Course Notes for Linear Algebra

Remkes Kooistra
The King's University

November 21, 2018

Contents

License

Welcome to the Course

This courses uses a pedagogical model known either as blended learning or the flipped classroom. In this model, the majority of the content will be delivered outside of class time through two avenues: these notes and a series of short videos. You will be required to watch one of the short videos before most class periods. We will begin each lecture period assuming you have watched the required video.

The videos are used to introduce the main ideas of the course. They are the explanation. Matched with each video will be a short section of notes. The notes are provided for reference, so that after you've watched the video, you can use the notes to remind yourself of the content and refer to useful ideas, concepts and formulae. The notes are not primarily written to explain the material; instead, they are written to provide a record of the ideas in the video for your reference. The notes are light on examples. The lecture time will be mostly devoted to necessary practice and examples.

The course is organized into twenty-two lectures; the videos and the activities are numbered to match these lectures. There is a detailed schedule on the course website showing when the various lectures happen over the term. Please use this schedule to ensure you watch the appropriate videos before class and bring the appropriate sections of the notes.

Other Resources

In addition to the notes associated to each lecture, there are a number of other resources which are distributed via the course website. These resources have been developed to assist you, so please make use of them.

- A complete formula sheet. All of the basic formulas and rules for calculation are included on this sheet. There are sections on algebra, trigonometry, derivatives and integrals. Later in the document are some pieces of linear algebra reference as well.
- A notation reference. This reference covers a number of notations used in the course which may be unfamiliar.
- A course outcomes sheet. This was originally developed as a study aid for students. It summarizes the main definitions and concepts of the course, as well as the types of questions your will encounter on assignments and exams. In particular, it gives a guide to the material on the exam. If you want to know whether a definition, topic or type of problem might show up on the exam, consult this sheet.

Perspectives and Philosophy

In addition to the strictly mathematical material of the course, I will try to share some ideas which give perspective and context to mathematics. This include the philosophy of mathematics, the aesthetics of mathematics, and how our own worldview and assumptions influence mathematical thought. Linear algebra is perhaps the most abstract mathematical course offered at King's; at least, it offers the most oppotunity for abstract thinking. Therefore, I use this course as a venue for talking about abstraction in the context of perspectives and philosophy.

Lecture 1

Themes of the Course

1.1 Algebra

Linear *algebra* is, as expected, an algebra course. (It is the only strictly algebra course taught at King's, though elsewhere, it would be the first of a sequence of algebra courses.)

The word 'algebra' does not mean the same thing in university mathematics as it did in high-school mathematics. Previously, the world 'algebra' usually signaled the use of variables: $3+5 = 8$ is arithmetic, but $3 + x = 8 \implies x = 5$ is algebra. From that point, algebra became the manipulation of equations (and inequalities) with variables, particularly focusing on solving polynomials and rational functions.

In academic mathematics, algebra has a much broader and deeper sense. It is a whole branch of mathematics, with many subdisciplines: linear algebra, abstract algebra, homological algebra, etc. One of the goals of this course will be to build as sense of what we mean by the term 'algebra'. I'll give a definition right now, at the start, but it's a difficult definition to understand; hopefully the course will provide the necessary elaboration.

Definition 1.1.1. *Algebra* is the study of sets with structure. It investigates possible structures on sets, the rules obeyed by those structures, and the interaction of various structures.

1.2 Abstraction

The definition of algebra just stated is an immensely abstract definition. I haven't specified what sets or what structures I'm talking about. (I could say that we are looking at *algebraic structures*, but that doesn't actually add any insight; it just delays the question.) The definition is very broad, applies to a huge variety of sets and structures. Abstraction is built into the very nature of algebra.

In the calculus courses, I often motivate the material by its applications. (The two most often used motivations are the Newtonian physics and percentage growth in ecology or finance.) In this sense, I'm treating the calculus as applied mathematics: mathematics driven by and designed for the solving of certain extra-mathematical problems. This is a reasonable approach for calculus, both for historical and pedagogical reasons.

In this course, there will be points where I emphasize the practical applications of linear algebra, of which there are many. However, my primary motivation for the course will be intrinsic and abstract: I want to study sets and their structure simply for the mathematical joy of it. This is the one course at King's which is most suited to presenting the goals and ideas of abstract mathematics, in and of themselves.

1.3 Geometry

Though the subjet is called 'algebra', linear algebra is a very geometric discpline. Given my background as a geometer, my approach to the course will be heavily geometric. We will start with vectors, which is a geometric discpline, though it has rich algebraic structre. Throughout the course, we will play the geometry and algebra against each other and use each to understand the other better. This interplay drives the discpline. In many ways, this will be a geomety course hiding under the title of an algebra course.

In addition, we will not restrict ourselves for conventional two and three dimensional geometry. The definitions of this course are very general, so we will try to do geometry in any (finite) dimension. A major challenge of the course is going from familiar, visible three dimensions to un-visualizable higher dimensional geometry.

1.4 Flatness

While a geometry course, the geometry we are considering will be very restricted. The 'linear' in linear algebra refers to lines, so we will start with the geometry of lines. However, the important property of a line, in our context, is the fact that it is straight: it doesn't curve. Linear algebra is the geomtery of flat things.

Therefore, curved, bent and broken objects are off limits. We will investigate the geometry of flat objects: points, lines, planes and higher dimensional analogue. A great advantage of this restriction is the fact that flat things are very accessable. The geometry of curved objects is an immensely complicated and difficult subject. By restricting to flat objects, we can produce a wide range of accessible mathematical results.

1.5 Symmetry

Other than flatness, our other major geometric theme is symmetry. Many of the definitions and propositions produced in this course can be stated in terms of some kind of symmetry. Speaking broadly, I want to thing of symmetry as preservation: the symmetry of a shape is some kind of operation which keeps the shape intact, preserves it. Throughout linear algebra, we will be talking about the preservation of various kinds of shapes and flat objects.

With the interplay between algebra and geometry mentioned above, we will also translate this idea of symmetry back to algebra to give a understanding of algebraic symmetry.

1.6 Proof

Proofs are another major difference between the grade-school understanding of algebra and the academic sense of the word. Algebra is originally introduced to solve problems involving unknown quantities, usually by solving equations. Algebra is a problem-solving technique. In academic mathematics, we are usually more concerned with conjectures about mathematical objects and the proofs of those conjecture. In this course, we'll introduce the idea of conjectures, propositions, theorems and proofs. We'll do a small amount of proving on the assignments to build a flavour for proof-based mathematics.

