output space of a matrix will be its column space. The input vector $\vec{x}$ contains the coefficients for each of the columns of the matrix to produce a vector $\vec{b}$ in $\text{col}(A)$.

2.2.1 Domain, Codomain, and Range

Any $\vec{x}$ such that $A\vec{x}$ is in the *domain* of matrix A. The set of all $\vec{b}$ such that $A\vec{x} = \vec{b}$ is referred to as the *range* of matrix A. The range of matrix A is just the column space of A. I will refer to these vectors $\vec{b} \in \text{col}(A)$ as reachable. If matrix A is $m \times n$, its domain is $\mathbb{R}^n$ and its *codomain* is $\mathbb{R}^m$. Note that not all vectors in its codomain may be reachable. It is important to not confuse the range and the codomain of a matrix. The range is always a subspace[6] of the codomain. A matrix with a domain of $\mathbb{R}^n$ and codomain of $\mathbb{R}^m$ is said to 'map' vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$, and this is denoted as the following.

$$T : \mathbb{R}^n \mapsto \mathbb{R}^m \tag{2.7}$$

2.3 Markov Chains

2.3.1 Stochastic Matrices

Probability vectors are just vectors that represent probability distribution.

2.4 Two-Dimensional Transformations

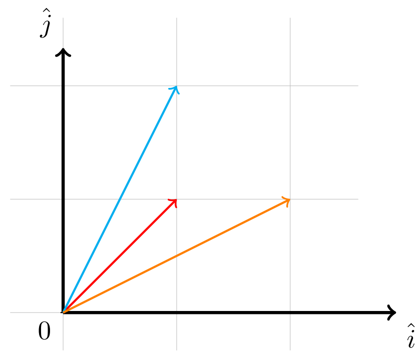
2.4.1 Some Types of Transformations

Scaling Transformation A scaling transformation scales any number of the components of the vectors that are inputted into the transformation. Let's say that we have two transformations, A and B. Let's say A scales the x component of its input vectors by 2, and

[5]There is no shame in having to draw this out so you remember the input and output spaces of a matrix. This *needs* to be clear.

[6]"Subspace" can mean any space inside another space or the space itself. A "proper subspace" refers to subspaces that are not the same as the original space. $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$, but it is not a *proper* subspace of $\mathbb{R}^3$.

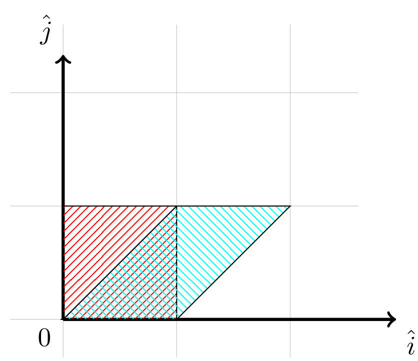
let's say B scales the y component of its input vectors by 2. Let's consider the input vector $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, seen on the graph in red. The orange vector represents the output for $A\vec{x}$, and the cyan vector represents the output for $B\vec{x}$.



We can easily generate scaling transformations, as they're pretty intuitive:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

The Sheer Transformation A sheer looks like the following, where the red region is the initial region, and the cyan region is the transformed region:



Don't be thrown off by the transformation of a region instead of a single vector. It is just a transformation of the vectors representing the corners of the region. This particular sheer is a simple one:

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

You can examine these transformations by looking at what they do to the basis vectors ($\hat{i}$ and $\hat{j}$) and the corners of the inital region. These are the following transformations:

$$A \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad A \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

These should be the points that define the cyan region in earlier the graph.

Rotations You need to be very familiar with the rotation matrix

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

where θ is measured counterclockwise (the normal way angles are measured). This rotates its input about the origin by angle θ counterclockwise. To get the clockwise matrix, we can just negate θ and use the odd and even properties of sin and cos respectively:

$$A_{\text{clockwise}} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

Translations and Homogeneous Coordinates We have learned previously that, for linear transformations, $\vec{0}$ must always map to $\vec{0}$. With that stipulation, how are we supposed to get translations? In two dimensions, if we add a third element, we can fake translations. Consider this example:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 + 2 \\ x_2 + 3 \\ 1 \end{bmatrix}$$

If we just look at the first two elements of the input and output vectors, we see that we translated the vector 2 units up and 3 units right. When using homogeneous coordinates for otherwise two-dimensional transformations, we just need to add an extra row and column to keep the added 1 in our input vector. For example, the rotation matrix in homogeneous coordinates is the following:

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

This is important when we compose transformations.

2.4.2 Composing Transformations

Let's say A and B are two different matrices representing different vector transformations. For simplicity let's say AB exists. The product AB represents the following:

1. The transformation associated with B is applied to the input $\vec{x}$, **then**

2. The transformation associated with A is applied to its input (now $B\vec{x}$)

It is important to note that we go from right to left when using matrices as functions. You might as well write matrix compositions like this:

$$AB\vec{x} = A\,(B\vec{x})$$

Let's do an important composition example: Find a matrix that rotates its input counterclockwise by $\frac{\pi}{2}$ radians about the point $(-5, 3)$. Because the rotation matrix rotates only about the origin, we're going to have to move $(-5, 3)$ to the origin, do the rotation, and then move it back. We do this by using a translation:

$$\begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{is this correct?}} \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -5 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Now we must perform the rotation. Because we used homogeneous coordinates to do the translation, the rotation must also be in homogeneous coordinates:

$$\begin{bmatrix} \cos\left(\frac{\pi}{2}\right) & -\sin\left(\frac{\pi}{2}\right) & 0 \\ \sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then we must translate the coordinates back the opposite transformation we did to translation $(-5, 3)$ to the origin:

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

We multiply all of these together, careful of the order. Remember, the matrix furthest to the *right* acts first.

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

That final matrix represents the transformation that was asked for. Now let's see where $(4, -3)$ maps:

$$\begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & 8 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 12 & 1 \end{bmatrix}$$

As we can see $(4, -3)$ maps to $(1, 12)$ under this transformation.

2.5 Describing Matrices and Their Solution Spaces

For a matrix with linearly independent columns (i.e. a matrix of nullity 0), there will be only one way to get a solution. This kind of matrix is classified as *one-to-one*: for every output, there can only be one input that produces it. One-to-one matrices are often stated to have a pivot in every column.

Now let's consider matrices with a pivot in every row. A matrix of this type is said to be *onto*.[7] Let's consider the following matrices:

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The column space of matrix A is clearly two-dimensional, yet we know that output vectors are elements of $\mathbb{R}^3$. Logically then, the column space of A is a two-dimensional subspace of $\mathbb{R}^3$.[8] Now let's look at B. The column space of B, by the same analysis as before, is a three-dimensional subspace of $\mathbb{R}^3$, but that's just $\mathbb{R}^3$. When the column space of an $m \times n$ matrix is $\mathbb{R}^m$, we say the matrix is onto (i.e. it gets everywhere). An onto matrix's codomain is equivalent to its range.