Lecture 2

Vectors

2.1 Definitions

Definition 2.1.1. In linear algebra, ordinary numbers (intergers, rational numbers or real numbers) are called *scalars*.

Definition 2.1.2. A *vector* is a finite ordered list of scalars. Vectors can be written either as columns or rows. If a vector is a list of the numbers 4, 15, π and $-e$ (in that order), we write the vector in one of two ways.

$$\begin{pmatrix} 4 \\ 15 \\ \pi \\ e \end{pmatrix} \qquad \text{or} \qquad (4, 15, \pi, -e)$$

In these notes, we will exclusively use column vectors.

Definition 2.1.3. Let n be a positive integer. *Real Euclidean Space* or *Cartesian Space*, written $\mathbb{R}^n$, is the set of all vectors of length n with real number entries. An arbitrary element of $\mathbb{R}^n$ is written as a column.

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

We say that $\mathbb{R}^n$ has *dimension* n.

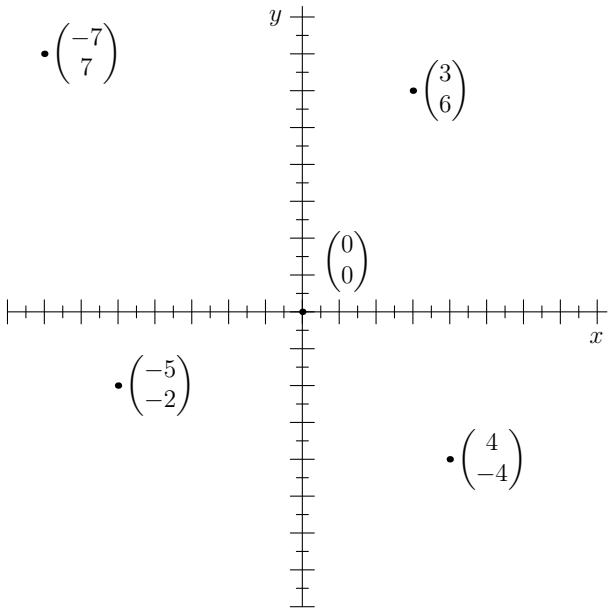

Figure 2.1: Points in the Cartesian Plane $\mathbb{R}^2$

Definition 2.1.4. The scalars x_i in a vector are called the *entries*, *coordinates* or *components* of that vector. Specifically, x_1 is the first coordinate, x_2 is the second coordinate, and so on. For $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^4$, we use the letters w, x, y, z instead of x_i.

$$\begin{pmatrix}x\\y\end{pmatrix} \in \mathbb{R}^2 \qquad \begin{pmatrix}x\\y\\z\end{pmatrix} \in \mathbb{R}^3 \qquad \begin{pmatrix}w\\x\\y\\z\end{pmatrix} \in \mathbb{R}^4$$

Definition 2.1.5. In any $\mathbb{R}^n$, the *origin* is the unique point given by a vector of zeros. It is also called the zero vector. It is considered the centre point of Cartesian space.

$$\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}$$

Cartesian space, particularly the Cartesian plane, is a familiar object from high-school mathematics. We usually visualize Cartesian space by drawing axes, one in each independent perpendicular direction. In this visualization, the vector $\begin{pmatrix}a\\b\end{pmatrix}$ corresponds to the unique point we get moving a units in the direction of the x axis and b units in the direction of the y axis. Figure 2.1 shows the location of several points in $\mathbb{R}^2$.

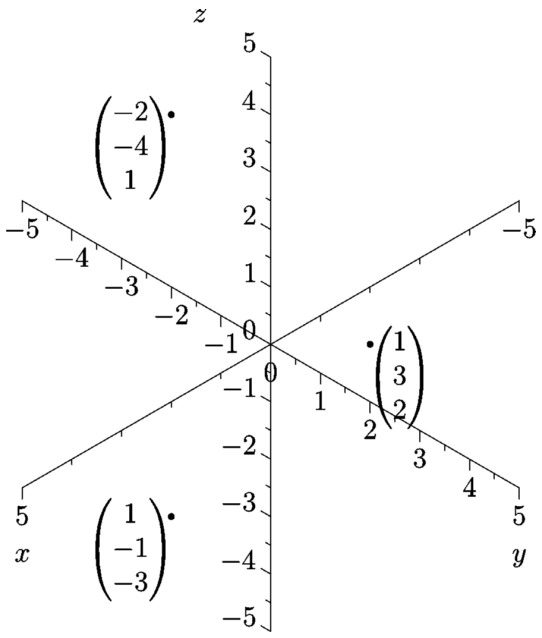

Figure 2.2: Points in Cartesian three-space $\mathbb{R}^3$

As with $\mathbb{R}^2$, the point $\begin{pmatrix}a\\b\\c\end{pmatrix} \in \mathbb{R}^3$ is the unique point we find by moving a units in the x direction, b units in the y direction and c units in the z direction. When we visualize $\mathbb{R}^2$, we conventionally write the x axis horizontally, with a positive direction to the right, and the y axis vertically, with a positive direction upwards. For $\mathbb{R}^3$, the x and y axes form a flat plane and the z axis extend vertically from that plan, as shown in Figure 2.2. Notice, in both cases, we needed to choose directions for the axes.

Definition 2.1.6. A choice of axis directions in a visualization of $\mathbb{R}^n$ is called an *orientation*.

While we can visualize $\mathbb{R}^2$ and $\mathbb{R}^3$ relatively easily and efficiently, but we can't visualize any higher $\mathbb{R}^n$. However, this doesn't prevent us from working in higher dimensions. We need to rely on the algebraic descriptions of vectors instead of the drawings and visualizations of $\mathbb{R}^2$ and $\mathbb{R}^3$.

In our visualizations of $\mathbb{R}^2$ and $\mathbb{R}^3$, we see the different axes as fundementally different perpendicular directions. We can think of $\mathbb{R}^2$ as the space with two independent directions and $\mathbb{R}^3$ as the space with three independent directions. Similarly, $\mathbb{R}^4$ is the space with four perpendicular, independent directions, even though it is impossible to visualize such a thing. Likewise, $\mathbb{R}^n$ is the space with n independent directions.