[7] Yes, a matrix can be *onto*. I know you're probably used to hearing 'onto' as an adverb, but in this context it is an adjective.

[8] It is important to not that $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$. There are no elements in $\mathbb{R}^2$ that are in $\mathbb{R}^3$.

An onto transformation is referred to as a *surjection* or a *surjective transformation*. A one-to-one transformation is referred to as an *injection* or an *injective transformation*. A transformation that is both onto and one-to-one is said to be a *bijection* or a *bijective transformation*.

A matrix can be both onto and one-to-one, but then it would have to be square and it's RREF would be in the form

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

This matrix is said to be diagonal, with only 1s along its main diagonal. This is also referred to as the *identity matrix*, denoted as I. For an onto and one-to-one $m \times m$ matrix A, its RREF is always the $m \times m$ identity matrix I_m. In other words, A is row-equivalent to the identity matrix. Matrices that are both onto and one-to-one are said to be invertible. We will discuss invertibility in section 3.

3 Matrix Operations

3.1 Transposition

Transposition is very simple: it is a flip along the main diagonal. Because of this definition, you cannot transpose non-square matrices. This is how transposition is defined:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \tag{3.1}$$

Here is an important transposition property that is very helpful when proving identities later on:

$$(AC)^T = C^T A^T \tag{3.2}$$

This can be proven for n-dimensional matrices using index notation, which we will discuss in section 6.1.

3.2 Inversion

3.2.1 In the Context of Linear Transformations

If we have a transformation $T(\vec{x}) = A\vec{x} = \vec{y}$, then $T^{-1}(\vec{y}) = \vec{x}$. This gives us another way to solve certain systems of equations: using the inverse of the coefficient matrix, assuming it is invertible. A matrix transformation can only have an inverse, as we said before, if it is both onto and one-to-one. If a matrix maps vectors from $\mathbb{R}^n$ to a vector space that is not $\mathbb{R}^n$, then the matrix cannot be invertible. Think about that using some examples like $\mathbb{R}^3 \mapsto \mathbb{R}^2$.

3.2.2 Singular Matrices

A matrix is *singular* or *degenerate* if it is square but does not have an inverse. So, for example, a matrix that maps $\mathbb{R}^n$ to $\mathbb{R}^n$ but is not onto will be singular. This can be seen if, when reducing a square matrix, a row of zeros appears at the bottom. This is a mark of a singular matrix. Singular matrices are said to be "rare", in that most randomly-generated square matrices are not singular (i.e. they are invertible). A matrix is *non-singular* if it is square and invertible.

3.2.3 Invertible Matrix Theorem (IMT)

When talking of an $n \times n$ matrix A, all of the following statements imply that it is invertible:

1. A is row-equivalent to I_n, the $n \times n$ identity matrix.
2. A has n pivot positions.
3. $A\vec{x} = \vec{0}$ has only one solution: the trivial solution.
4. The columns of A are linearly independent.
5. The linear transformation $\vec{x} \mapsto A\vec{x}$ is one-to-one, onto, a surjection, an injection, or a bijection.
6. $\forall \vec{b} \in \mathbb{R}^n, A\vec{x} = \vec{b}$ has exactly one solution.
7. $\text{row}(A) = \mathbb{R}^n$.
8. $\text{col}(A) = \mathbb{R}^n$.
9. The columns of A form a basis for $\mathbb{R}^n$.
10. The columns of A span $\mathbb{R}^n$.
11. There exists C such that $CA = I_n$.
12. There exists D such that $AD = I_n$.
13. A^T is invertible.
14. The dimension of the column space of A is n.
15. The rank of A is n.
16. The null space of A is $\{\vec{0}\}$.
17. The dimension of the null space of A is $\vec{0}$.
18. $\det A$ is non-0.
19. 0 is not an eigenvalue of A.
20. The orthogonal complement of $\text{col}(A)$ is $\{0\}$.
21. The orthogonal complement of $\text{Null}(A)$ is $\mathbb{R}^n$.

We will discuss eigenvalues and orthogonal complements later on in this discussion, so don't worry about those last three for now. It is important that the IMT is intuitive. You should *not* have to memorize these; most of them should just make sense.

3.2.4 Calculating Inverses

Inverting 2×2 matrices is easy, and you may be asked to do this many times on any Linear Algebra test. Inverting 3×3 matrices can be done in reasonable time, and 4×4 matrices will be largest matrices you will reasonably be asked to invert.

Inverse by Determinant Here we will focus only on 2×2 matrices. For a 2×2 matrix A,

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

Remember: "swap and negate" *in that order*. Swap a with d and then negate b and c. After doing that, multiply the matrix by reciprocal of the determinant: $\frac{1}{ad-bc}$. You're just going to have to memorize that.

Method of Row Reduction For an $n \times n$ matrix A, the following holds:

$$[A \mid I_n] \sim \left[I_n \mid A^{-1}\right], \tag{3.3}$$

which means that you can row reduce with the identity matrix as the augment to get A^{-1}. This can be a real chore to do by hand, but it is the most efficient way for large martices. It is important to note that if, for example, while you are row reducing, you get a row of 0s, you can conclude the matrix you are trying to invert is *not* invertible.

Method of Cofactor Expansion This is the method that is perhaps most efficient for 3×3 matrices. Cofactor expansion (also called "expansion by minors", not to be confused with the method of calculating a determinant) relies on the following identity:

$$A^{-1} = \frac{C^T}{\det A} \tag{3.4}$$

C is defined as the matrix of cofactors, and M is the matrix of minors. For matrix A where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

M is defined as

$$M = \begin{bmatrix} \begin{vmatrix} a_{22}a_{23} \\ a_{32}a_{33} \end{vmatrix} & \begin{vmatrix} a_{21}a_{23} \\ a_{31}a_{33} \end{vmatrix} & \begin{vmatrix} a_{21}a_{22} \\ a_{31}a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{12}a_{13} \\ a_{32}a_{33} \end{vmatrix} & \begin{vmatrix} a_{11}a_{13} \\ a_{31}a_{33} \end{vmatrix} & \begin{vmatrix} a_{11}a_{12} \\ a_{31}a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{22}a_{23} \\ a_{22}a_{23} \end{vmatrix} & \begin{vmatrix} a_{21}a_{23} \\ a_{21}a_{23} \end{vmatrix} & \begin{vmatrix} a_{21}a_{22} \\ a_{21}a_{22} \end{vmatrix} \end{bmatrix}. \tag{3.5}$$