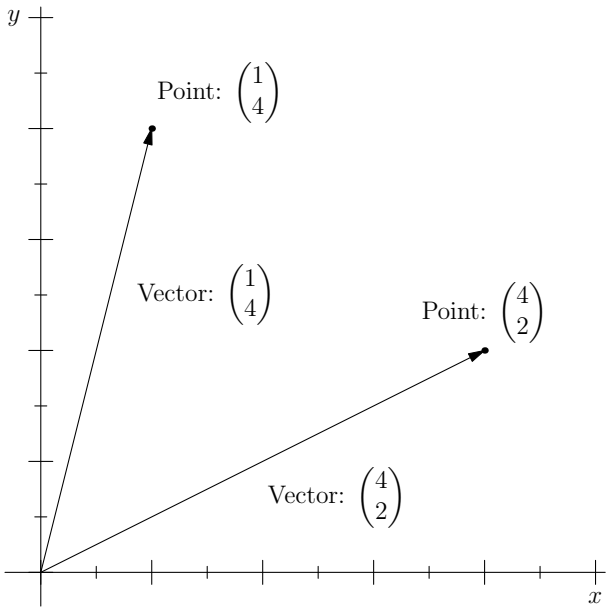

Figure 2.3: Vectors as Points and Directions

2.2 Points or Directions?

We can think of a element of $\mathbb{R}^2$, say $\begin{pmatrix}1\\4\end{pmatrix}$, as both the point located at $\begin{pmatrix}1\\4\end{pmatrix}$ and the vector drawn from the origin to the point $\begin{pmatrix}1\\4\end{pmatrix}$, as shown in Figure 2.3. Though these two ideas are distinct, we will frequently change perspective between them. Part of becoming proficient in vector geometry is becoming accustomed to the switch between the perspectives of points and directions.

2.3 Linear Operations

The environment for linear algebra is $\mathbb{R}^n$. Algebra is concerned with operations on sets, so we want to know what operations we can perform on $\mathbb{R}^n$. There are several.

Definition 2.3.1. The *sum* of two vectors u and v in $\mathbb{R}^n$ is the sum taken componentwise.

$$u + v = \begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix} + \begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix} = \begin{pmatrix}u_1 + v_1\\u_2 + v_2\\\vdots\\u_n + v_n\end{pmatrix}$$

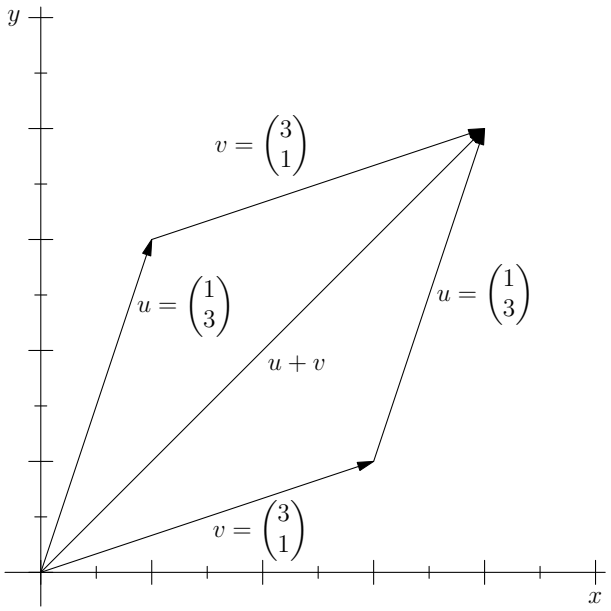

Figure 2.4: Visualizing Vector Addition

The sum is visuzliazed by placing the start of the second vector at the end of the first, as in Figure 2.4. Note that we can only add two vectors in the same dimension. We can't add a vector in $\mathbb{R}^2$ to a vector in $\mathbb{R}^3$.

Definition 2.3.2. If u is a vector in $\mathbb{R}^n$ and $a \in \mathbb{R}$ is a real number, then the *scalar multiplication* of u and a is multiplication by a in each component of u. By convention, scalar multiplication is written with the scalar on the left of the vector.

$$au = a \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} au_1 \\ au_2 \\ \vdots \\ au_n \end{pmatrix}$$

Though there will be other 'multiplications' to come, we generally say that we can't multiply vectors together in any way reminiscent of numbers. Instead, we can only multiply by scalars. Scalar multiplication is visualizing by scaling the vector by the value of the scalar. (Hence the term 'scalar'!) If the scalar is negative, the direction is also reversed, as in Figure 2.5.

Scalar multiplication also lets us define the difference between vectors.

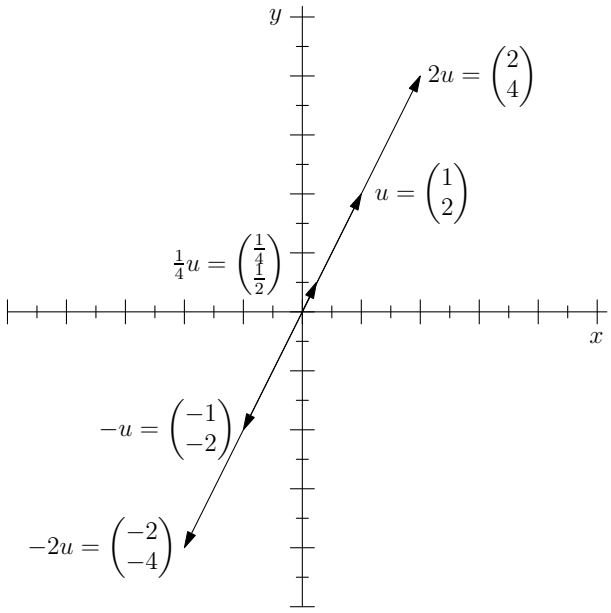

Figure 2.5: Visualizing Scalar Multiplication

Definition 2.3.3. The difference between two vectors u and v is the vector $u + (-1)v$, defined using addition and scalar multiplication. This works out to be componentwise subtraction.

$$u - v = u + (-1)v = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + (-1) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}$$

Definition 2.3.4. With respect to some set of scalars (such as $\mathbb{R}$), whenever we find a mathematical structure which has the two properties of addition and scalar multiplication, we call the structure *linear*. $\mathbb{R}^n$ is a linear space, because vectors allow for addition and scalar multiplication.