The minor associated with every element in the initial matrix is the determinant of all the elements not in that element's row or column. Converting to a cofactor matrix (as opposed to a matrix of minors without signs) from this just means multiplying each element in M by alternating $+$ and $-$ signs.

$$C = \begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix} = \begin{bmatrix} \begin{vmatrix} a_{22}a_{23} \\ a_{32}a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21}a_{23} \\ a_{31}a_{33} \end{vmatrix} & \begin{vmatrix} a_{21}a_{22} \\ a_{31}a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12}a_{13} \\ a_{32}a_{33} \end{vmatrix} & \begin{vmatrix} a_{11}a_{13} \\ a_{31}a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11}a_{12} \\ a_{31}a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{22}a_{23} \\ a_{22}a_{23} \end{vmatrix} & -\begin{vmatrix} a_{21}a_{23} \\ a_{21}a_{23} \end{vmatrix} & \begin{vmatrix} a_{21}a_{22} \\ a_{21}a_{22} \end{vmatrix} \end{bmatrix} \tag{3.6}$$

This can be done for matrices of dimensions larger than 3×3, but instead of doing determinants of 2×2, you'll end up having to do determinants of 3×3 or $n \times n$ which will require tons of work. Putting this result for C back into equation (3.4), we can solve for A^{-1}. Sometimes C^T is referred to as the *adjugate* or *adjunct* of matrix A.[9]

3.3 Determinants and Their Meaning

3.3.1 Invertibility

The determinant has been designed to give us an easy way to figure out if a matrix is or is not invertible. By design and definition, the determinant of an invertible matrix is non-0, and the determinant of a non-invertible matrix is 0.

We can illustrate the properties of a determinant with this example: Suppose we had a *linearly independent* set of vectors

$$S = \left\{ \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ 10 \end{bmatrix} \right\},$$

and we wanted to see if

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \text{span}(S).$$

To do this, let's consider the matrix

$$A = \begin{bmatrix} 1 & 4 & 1 \\ 3 & -2 & 1 \\ -1 & 10 & 1 \end{bmatrix}$$

Because we know set S is linearly independent, we know that

$$A = \begin{bmatrix} 1 & 4 & 1 \\ 3 & -2 & 1 \\ -1 & 10 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & 0 \end{bmatrix} \text{ OR } \begin{bmatrix} 1 & 4 & 1 \\ 3 & -2 & 1 \\ -1 & 10 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In the former case, the determinant of A will be 0 and $\vec{u}$ will be in the span of the columns of A. In the latter case, the determinant of A will be non-0 and $\vec{u}$ will not be in the span of

[9] The terms "adjugate" and "adjunct" are not to be confused with the term "adjoint." The adjoint of matrix A is its complex conjugate transpose, $A^\dagger$. When working with strictly real matrices (as we are in this section), $A^\dagger = A^T$.

the columns of A. It turns out

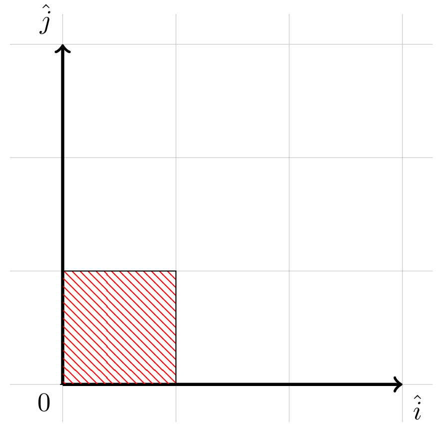
$$\begin{vmatrix} 1 & 4 & 1 \\ 3 & -2 & 1 \\ -1 & 10 & 1 \end{vmatrix} = 0.$$

Because the determinant is 0, we know that $\vec{u}$ is in the span of the columns of A, answering our question.

3.3.2 Area Scaling

Determinants also have another property: their values correspond to how much any given area will be scaled. Let's look at a 2×2 example:

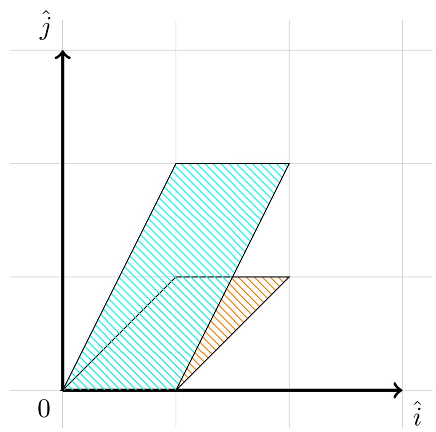
Let's say our area of interest is just a 1 by 1 square in the first quadrant:



Let's perform a sheer transformation and then scale the $\hat{j}$ value:

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

Remember that we can compose matrix transformations in this fashion. Having A act on the following area, we get the following:



The orange region is what comes after transforming the initial square with the sheer transformation $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The cyan region is the orange region scaled by 2 along the $\hat{\jmath}$ axis. We can determine the area of both the orange and the cyan regions using the determinant:

$$\begin{aligned} \text{Area}_{\text{orange}} &= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \\ &= 1 \\ \text{Area}_{\text{cyan}} &= \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} \\ &= 2 \end{aligned}$$

Keep in mind the determinant represents how much the area is scaled, so, because our 1×1 square had an area of 1, the area was just the same as the scale factor. If the initial square were a 2×2 square, the area of the cyan region would have $4 \cdot 2 = 8$.

Distributive Property It should make sense, then, that the following applies:

$$\det(AB) = \det(A) \cdot \det(B) \tag{3.7}$$

If we look at the previous example, we see that the cyan area has a determinant of 2 because the scaling transformation has a determinant of 2:

$$\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) \cdot \det\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) = 2 \cdot 1 = 2$$

We can expand equation (3.7) to get the following identity:

$$\det(AB) = \det(A) \cdot \det(B) = \det(B) \cdot \det(A) = \det(BA)$$

Scaling If you multiple any row in an $n \times n$ matrix by a factor c, the determinant of that matrix will also be multiplied by c. If you multiply all rows by c, the determinant of that matrix will be multiplied by c^n. Therefore, we get the following identity:

$$\det(cA) = c^n \det(A)$$

Scaling an individual row, just with a 3×3 example for simplicity:

$$c \det(A) = \begin{vmatrix} ca_{11} & ca_{12} & ca_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ ca_{31} & ca_{32} & ca_{33} \end{vmatrix}$$

3.4 Calculating Determinants

3.4.1 Properties of the Determinant

Row and Column Reduction To begin our discussion, I need to make it clear that in the context of determinants, row reduction and column reduction mean the same thing. This is due to the fact that the determinant of the tranpose of a matrix is the determinant of the original matrix. We know that the determinant, by design, does not change under a certain type of row reduction. Allow row operations while taking determinants must replace the altered row with itself, *not* a multiple of itself. So, for example, you cannot replace R_2 with $2R_2 - 3R_3$, but you can replace it with $R_2 - \frac{3}{2}R_3$. This is very important. If you scale a row, the determinant will be scaled by that same factor. Similarly, if you *flip* two rows, the determinant will be negated. Sometimes it is easier to flip rows, calculate the determinant, and then negate the determinant again. From now on, row reduction in the context of determinants will refer only to this "special" type of row reduction: leaving coefficients of altered rows as 1.