Definition 2.3.5. The *length* of a vector u in $\mathbb{R}^n$ is written $|u|$ and is given by a generalized form of the Pythagorean rule for right triangles.

$$|u| = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$$

This length is also called the *norm* of the vector. A vector of length one is called a *unit vector*.

If we think of vectors as directions from the origin towards a point, this definition of length gives exactly what we expect: the physical length of that arrow in $\mathbb{R}^2$ and $\mathbb{R}^3$. Past $\mathbb{R}^3$, we don't have a natural notion of length. This definition serves as a reasonable generalization to $\mathbb{R}^4$ and higher dimensions

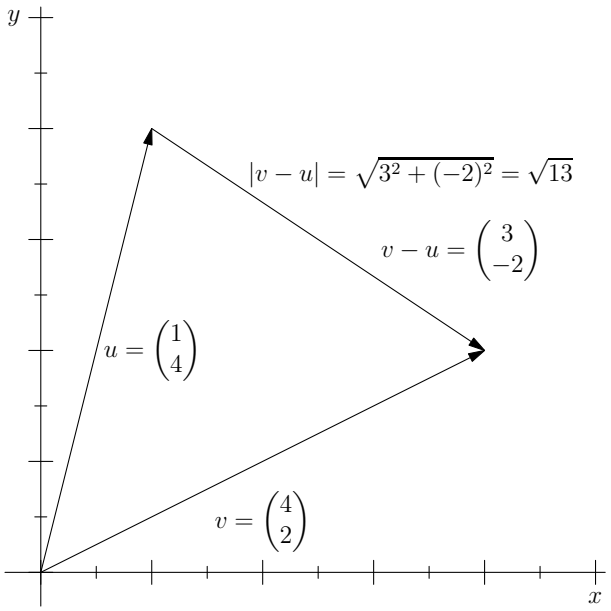

Figure 2.6: Visualizing Distance Between Vectors

which we can't visualize. Note also that $|u| = 0$ only if u is the zero vector. All other vectors have positive length.

Often the square root is annoying and we find it convenient to work with the square of length.

$$|u|^2 = u_1^2 + u_2^2 + \ldots + u_n^2$$

The notions of length and difference allow us to define the distance between two vectors.

Definition 2.3.6. The *distance between two vectors* u and v in $\mathbb{R}^n$ is the length of their difference: $|u - v|$.

You can check from the definition that $|u - v| = |v - u|$, so distance doesn't depend on which comes first. If $|\cdot|$ were absolute value in $\mathbb{R}$, this definition would match the notion of distance between numbers on the number line. Difference and length are visualized in Figure 2.6.

Proposition 2.3.7. *We briefly state two properties of vector lenghts without proof.*

$$|u + v| \leq |u| + |v| \qquad \textit{Triangle Inequality}$$
$$|au| = |a||u|$$

The last line deserves some attention for the notation. When we write $|a||u|$, $|a|$ is an absolute value of a real number and $|u|$ is the length of a vector. The fact that they have the same notation is frustrating, but these notations are common. (Some text use double bars for the length of a vector, $||v||$, to avoid this particular issue).

Lecture 3

The Dot Product

3.1 Definition

Earlier we said that we can't multiply two vectors together. That's mostly true, in the sense that there is no general product of two vectors uv which is still a vector. However, there are other kinds of 'multiplication' which combine two vectors. The operation defined in this lecture multiplies two vectors, but the result is a scalar.

Definition 3.1.1. The *dot product* or *inner product* or *scalar product* of two vectors u and v is given by the following formula.

$$u \cdot v = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1v_1 + u_2v_2 + \ldots + u_nv_n$$

We can think of the dot product as a scalar measure of the similarity of direction between the two vectors. If the two vectors point in a similar direction, their dot product is large, but if they point in very different directions, their dot product is small. However, we already have a measure, at least in $\mathbb{R}^2$, of this difference: the angle between two vectors. Thankfully, the two measures of difference agree and the dot product can be expressed in terms of angles.

Definition 3.1.2. The *angle* θ between two non-zero vectors u and v in $\mathbb{R}^n$ is given by the equation

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

This definition agrees with the angles in $\mathbb{R}^2$ and $\mathbb{R}^3$ which we can visualize. However, this serves as a new definition for angles between vectors in all $\mathbb{R}^n$ when $n \geq 4$. Since we can't visualize those spaces, we don't have a way of drawing angles and calculating them with conventional trigonometry. This definition allows us to extend angles in a completely algebraic way. Notes that $\theta \in [0, \pi]$, since we always take the smallest possible angle between two vectors.

Definition 3.1.3. Two vectors u and v in $\mathbb{R}^n$ are called *orthogonal* or *perpendicular* or *normal* if $u \cdot v = 0$.

3.2 Properties of the Dot Product

There are many pairs of orthogonal vectors. Thinking of the dot product as a multiplication, we have uncovered a serious difference between the dot product and conventional multiplication of numbers. If $a, b \in \mathbb{R}$ then $ab = 0$ implies that one of a or b must be zero. For vectors, we can have $\binom{1}{0} \cdot \binom{0}{1} = 0$ even though neither factor in the product is the zero vector. We have a definition to keep track of this new property.

Definition 3.2.1. Assume A is a set with addition and some kind of multiplication and that $0 \in A$. If $u \neq 0$ and $\neq 0$ but $uv = 0$, we say that u and v are *zero divisors*.

An important property of ordinary numbers, as we just noted, is that there are no zero divisors. Other algebraic structures, such as vectors with the dot product, may have many zero divisors.

Now that we have a new operation, it is useful to see how it interact with previously defined structure. The following list shows some interactions between addition, scalar multiplication, length and the dot product. Some of these are easy to establish from the definition and some take more work.

Proposition 3.2.2. *Let u, v, w be vectors in $\mathbb{R}^n$ and let a be a scalar in $\mathbb{R}$.*

$u + v$	$= v + u$	*Commutative Law for Vector Addition*
$a(u + v)$	$= au + av$	*Distributive Law for Scalar Multiplication*
$u \cdot v$	$= v \cdot u$	*Commutative Law for the Dot Product*
$u \cdot u$	$= \lvert u \rvert^2$	
$u \cdot (v + w)$	$= u \cdot v + u \cdot w$	*Distributive Law for the Dot Product*
$u \cdot (av)$	$= (au) \cdot v = a(u \cdot v)$	

In $\mathbb{R}^2$, norms and dot products allow us to recreate some well-known geometric constructions. For example, now that we have lengths and angles, we can state the cosine law in terms of vectors. The visualization of the vector relationships of the cosine law is shown in Figure 3.1.