Transposition Most of the tricks to calculate determinants come from the fact that the determinant of a matrix is the same as the determinant of the transpose of that matrix.

$$|A| = \left|A^T\right|$$

Because of the previous two properties, we know that the determinant doesn't change under (special) column reduction either, because column reduction is just row reduction of the transpose. It must be clear that we can do any valid manipulations in any order. For example, the following procedure is valid: row reduce a little bit, transpose the matrix, row reduce more, transpose again, column reduce. None of those operations change the determinant.

3.4.2 Expansion by Minors

When calculating the determinant, we can "expand" along any row or column. Most people are used to just expanding along the top row, but due to how the determinant is designed, we can actually expand along anything. When expanding, each element x in the row of column you choose is multiplied by the determinant of the minor matrix that is produced when you remove the column and row that the current element x belongs to. Let's use the following matrix as an example:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 10 & -2 & 4 \\ 3 & 7 & 5 & 5 \\ 4 & 8 & -1 & 3 \end{bmatrix}$$

Upon first examination, we can see that the second column looks almost like a scalar multiple of the first column, so let's do some column reduction:

$$A \sim \begin{bmatrix} 1 & 0 & 3 & 4 \\ 5 & 0 & -2 & 4 \\ 3 & 1 & 5 & 5 \\ 4 & 0 & -1 & 3 \end{bmatrix}$$

Now that we have a column of almost all zeros, we can expand down that column. When expanding you need to remember the matrix of signs that must be followed. For a 4×4 matrix, it is the following:

$$\begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}$$

For any size, you can calculate the signs just by going from the top left, taking it as +, and then alternating as you move vertically or horizontally (not diagonally) through the matrix. Let's expand down the second column:

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

$$\begin{aligned} &= (-1)\cdot \begin{vmatrix} 1 & 3 & 4 \\ 5 & -2 & 4 \\ 4 & -1 & 3 \end{vmatrix} \\ &= (-1)\cdot \left[1 \cdot \begin{vmatrix} -2 & 4 \\ -1 & 3 \end{vmatrix} + -5 \cdot \begin{vmatrix} 3 & 4 \\ -1 & 3 \end{vmatrix} + 4 \cdot \begin{vmatrix} 3 & 4 \\ -2 & 4 \end{vmatrix} \right] \end{aligned}$$

Before we continue, we should find the general determinant of a 2×2 matrix. We have seen this already, but we can prove it using expansion by minors. Consider the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

We can expand own the first row, giving us

$$\begin{aligned} \det(A) &= a \cdot \begin{vmatrix} d \end{vmatrix} - b \cdot \begin{vmatrix} c \end{vmatrix} \\ &= ad - bc \end{aligned}$$

Going back to our determinant example:

$$\begin{aligned} &= (-1)\cdot [1 \cdot (-6+4) + (-1)\cdot 3 \cdot (15-16) + 4 \cdot (-5+8)] \\ &= (-1)\cdot 13 = -13 \end{aligned}$$

3.4.3 Triangular Matrices

If you are able to reduce a matrix to an upper- or lower-triangular matrix[10], the determinant is just equal to the product of the elements along the diagonal of the reduced matrix. You can easily prove this by doing expansion by minors for a general upper- and lower-triangular matrix.

[10]Upper-triangular matrices are always in row-echelon form. Lower-triangular matrices are just the transpose of a matrix in row-echelon form.

3.5 Applications of the Determinant

3.5.1 Cramer's Rule

Given the following linear system with square $n \times n$ matrix A

$$A\vec{x} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix} \vec{x} = \vec{b},$$

Cramer's Rule tells us that the i^{th} component of the solution is given by replacing the i^{th} column with vector b. After the vectors have been replaced, you may call the new matrix A_i. In other words:

$$x_i = \frac{\det\left(\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_{i-1} & \vec{b} & \vec{u}_{i+1} & \cdots & \vec{u}_n \end{bmatrix}\right)}{\det(A)} = \frac{\det(A_i)}{\det(A)}$$

Let's solve a system of equations using this method.

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 2 & 1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

Cramer's rule tells us the following:

$$x_1 = \frac{\begin{vmatrix} 3 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & 2 & 1 \end{vmatrix}}{\det(A)}, \quad x_2 = \frac{\begin{vmatrix} 2 & 3 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \end{vmatrix}}{\det(A)}, \quad x_3 = \frac{\begin{vmatrix} 2 & 1 & 3 \\ 1 & -1 & 0 \\ 1 & 2 & 0 \end{vmatrix}}{\det(A)}$$

Let's solve for each of the determinants:

$$\begin{aligned} \det(A) &= 2 \cdot (-1 + 2) - 1 \cdot (1 + 1) + 1 \cdot (2 + 1) = 3 \\ \det(A_1) &= 3 \cdot (-1 + 2) = 3 \implies x_1 = 1 \\ \det(A_2) &= -3 \cdot (1 + 1) = -6 \implies x_2 = -2 \\ \det(A_3) &= 3 \cdot (2 + 1) = 9 \implies x_3 = 3 \end{aligned}$$

This gives us the solution vector

$$\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

4 Vector Spaces

4.1 Subspace

A subset S of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ if S satisfies the following:

1. S contains $\vec{0}$
2. If $\vec{u}$ and $\vec{v}$ are in S, $\vec{u} + \vec{v}$ is in S. This is called being "closed under addition".

3. If r is a real number and $\vec{u}$ is in S, $r\vec{u}$ is also in S. This is called being "closed under scalar multiplication".

It is important to know that any subspace is also a vector space. Let's look at an example: Let S be the set of vectors

$$\begin{bmatrix} a \\ b \end{bmatrix},$$

where a and b are integers. This is *not a subspace*. Just multiply the vectors by $\frac{1}{2}$, and you'll have problems. Let's look at another example: Let S be the set of vectors of the form

$$\begin{bmatrix} 2a-b \\ 3b \\ a+5b \end{bmatrix}.$$

Let's deconstruct this into the following:

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

It logically follows that

$$S = \text{span}\left\{\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}\right\}$$

Clearly S has a 0, is closed under addition, and is closed under scalar multiplication.[11] If any set of vectors S can be written as a span of another set of vectors R, then S is always a vector space with a spanning set R. Let's do another example: Let's say S is the set of all vectors in $\mathbb{R}^3$ of the form

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix},$$

such that $v_1 + v_2 + v_3 = 0$. Is S a subspace of $\mathbb{R}^3$? We can set $v_3 = -v_1 - v_2$, and then we can substitute this in.

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ -v_1 - v_2 \end{bmatrix}.$$

Clearly this means that

$$S = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right\},$$

so S is a vector space and a subspace of $\mathbb{R}^3$. If you can show that S can be represented as a span of vectors, you don't need to rigorously show that S is a vector space because we already know that all spans are vector spaces. Let's look at our final example: Let S be the set of vectors of the form

$$\begin{bmatrix} 2a-b \\ 3b+1 \\ a+5b \end{bmatrix}.$$

[11] I recommend looking up a proof if you don't think this is intuitive.

Let's desconstruct this into the following:

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

So now whether or not this space can be represented as a span depends on whether or not

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \in \text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix} \right\}$$

We could row-reduce to see if this is true, but we could also take the determinant.

$$\begin{vmatrix} 0 & 2 & -1 \\ 1 & 0 & 3 \\ 0 & 1 & 5 \end{vmatrix} = -2(4) - 1 = -9 \implies \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \notin \text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix} \right\}$$

This means that S cannot be represented as a span, so S cannot be a vector space and therefore is not a subspace of $\mathbb{R}^3$.

4.1.1 Null Space Revisited

It is important to note that the null space of any matrix is always a vector space. Clearly, for any *linear* transformation T, $T(\vec{0}) = \vec{0}$, so we know that the zero vector is included. The other properties of vector spaces will clearly be present in any null space problem you do. In fact, the null space of any matrix can always be represented as a span of a set of vectors.

4.1.2 Isomorphism

Any vector space V_1 that is the same dimension as another vector space V_2 is said to be *isomorphic* to V_2. It may also be stated in this way: you can create an isomorphism from V_1 to V_2. An isomorphism is effectively a bijective (invertible) map from any element in V_1 to V_2. For example, any two-dimensional subspace of $\mathbb{R}^3$ may have an isomorphism created with $\mathbb{R}^2$. Indeed any two-dimensional vector space is isomorphic to $\mathbb{R}^2$.

4.2 Change of Basis

Let's say we have a vector space that is represented with a basis. We know that there are infinitely many bases that could be used. Consider two linearly independent vectors in $\mathbb{R}^2$. No matter which two vectors you choose (as long as they are linearly independent), we can represent all vectors in $\mathbb{R}^2$ with those two vectors. So let's say we have these two linearly independent vectors:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Now let's say we wanted to represent the vector $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$ in terms of those two vectors. First, we must recognize that the vector $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$ has been written in the e basis. Recall that the e basis for $\mathbb{R}^2$ is the following:

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

This means that, when we write $\begin{bmatrix}5\\3\end{bmatrix}$ we really intend it to mean the following:

$$\begin{bmatrix}5\\3\end{bmatrix} = 5\vec{e}_1 + 3\vec{e}_2 = 5\begin{bmatrix}1\\0\end{bmatrix} + 3\begin{bmatrix}0\\1\end{bmatrix}$$

You can think about this as reaching a point in a plane using only the $\hat{\imath}$ and $\hat{\jmath}$ unit vectors. If we didn't have the so-called e basis to express vectors in, we wouldn't be able to express these vectors. Similarly, if the unit vectors $\hat{\imath}$ and $\hat{\jmath}$ didn't exist, we wouldn't be able to describe points in a Cartesian plane. I hope this makes sense. We also must recognize that the vectors describing our new basis also have been written in the e basis.

$$\begin{bmatrix}1\\1\end{bmatrix} = 1\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}0\\1\end{bmatrix} \qquad \begin{bmatrix}2\\1\end{bmatrix} = 2\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}0\\1\end{bmatrix}$$

What does it mean to write a vector in another basis? It follows from what we just talked about with the e basis. Let's define our new basis B as the following (using the vectors from before):

$$B = \left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\} = \left\{\vec{b}_1, \vec{b}_2\right\}$$

Now let's say we have a vector in basis B (denoted with a B subscript). It is defined in the following way:

$$\begin{bmatrix}x_1\\x_2\end{bmatrix}_B = x_1\vec{b}_1 + x_2\vec{b}_2 = x_1\begin{bmatrix}1\\1\end{bmatrix} + x_2\begin{bmatrix}2\\1\end{bmatrix} \tag{4.1}$$

The result of this calculation is in the e basis, because it is just a linear combination of vectors in the e basis (and these vectors are the basis vectors of basis B). When no subscript is provided, we assume the vector is written in the e basis as we have been since the beginning of our exploration into Linear Algebra, you probably just didn't realize it. For clarity, sometimes vectors will have e as a subscript, explicitly stating that they have been written in the e basis even if it would already be implied. Given our previous result, if we have a vector written in this new basis (we will call this basis B), then we can convert it to the e basis by multiplying it like this:

$$\begin{bmatrix}\vec{b}_1 & \vec{b}_2\end{bmatrix}_{eB}\begin{bmatrix}x_1\\x_2\end{bmatrix}_B = \begin{bmatrix}1 & 2\\1 & 1\end{bmatrix}_{eB}\begin{bmatrix}x_1\\x_2\end{bmatrix}_B$$

This comes straight from the definition, equation (4.1). Here the eB subscript makes it clear that we are converting from the B basis to the e basis. We don't need this, but it makes doing change of basis problems easier to comprehend and harder to mess up.

So converting *to* the e basis is easy, but what about converting *from* the e basis into another basis? The transformation from the B basis to the e basis is the last example was a linear transformation, and they always will be. Because of this, we can just invert the matrix, and it will give us a transformation from the e to the B basis or whichever basis we want. Using our last example:

$$\begin{bmatrix}1 & 2\\1 & 1\end{bmatrix}_{eB}^{-1} = \frac{1}{1-2}\begin{bmatrix}1 & -2\\-1 & 1\end{bmatrix}_{Be} = \begin{bmatrix}-1 & 2\\1 & -1\end{bmatrix}_{Be}$$

Let's test this. Consider the vector $\begin{bmatrix}1\\1\end{bmatrix}_e$. This should become $\begin{bmatrix}1\\0\end{bmatrix}_e$ because it's just our first basis vector $\vec{b}_1$. Let's check it:

$$\begin{bmatrix}-1 & 2\\1 & -1\end{bmatrix}_{Be}\begin{bmatrix}1\\1\end{bmatrix}_e = \begin{bmatrix}1\\0\end{bmatrix}_B$$

This process works for all dimensions, it just gets increasingly hard to invert matrices the higher you go. Keep in mind that we can only perform a change of basis when we have two bases that span the same space, and that space must be a Euclidian space (i.e. $\mathbb{R}^n$). Otherwise our change of basis would neither be able to be completed nor make any sense.