Proposition 3.2.3 (The Cosine Law)**.** *Let u and v be vectors in $\mathbb{R}^n$.*

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

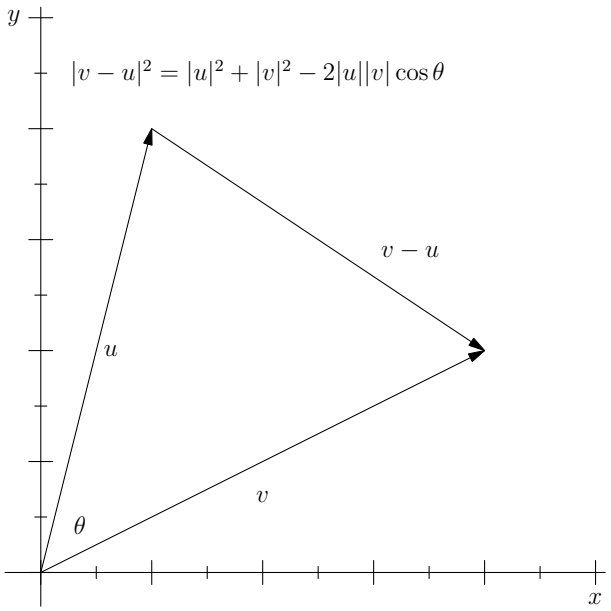

Figure 3.1: The Cosine Law

Lecture 4

More Vector Structures

4.1 The Cross Product

The dot product is an operation which can be performed on any two vectors in $\mathbb{R}^n$ for any $n \geq 1$. There are no other conventional products that work in all dimensions. However, there is a special product that works in three dimensions.

Definition 4.1.1. Let $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ be two vectors in $\mathbb{R}^3$. The *cross product* of u and v is written $u \times v$ and defined by the following formula.

$$u \times v = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}$$

The cross product differs from the dot product in several important ways. First, it produces a new vector in $\mathbb{R}^3$, not a scalar. For this reason, when working in $\mathbb{R}^3$, the dot product is often refered to as the *scalar product* and the cross product as the *vector product*. Second, the dot product measures, in some sense, the similarity of two vectors. The cross product measures, in some sense, the difference between two vectors. The cross product has greater magnitude if the vectors are closer to being perpendicular. If θ is the angle between u and v, the dot product was expressed in terms of $\cos\theta$. This measures similarity, since $\cos 0 = 1$. There is a similar identity for the cross product:

$$|u \times v| = |u||v| \sin\theta$$

This identity tells us that the cross product measures difference in direction, since $\sin 0 = 0$. In particular, this tells us that $|u \times u| = 0$, implying that $u \times u = 0$ (the zero vector is the only vector which has zero length). This is another new and strange property: in this particular multiplication,

everything squares to zero. The cross product is obviously very different from multiplication of scalars, where $a^2 = 0$ cannot happen unless $a = 0$.

Also consider the relationship between u and $u \times v$ as calculated through the dot product.

$$\begin{aligned} u \cdot (u \times v) &= \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \cdot \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix} \\ &= u_1u_2v_3 - u_1u_3v_2 + u_2u_3v_1 - u_2u_1v_3 + u_3u_1v_2 - u_3u_2v_1 = 0 \end{aligned}$$

A similar calculation shows that $v \cdot (u \times v) = 0$. Since a dot product of two vectors is zero if and only if the vectors are perpendicular, the vector $v \times u$ is perpendicular to both u and v. This turns out to be a very useful property of the cross product.

Finally, a calculation from the definition shows that $u \times v = -(v \times u)$. So far, multiplication of scalars and the dot product of vectors have not depended on order. The cross product is one of many products in mathematics which depends on order. If we change the order of the cross product, we introduce a negative sign.

Definition 4.1.2. Products which do not depend on the order of the factors, such as multiplication of scalars and the dot product of vectors, are called *commutative products*. Products where changing the order of the factors introduces a negative sign are caled *anti-commutative products*. The cross product is an anti-commutative product. Other products which have neither of these properties are called *non-commutative products*.

4.2 Angular Motion

An important application of the cross product is found in describing rotational motion. Linear mechanics describes the motion of an object through space but rotational mechanics describes the rotation of an object independent of its movement through space. A force on an object can cause both kinds of movement, obviously. The following table summarizes the parallel questions of linear motion and rotational motion in $\mathbb{R}^3$.

Linear Motion	Rotational Motion
Straight line in a vacuum	Continual spinning in a vacuum
Direction of motion	Axis of spin
Force	Torque
Momentum	Angular Momentum
Mass (resistance to motion)	Moment of Intertia (resistance to spin)
Velocity	Frequency (Angular Velocity)
Acceleration	Angular Acceleration

How do we describe torque? If there is a linear force applied to an object which can only rotate around an axis, and if the linear force is applied at a distance r from the axis, we can think of the force F and

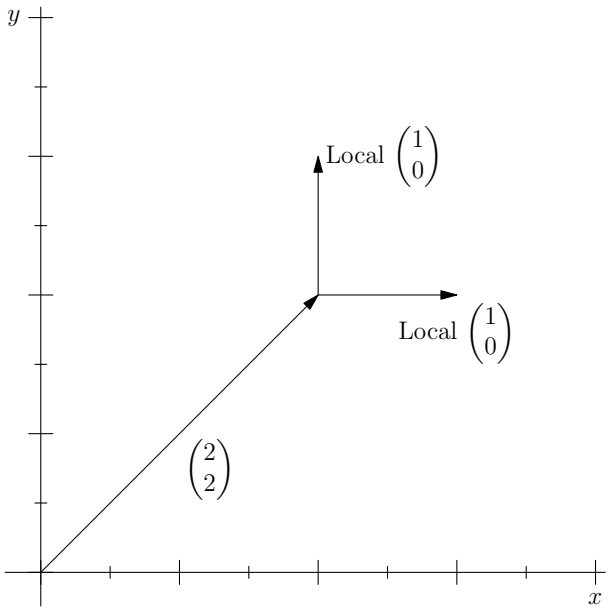

Figure 4.1: Local Direction Vectors

the distance r as vectors. The torque is then $\tau = r \times F$. Notice that $|\tau| = |r||F| \sin\theta$, indicating that linear force perpendicular to the radius gives the greatest angular acceleration. That makes sense. If F and r share a direction, then we are pushing directly along the axis and no rotation can occur.

The use of cross products in rotational dynamics is extended in many interesting ways. In fluid dynamics, local rotational products of the fluid result in turbulence, vortices and similar effects. Tornadoes and hurricanes are particularly extreme examples of vortices in the fluid which is our atmosphere. All the descriptions of the force and motion of these vortices involve cross products in the vectors describing the fluid.

4.3 Local Direction Vectors

We've already spoken about the distinction between elements of $\mathbb{R}^n$ as points and vectors. There is another important subtlety that shows up all throughout vector geometry. In addition to thinking of vectors as directions starting at the origin, we can think of them as directions starting anywhere in $\mathbb{R}^n$. We call these local direction vectors.

For example, as pictured in Figure 4.1, at the point $\begin{pmatrix}2\\2\end{pmatrix}$ in $\mathbb{R}^2$, we could think of the local directions $\begin{pmatrix}1\\0\end{pmatrix}$ or $\begin{pmatrix}0\\1\end{pmatrix}$. These are not directions starting from the origin, but starting from $\begin{pmatrix}2\\2\end{pmatrix}$ *as if that were the origin.*

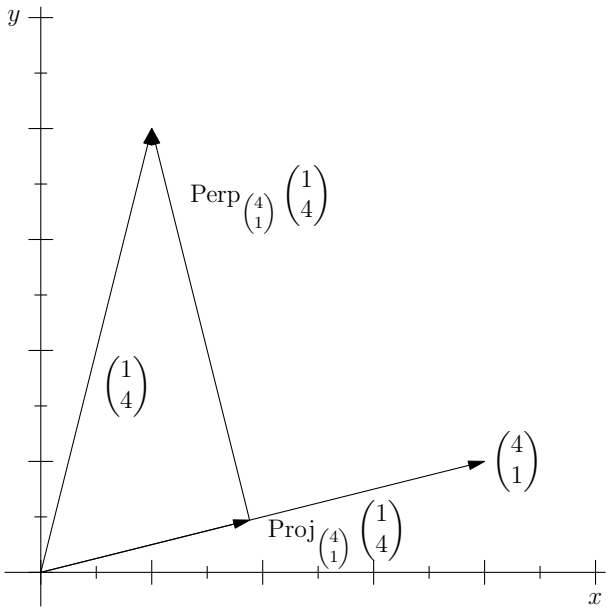

Figure 4.2: Projection and Perpendicular Vectors

Using vectors to define local directions is a particularly useful tool. A standard example is camera location in a three dimensional virtual environment. First, you need to know the location of the camera, which is an ordinary vector starting from the origin. Second, you need to know what direction the camera is pointing, which is a local direction vector which treats the camera location as the current origin.

One of the most difficult things about learning vector geometry is becoming accustomed to local direction vectors. We don't always carefully distinguish between vectors at the origin and local direction vectors; often, the difference is implied and it is up to the reader/student to figure out how the vectors are being used.

4.4 Projections

Definition 4.4.1. Let u and v be two vectors in $\mathbb{R}^n$. The *projection* of u onto v is a scalar multiple of the vector v given by the following formula.

$$\text{Proj}_v u = \left(\frac{u \cdot v}{|v|^2} \right) v$$

Note that the bracketed term involves the dot product and the norm, so it is a scalar. Therefore, the result is a multiple of the vector v. Projection is best visualized as the shadow of the vector u on the the vector v.

Definition 4.4.2. Let u and v be two vectors in $\mathbb{R}^n$. The part of u which is *perpendicular* to v is given by the following formula.

$$\text{Perp}_v u = u - \text{Proj}_v u$$

We can rearrange the previous definition to solve for the original vector u.

$$u = \text{Proj}_v u + \text{Perp}_v u$$

For any vector v, every vector u can be decomposed into a sum of two unique vectors: one in the direction of v and one perpendicular to v. If u and v are already perpendicular, then the projection term is the zero vector. If u is a multple of v, then the perpendicular term is the zero vector. We think of this decomposition as capturing two pieces of the vector u: the part that aligns with the direction of v and the part that has nothing to do with the direction of v. A vectors with its projection and perpendicular onto another vector is shows in Figure 4.2.

Lecture 5

Polyhedra in $\mathbb{R}^n$

Now that we have defined vectors, we want to investigate more complicated objects in $\mathbb{R}^n$. The major objects for the course work (linear and affine subspaces) will be defined in the coming lectures. In this lecture, however, we take a short detour to discover how familiar shapes and solids extend into higher dimensions. I'll use the standard term *polyhedron* (plural polyhedra) to refer to a straight-edged objects in any $\mathbb{R}^n$. First, however, we start with a familiar without straight edges.

5.1 Spheres

Spheres are, in some way, the easiest objects to generalize. Spheres are all things in $\mathbb{R}^n$ which are exactly one unit of distance from the origin. The 'sphere' in $\mathbb{R}$ is just the points -1 and 1. The 'sphere' in $\mathbb{R}^2$ is the circle $x^2+y^2=1$. The sphere in $\mathbb{R}^3$ is the conventional sphere, with equation $x^2+y^2+z^2=1$. The sphere in $\mathbb{R}^4$ has equation $x^2+y^2+z^2+w^2=1$. The sphere in $\mathbb{R}^n$ has equation $x_1^2+x_2^2+\ldots+x_n^2=1$. For dimensional reasons, since spheres are usually considered hollow objects, the sphere in $\mathbb{R}^n$ is called the $(n-1)$-sphere. That means the circle is the 1-sphere and the conventional sphere is the 2-sphere.

A sphere relate to the lower dimension spheres by looking at slices. Any slice of a sphere produces a circle. Likewise, any slice of a 3-sphere produces a 2-sphere.

5.2 Simplicies

The simplex is the simplest straight-line polyhedra. I'll define them without specifying their vector definition in $\mathbb{R}^n$, since that definition is slightly more technical than necessary.