If you wanted to change a vector from basis B_1 to basis B_2, you use the e basis (or whichever basis is used to write down the basis vectors for B_1 and B_2) as the middle man. First, convert the the vector in B_1 to a vector in the e basis. Then, do a conversion from e to B_2. Both of these transformations are linear transformations, so they can be composed. Let's say U converts from e to B_1 and V converts from e to B_2. That means that you can convert from B_1 to B_2 the following way:

$$V_{B_2e}U^{-1}_{eB_1}\vec{x}_{B_1} = \vec{x}_{B_2}$$

We can also invert this composition like any other matrix

$$\left(V_{B_2e}U^{-1}_{eB_1}\right)^{-1}\vec{x}_{B_2} = U_{B_1e}V^{-1}_{eB_2}\vec{x}_{B_2} = \vec{x}_{B_1}$$

This is the exact result we would expect if we wanted to generate a matrix that converts from B_2 to B_1.

4.2.1 Why?

You may be quite rightly asking yourself, "Why would I ever want to stop using the e basis?" Well I'm glad you asked. There is no special or preferred basis in Mathematics. All the rules we've come up with work no matter which basis the vectors are written in. There are some problems where it is helpful to change your basis into one that can be more easily worked with. For example, in Quantum Mechanics, it is common to rewrite quantum states as linear combinations of the energy eigenstates of the system. This is done so that it is possible to perform additional analysis. We will go over an example that pertains to Linear Algebra in section 5.2 after we learn and understand eigenvalues and eigenvectors.

4.3 Inner Product

In previous math studies, you have probably heard of the dot product. The idea of the dot product gives us the idea of lengths of vectors. We are going to generalize that concept by introducting the *inner product*. The standard inner product (the dot product) of two column vectors of the same size $\vec{u}$ and $\vec{v}$ whose components are strictly real is written the following way:

$$\vec{u}^T\vec{v} = c.$$

Because of how multiplication of a row and column vector works, the dot product will always return a scalar. However, the dot product does not represent all inner products. To represent general inner products, we are going to introduce another notation for vectors:

Dirac's bra-ket notation. It is the standard notation used in Quantum Mechanics, and it is very useful to be exposed to. This involves, as the name suggests, bras (row vectors) and kets (column vectors). A bra is written the following way:

$$\langle u| = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix},$$

and a ket is written like this:

$$|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

These two notations for vectors is the same as writing $\vec{u}$ or $\vec{v}$, just we know that a ket is strictly a column vector and a bra is strictly row vector. Just as how "vector" usually refers to column vectors, "vector" when working with bra-ket notation, usually refers to kets. We can write the sum of vectors like this:

$$|u\rangle + |v\rangle = |u + v\rangle$$
$$\langle u| + \langle v| = \langle u + v|$$

A bra can always be converted to a ket by taking its complex conjugate transpose (represented using the † ("dagger") superscript).

$$|v\rangle = \langle v|^\dagger$$

$$|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \implies \langle v| = \begin{bmatrix} \overline{v}_1 & \overline{v}_2 & \cdots & \overline{v}_n \end{bmatrix}$$

where, for a complex number z, $\overline{z}$ is the complex conjugate of z.

$$z = a + bi \implies \overline{z} = a - bi$$

When dealing with vectors of strictly real components (as we have been and will continue to do in these notes), the complex conjugate transpose is just the transpose. However, for completeness sake, I will be using the complex conjugate transpose in this section. The inner product between the two (sometimes referred to as a "bracket") is written like this:

$$\langle u|v\rangle = (\langle u|)\,(|v\rangle) = |u\rangle^\dagger\,|v\rangle = c,$$

and it will always return a scalar value for the same reason the dot product always returns a scalar. The inner product is a linear operator (as with most products). That means it satisifes the following properties:

$$\langle x|y_1 + y_2\rangle = \langle x|y_1\rangle + \langle x|y_2\rangle$$
$$\langle x|cy\rangle = c\,\langle x|y\rangle$$

The inner product also has conjugate symmetry:

$$\langle x|y\rangle = \overline{\langle y|x\rangle},$$

where the bar represents the complex conjugate. The inner product of two vectors with real components is symmetric:

$$\langle x|y\rangle = \langle y|x\rangle\,, \quad |x\rangle\,,|y\rangle \in \mathbb{R}^n$$

The inner product of a non-zero vector with itself is positive definite[12]:

$$\langle x|x\rangle > 0 \iff |x\rangle \neq |0\rangle$$

When the inner product of a vector with itself is 0, we know that the vector itself must be the zero vector.

$$\langle x|x\rangle = 0 \implies x = |0\rangle$$

We can define any inner product that has these properties for any vector space we want. When we extend a Euclidean space to also include an inner product, that space is called a *Hilbert space.*

4.3.1 What the Inner Product Means

The most important thing about an inner product is that it gives us a definition of magnitude. The magnitude of a vector $|x\rangle$ is always given by the following

$$\left||x\rangle\right|^2 = \langle x|x\rangle$$
$$\left||x\rangle\right| = \sqrt{\langle x|x\rangle}$$

The inner product also gives us a definition of orthogonality. When the inner product of two non-0 vectors is 0, those vectors are said to be orthogonal.

$$\langle u|v\rangle = 0 \implies |u\rangle \perp |v\rangle$$

4.3.2 Various Examples of "Inner Products" in Real Vector Spaces

We can define any inner product between two vectors as long as they have the properties we defined earlier. We are going to limit ourselves to only real vector spaces here for simplicity. It is unlikely you will be asked about inner products of complex vector spaces in an introductory Linear Algebra course. Let's briefly summarize the required properties of a inner products of real vector spaces:

1. Linearity: $\langle x|y_1 + y_2\rangle = \langle x|y_1\rangle + \langle x|y_2\rangle$ and $\langle x|cy\rangle = c\,\langle x|y\rangle$.
2. Positive definite: $\langle x|x\rangle > 0$ iff $|x\rangle \neq |0\rangle$.
3. Symmetric: $\langle x|y\rangle = \langle y|x\rangle$.

The Standard Inner Product

$$\langle x|y\rangle \coloneqq \sum_{i=1}^{n} x_i y_i$$

Let's check each of the properties:

[12]The term "positive definite" means greater than 0. "Positive semi-definite" means greater than or equal to 0.

1. Linearity:

$$\begin{aligned}\langle x|cy + \tilde{c}\tilde{y}\rangle &= \sum_{i=1}^{n} x_i(cy_i + \tilde{c}\tilde{y}_i) \\ &= \sum_{i=1}^{n} [x_i c y_i + x_i \tilde{c}\tilde{y}_i] \\ &= c\sum_{i=1}^{n} [x_i y_i] + \tilde{c}\sum_{i=1}^{n} [x_i \tilde{y}_i] \\ &= c\,\langle x|y\rangle + \tilde{c}\,\langle x|\tilde{y}\rangle\end{aligned}$$

   So we can see that the property of linearity is satisfied.

2. Positive definite:

$$\langle x|x\rangle = \sum_{i=1}^{n} x_i x_i = \sum_{i=1}^{n} (x_i)^2$$

   Clearly then, $\langle x|x\rangle$ will always be positive when $|x\rangle$ is non-0, and it will only be 0 when $|x\rangle$ is 0.

3. Symmetric:

$$\langle x|y\rangle = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i = \langle y|x\rangle$$

   So this inner product is symmetric. We now know that this inner product (the standard inner product) is valid.

Broken Inner Products Let's look at some inner products that don't work.

1. Violating Linearity:

$$\langle x|y\rangle \coloneqq \sum_{i=1}^{n} x_i y_i^2$$

   I'm going to use the same test for linearity we used before, but I will set the constants

c and $\tilde{c}$ to 1 for simplicity.

$$\begin{aligned}
\langle x|y+\tilde{y}\rangle &= \sum_{i=1}^{n} x_i \left[y_i + \tilde{y}_i\right]^2 \\
&= \sum_{i=1}^{n} x_i \left[y_i^2 + 2y_i\tilde{y}_i + \tilde{y}^2\right] \\
&= \sum_{i=1}^{n} x_i y_i^2 + 2\sum_{i=1}^{n} x_i y_i \tilde{y}_i + \sum_{i=1}^{n} x_i \tilde{y}_i^2 \\
&= \langle x|y\rangle + 2\sum_{i=1}^{n} x_i y_i \tilde{y}_i + \langle x|\tilde{y}\rangle \\
&= \langle x|y\rangle + \langle x|\tilde{y}\rangle + 2\sum_{i=1}^{n} x_i y_i \tilde{y}_i \\
&\neq \langle x|y\rangle + \langle x|\tilde{y}\rangle
\end{aligned}$$

Linearity requires that each component appears only as a linear term, so it makes sense that the squared y terms cause this “inner product” to not work. This “inner product” turns out to fail to satisfy all properties required of a valid inner product.

2. Violating Positivity:

$$\langle x|y\rangle := \sum_{i=1}^{n} x_i$$

It doesn’t take much thought to see that this inner product isn’t positive definite. Let’s say we are in $\mathbb{R}^4$. Let’s find an example of a non-zero vector $|x\rangle$ such that $\langle x|x\rangle$ is 0.

$$|x\rangle = \begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix} \implies \langle x|x\rangle = -2+0+2+0 = 0$$

3. Violating Symmetry: It can be more easily demonstrated that a linear “inner product” fails symmetry if we limit ourselves to a certain vector space. We will choose $\mathbb{R}^3$.

$$\langle x|y\rangle := x_1y_3 + 2x_2y_2 + 3x_3y_1$$

$$\begin{aligned}
\langle x|y\rangle &= x_1y_3 + 2x_2y_2 + 3x_3y_1 \\
\langle y|x\rangle &= y_1x_3 + 2y_2x_2 + 3y_3x_1 \\
&= 3x_1y_3 + 2y_2x_2 + x_3y_1 \\
&\neq \langle x|y\rangle
\end{aligned}$$

It is important to keep in mind that, if the coefficients were all 1, symmetry wouldn’t be violated.

4.3.3 Projection and Rejection

5 Eigenvalues and Eigenvectors

5.1 Definitions

Given an $n \times n$ matrix A, $\vec{u}$ is an eigenvector associated with eigenvalue λ (scalar) when the following equation holds:

$$A\vec{u} = \lambda\vec{u} \tag{5.1}$$

However, $\vec{u}$ *cannot* be $\vec{0}$. This is just a restriction we place, and it will be clear why we do this later on. Also, we can scale $\vec{u}$ by any non-0 number to find another eigenvector.

$$A(c\vec{u}) = cA\vec{u} = c\lambda\vec{u} = \lambda(c\vec{u})$$

So we know that if we find one eigenvector, we can just as easily say that it forms a basis for an eigenspace associated with eigenvalue λ. But how do we solve for the eigenvalues and eigenvectors? We can perform some additional analysis on equation (5.1).

$$\begin{aligned} A\vec{u} &= \lambda I_n\vec{u} \\ A\vec{u} - \lambda I_n\vec{u} &= 0 \\ [A - \lambda I_n]\vec{u} &= 0 \end{aligned} \tag{5.2}$$

Now we know that, if $\vec{u}$ must be non-0, $A - \lambda I_n$ must have a non-trivial null space. By the definition of the determinant, we also know that $\det(A - \lambda I_n)$ cannot be equal to 0. You can solve for the eigenvalues by checking to see which values for λ make the determinant of $A - \lambda I_n$ non-0. Let's do an example:

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

Let's find the determinant:

$$\begin{vmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda - 1 & 1 & 1 \\ 0 & -\lambda & 1 \\ 1+\lambda & 1 & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda - 1 & 1 & 1 \\ 0 & -\lambda & 1 \\ 0 & 2 & -\lambda + 1 \end{vmatrix} = 0$$

$$(-\lambda - 1)(-\lambda(-\lambda + 1) - 2) = 0 \tag{5.3}$$

Equation (5.3) is what is referred to as the *characteristic equation* for A. In other words, the characteristic equation is just $\det(A - \lambda I_n) = 0$. The eigenvalues are the *real* roots of the characteristic equation. Because there are characteristic equations where only complex roots exist, a matrix is not always guaranteed to have any eigenvalues. The two-dimensional rotation matrix is an example of this. The *characteristic polynomial* is just the left hand side of the characteristic equation: $\det(A - \lambda I_n)$. Now let's keep going, solving for λ.

$$\begin{aligned} (-\lambda - 1)(-\lambda(-\lambda + 1) - 2) &= 0 \\ (\lambda + 1)(-\lambda(-\lambda + 1) - 2) &= 0 \\ (\lambda + 1)(\lambda^2 - \lambda - 2) &= 0 \\ (\lambda + 1)(\lambda - 2)(\lambda + 1) &= 0 \\ (\lambda + 1)^2(\lambda - 2) &= 0 \end{aligned}$$

Now we can see that we have two eigenvalues: -1 and 2. It is very important to notice that the eigenvalue -1 is said to have multiplicity 2. A 3×3 matrix can have up to three eigenvalues. This is because each eigenspace associated with distinct eigenvalues must be at least one-dimensional, so we can have up to three of them. However, sometimes we do not have three distinct eigenvalues as in this example. Eigenspaces can have a dimension of up to the multiplicty of its associated eigenvalue. So the eigenspace associated with $\lambda = -1$ can have either dimension 1 or 2. Looking back at equation (5.2), we know that the eigenspace associated with eigenvalue λ is the nullspace of matrix $A - \lambda I_n$. Let's solve this for $\lambda = -1$ and $\lambda = 2$.