- The 1-simplex is a line segment.
- The 2-simplex is an equilateral triangle. It is formed from the 1-simplex by adding one new vertex and drawing edges to the existing vertex such that all edges have the same length.
- The 3-simplex is a tetrahedron (or triangular pyramid). Again, it is formed by adding one new vertex in a new direction and drawing lines to all existing vertices such that all line (new and old) have the same length.
- This process extends into higher dimensions. In each stage a new vertex is added in the new direction and edges connect it to all previous vertices, such that all edges have the same length.

5.3 Cross-Polytopes

Another family of regular solids which extends to all dimensions is the cross-polytopes.

- In $\mathbb{R}^2$, the cross-polytope is the diamond with vertices $\binom{1}{0}$, $\binom{-1}{0}$, $\binom{0}{1}$, and $\binom{0}{-1}$. In each dimension, two vertices are added at ± 1 in the new axis direction, and edges are added connecting the two new vertices to each existing vertex (but the two new vertices are not connected to each other)..
- In $\mathbb{R}^3$, the cross-polytope is the octahedron, with vertices $\begin{pmatrix}1\\0\\0\end{pmatrix}$, $\begin{pmatrix}-1\\0\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\\0\end{pmatrix}$, $\begin{pmatrix}0\\-1\\0\end{pmatrix}$, $\begin{pmatrix}0\\0\\1\end{pmatrix}$, and $\begin{pmatrix}0\\0\\-1\end{pmatrix}$.
- In higher dimensions, the vertices are all vectors with ± 1 in one component and zero in all other components. Each vertex is connected to all other vertices except its opposite.

5.4 Cubes

Finally, we have the family of cubes.

- The 'cube' n $\mathbb{R}$ is the interval $[0, 1]$. It can be defined by the inequality $0 \leq x \leq 1$ for $x \in \mathbb{R}$.
- In $\mathbb{R}^2$, the 'cube' is the just the ordinary (solid) square. It is all vectors $\binom{x}{y}$ such that $0 \leq x \leq 1$ and $0 \leq y \leq 1$. The square can be formed by taking two intervals and connecting the matching vertices. It has four vertices and four edges.

- In $\mathbb{R}^3$, the square object is the (solid) cube. It is all vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, such that each coordinate is in the interval $[0, 1]$. It can also be seen as two squares with matching vertices connected. Two square gives eight vertices. Eight square edges plus four connecting edges gives the twelve edges of the cube. It also has six square faces.
- Then we can simply keep extending. In $\mathbb{R}^4$, the set of all vectors $\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$ where all coordinates are in the interal $[0, 1]$ is called the hypercube or 4-cube. It can be seen as two cubes with pairwise edges connected. The cubes each have eight vertices, so the 4-cubes has sixteen vertices. Each cube has 12 edges, and their are 8 new connecting edges, so the 4-cube has 32 edges. It has 24 faces and 8 cells. A cell (or 3-face) here is a three dimensional 'edge' of a four (or higher) dimensional object.
- There is an n-cube in each $\mathbb{R}^n$, consisting of all vectors where all components are in the interval $[0, 1]$. Each can be constructed by joining two copies of a lower dimensional $(n-1)$-cube with edges between matching vertices.

5.5 Other Platonic Solics

A regular polyhedron is one where all edges, faces, cells, n-cells are the same size/shape and, in addition, the angles between all the edges, faces, cells, n-cells are also the same whever the various objects meet. In addition, the polyhedron is called convex if all angles are greater that $\frac{\pi}{2}$ radians. The study of convex regular polyhedra is an old and celebrated part of mathematics.

In $\mathbb{R}^2$, there are infinitely many convex regular polyhedra: the regular polygons with any number of sides. In $\mathbb{R}^3$, in addition to the cube, tetrahedron and octahedron, there are only two others: the dodecahedron and the icosahedron. These were well known to the ancient Greeks and are called the *Platonic Solids*.

The three families (cube, tetrahedron, cross-polytope) extend to all dimensions, but the dodecahedron and icoahedron are particular to $\mathbb{R}^3$. It is a curious and fascinating question to ask what other special, unique convex regular polyhedra occur in higher dimesions.

In $\mathbb{R}^4$, there are three others. They are called the 24-cell, the 120-cell and 600-cell, named for the number of 3-dimensional cells they contain (the same way the 4-cubes contains 8 3-cubes). The 24-cell is built from 24 octahedral cells. The 120-cell is built from dodecahedral cells and the 600-cell is built from tetrahedral cells. The 120-cell, in some ways, extends the dodecahedron and the 600-cell extends the icosahedron. The 24-cell is unique to $\mathbb{R}^4$.

It is an amazing theorem of modern mathematics that in dimensions higher than 4, there are no regular polyhedra other than the three families. Neither the icosahedron nor the dodecahedron extend, and there are no other erratic special polyhedra found in any higher dimension.

Lecture 6

Linear and Affine Subspaces

6.1 Definitions

In addition to vectors, we want to consider various geometric objects that live in $\mathbb{R}^n$. Since this is linear algebra, we will be restricting ourselves to flat objects.

Definition 6.1.1. A *linear subspace* of $\mathbb{R}^n$ is a non-empty set of vectors L which satisfies the following two properties.

- If $u, v \in L$ then $u + v \in L$.
- If $u \in L$ and $a \in \mathbb{R}$ then $av \in L$.

There are two basic operations on $\mathbb{R}^n$: we can add vectors and we can multiply by scalars. Linear subspaces are just subsets where we can still perform both operations and remain in the subset.

Geometrically, vector addition and scalar multiplication produce flat objects: lines, planes, and their higher-dimensions analogues. Also, since we can take $a = 0$, we must have $0 \in L$. So linear subspaces can be informally defined as flat subsets which include the origin.

Definition 6.1.2. An *affine subspace* of $\mathbb{R}^n$ is a non-empty set of vectors A which can be described as a sum $v + u$ where v is a fixed vector and u is any vector in some fixed linear subspace L. With some abuse of notation, we write this as a sum of a fixed vector and a linear subspace.

$$A = v + L$$

We think of affine subspaces as flat spaces that may be offset from the origin. The vector v is called the *offset vector*. Affine spaces include linear spaces, since we can also take u to be the zero vector and have $A = L$. Affine objects are the lines, planes and higher dimensional flat objects that may or may not pass through the origin.

Notice that we defined both affine and linear subspaces to be non-empty. The empty set $\emptyset$ is *not* a linear or affine subspace. The smallest linear subspace if $\{0\}$: just the origin. The smallest affine subspace is any isolated point.