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

Starting with $\lambda = -1$:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \implies x_3 = s,\ x_2 = t,\ x_1 = -s - t$$

So we see that the eigenspace associated with $\lambda = -1$ is

$$\text{span}\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

We see that the dimension of this space is 2, not 1. Keep in mind, the multiplicity of $\lambda = -1$ does *not* determine the dimension of its associated eigenspace, just its eigenspace's upper bound. Now let's solve for the eigenspace associated with $\lambda = 2$. We know it will have dimension 1.

$$\left[\begin{array}{ccc|c} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & 1 & -2 & 0 \end{array}\right] \xrightarrow[R_3 \to 2R_3 + R_1]{R_2 \to 2R_2 + R_1} \left[\begin{array}{ccc|c} -2 & 1 & 1 & 0 \\ 0 & -3 & 3 & 0 \\ 0 & 3 & -3 & 0 \end{array}\right] \xrightarrow{R_3 \to R_3 + R_2}$$

$$\left[\begin{array}{ccc|c} -2 & 1 & 1 & 0 \\ 0 & -3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1 \to 3R_1 + R_2} \left[\begin{array}{ccc|c} -6 & 0 & 6 & 0 \\ 0 & -3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This gives us the following solutions:

$$x_3 = s,\ x_2 = s,\ x_1 = s,$$

then the eigenspace is

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

We can check that this is right by putting it back into the original equation (5.1).

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

I hope this makes you believe that this process works. Combining our eigenspaces, we create an *eigenbasis* for the range of A. This gives us an eigenbasis of

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Notice that the eigenbasis for A is three-dimensional, the same dimension as its codomain.

5.1.1 Triangular Matrices

Because we solve for eigenvalues of a matrix by taking the determinant of matrix $A - \lambda I_n$, we know that, if the matrix is triangular, the eigenvalues are just the diagonal elements. Remember, we can easily take the determinant of a triangular matrix: it's the product of the diagonal elements. Consider the following matrix as an example:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\det(A - \lambda I_3) = \begin{vmatrix} 1-\lambda & 1 & 2 \\ 0 & 3-\lambda & 4 \\ 0 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda)(2-\lambda) = 0$$

$$\implies \lambda_1 = 1,\ \lambda_2 = 3,\ \lambda_3 = 2$$

5.2 Diagonalization

We can say that a matrix A is diagonalizable if there exist a diagonal matrix D and invertible matrix P such that the following relation holds:

$$A = PDP^{-1}$$

How do we find what P and D are for a given A? Let's consider the properties of a diagonal matrix:

$$D = \begin{bmatrix} d_{11} & 0 & 0 & \cdots & 0 \\ 0 & d_{22} & 0 & \cdots & 0 \\ 0 & 0 & d_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_{nn} \end{bmatrix} \qquad D\vec{x} = \begin{bmatrix} d_{11}x_1 \\ d_{22}x_2 \\ d_{33}x_3 \\ \vdots \\ d_{nn}x_n \end{bmatrix}$$

Now let's say matrix A has eigenvectors that form a basis for $\mathbb{R}^n$. That means we can write every vector x in $\mathbb{R}^n$ as a vector in the eigenvector basis, $\mathscr{B}$.

$$\mathscr{B} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$$

$$\vec{x} = c_1\vec{u}_1 + c_2\vec{u}_2 + \ldots + c_n\vec{u}_n \implies \vec{x}_{\mathscr{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_{\mathscr{B}}$$

Because each of the basis vectors are eigenvectors of matrix A, we can write the following relation:

$$\begin{aligned} A\vec{x} &= Ac_1\vec{u}_1 + Ac_2\vec{u}_2 + \ldots + Ac_n\vec{u}_n \\ &= \lambda_1 c_1\vec{u}_1 + \lambda_2 c_2\vec{u}_2 + \ldots + \lambda_n c_n\vec{u}_n \\ &= \begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \\ \vdots \\ \lambda_n c_n \end{bmatrix}_{\mathscr{B}} \end{aligned}$$

We can rewrite this as the following:

$$D\vec{x}_{\mathscr{B}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots 0 \quad 0 \quad 0 \quad \cdots \quad \lambda_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{bmatrix}_{\mathscr{B}} = \begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \\ \vdots \\ \lambda_n c_n \end{bmatrix}_{\mathscr{B}} = A\vec{x}$$

This shows that we can compute $A\vec{x}$ by first changing our basis to the eigenbasis, performing the diagonal matrix multiplication, and converting back. We can define a matrix P from the eigenbasis to the standard basis just by using the eigenvectors of A:

$$P_{\mathscr{B}} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots \vec{u}_n \end{bmatrix}$$

Similarly, we can invert this matrix to get us back from the standard basis to the eigenbasis. This means that we can write the following relation:

$$\vec{x}_B = P_{\mathscr{B}}^{-1}\vec{x}$$

This means that

$$A\vec{x} = P_{\mathscr{B}} D P_{\mathscr{B}}^{-1}\vec{x},$$

or, more concisely, that

$$A = PDP^{-1}$$

where P is the matrix of linearly indepedent eigenvectors associated with A with matching eigenvalues in D. Keep in mind, P must have as many columns as A, so the eigenvectors of A must span $\mathbb{R}^n$, and this may not happen. When A does not have a large enough eigenspace, it is said to not be diagonalizable.

Small Example Given A has the following eigenvalues and eigenvectors:

$$\lambda_1 = 3,\ \lambda_2 = \lambda_3 = -1,\ \vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ \vec{u}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix},\ \vec{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix}$$

This gives us the following equations for P and D:

$$P = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 3 & 5 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Order *does* matter here, except in the case of repeated eigenvalues (the second and last columns of P could be flipped, and the diagonalization would stay the same).

5.2.1 Matrix Powers

We can use this diagonalized form of matrix A to do special things.

$$A = PDP^{-1}$$

We can easily find A^2:

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD^2P^{-1} \end{aligned}$$

What about A^3?

$$\begin{aligned} A^3 &= (PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= (PD^2P^{-1})(PDP^{-1}) \\ &= PD^3P^{-1} \end{aligned}$$

In general we get the following equation for A^n:

$$A^n = PD^nP^{-1}$$

This makes sense because the basis in which A acts doesn't make a difference: we convert once, apply the transformation many times, and then convert back when we're done. Converting to the eigenbasis makes it easy for us to apply a transformation multiple times. We can also take roots of matrices.

6 Beyond the Curriculum

6.1 Index Notation