We need ways to algebraically describe linear and affine substapces. There are two main approaches: loci and spans.

6.2 Loci

Definition 6.2.1. Consider any set of linear equations in the variables $x_1, x_2, \ldots, x_n$. The *locus* in $\mathbb{R}^n$ of this set of equations is the set of vectors which satisfy *all* of the equations. The plural of locus is *loci*.

In general, the equations can be of any sort. The unit circle in $\mathbb{R}^2$ is most commonly defined as the locus of the equation $x^2 + y^2 = 1$. The graph of a function is the locus of the equation $y = f(x)$. However, in linear algebra, we exclude curved objects. We're concerned with linear/affine objects: things which are straight and flat.

Definition 6.2.2. Let a_i and c be real numbers. A *linear equation* in variables $x_1, x_2, \ldots x_n$ is an equation of the following form.

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = c$$

Proposition 6.2.3. *Any linear or affine subspace of $\mathbb{R}^n$ can be described as the locus of finitely many linear equations. Likewise, the locus of any number of linear equations is either an affine subspace of $\mathbb{R}^n$ or the empty set.*

The best way to think about loci is in terms of restrictions. We start with all of $\mathbb{R}^n$ as the locus of no equations, or of the equation $0 = 0$. There are no restrictions. Then we introduce equations. Each equation is a restriction on the available points. If we work in $\mathbb{R}^2$, adding the equation $x = 3$ restricts us to a vertical line passing through the x-axis at $\binom{3}{0}$. Likewise, if we were to use the equation $y = 4$, we would have a horizontal line passing through the y-axis at $\binom{0}{4}$. If we consider the locus of *both* equations, we have only one point remaining: $\binom{3}{4}$ is the only point that satisfies both equations. In this way, each additional equation potentially adds an additional restrictions and leads to a smaller linear or affine subspaces. The next three definitions give the familiar names for loci of one restriction.

Definition 6.2.4. A *line* in $\mathbb{R}^2$ is the locus of the equation $ax + by = c$ for $a, b, c \in \mathbb{R}$. In general, the line is affine. The line is linear if $c = 0$.

Definition 6.2.5. A *plane* in $\mathbb{R}^3$ is the locus of the linear equation $ax + by + cz = d$. In general, the plane is affine. The plane is linear if $d = 0$.

If we think of a plane in $\mathbb{R}^3$ as the locus of one linear equation, the important dimensional fact about a plane is not that it has dimension two but that it has dimension one less than its ambient space $\mathbb{R}^3$.

Definition 6.2.6. A *hyperplane* in $\mathbb{R}^n$ is the locus of one linear equation: $a_1x_1 + a_2x_2 + \ldots + a_nx_n = c$. It has dimension $n-1$. It is, in general, affine. The hyperplane is linear if $c = 0$.

All this discussion is pointing towards the notion of dimension. The dimension of $\mathbb{R}^n$ is n; it is the number of independent directions or degrees of freedom of movement. For linear or affice subspaces, we also want a well-defined notion of dimension. So far it looks good: the restriction of a linear equation should drop the dimension by one. A line in $\mathbb{R}^3$, which is one dimensional in a three dimensional space, should be the locus of two different linear equations.

We would like the dimension of a locus to be simply determined by the dimension of the ambient space minus the number of equations. However, there is a problem with the naïve approach to dimension. In $\mathbb{R}^2$, adding two linear equations should drop the dimension by two, giving a dimension zero subspace: a point. However, consider the equations $3x + 4y = 0$ and $6x + 8y = 0$. We have two equations, but the second is redundant. All points on the line $3x + 4y = 0$ are already satisfied by the second equation. So, the locus of the two equations only drops the dimension by one.

In $\mathbb{R}^3$ the equations $y = 0$, $z = 0$ and $y + z = 0$ have a locus which is the x axis. This is one dimensional in a three dimensional space, so the dimension from the three equations has only dropped by two. One of the equations is redundant.

This problem scales into higher dimensions. In $\mathbb{R}^n$, if we have several equations, it is almost impossible to see, at a glance, whether any of the equations are redundant. We need methods to calculate dimension. Unfortunately, those methods will have to wait until later in these notes.

6.3 Intersection

Definition 6.3.1. If A and B are sets, their intersection $A \cap B$ is the set of all points they have in common. The intersection of affine subspaces is also an affine subspace. If A and B are both linear, the intersection is also linear.

Example 6.3.2. Loci can easily be understood as intersections. Consider the locus of two equations, say the example we have from $\mathbb{R}^2$ before: the locus of $x = 3$ and $y = 4$. We defined this directly as a single locus. However, we could just as easily think of this as the intersection of the two lines given by $x = 3$ and $y = 4$ seperately. In this way, it is the intersection of two loci. Similarly, all loci are the intersection of the planes or hyperplanes defined by each individual linear equation.

Lecture 7

Spans

7.1 Definitions

In addition to presenting the second description of linear subspaces, this chapter also introduces linear combinations, span and linear (in)dependence. These are some of the most important and central definition in linear algebra.

Definition 7.1.1. A *linear combination* of a set of vectors $\{v_1, v_2, \ldots, v_k\}$ is a sum of the form $a_1v_1 + a_2v_2 + \ldots a_kv_k$ where the $a_i \in \mathbb{R}$.

Definition 7.1.2. The *span* of a set of vectors $\{v_1, v_2, \ldots, v_k\}$, written Span$\{v_1, v_2, \ldots, v_k\}$, is the set of *all* linear combinations of the vectors.

After loci, spans are the second way of defining linear subspaces. Span are never affine: since linear combinations allow for all the coefficients a_i to be zero, spans always include the origin. To use spans to define affine subspaces, we have to add an offset vector.

Definition 7.1.3. An *offset span* is an affine subspace formed by adding a fixed vector u, called the *offset vector*, to the span of some set of vectors.

Loci are built top-down, by starting with the ambient space and reducing the number of points by imposing restrictions in the form of linear equations. Their dimension, at least ideally, is the dimension of the ambient space minus the number of equations or restrictions. Spans, on the other hand, are built bottom-up. They start with a number of vectors and take all linear combinations: more starting vectors leads to more independent directions and a larger dimension.

In particular, the span of one non-zero vector is the line (through the origin) consisting of all multiples of that vector. Similarly, we expect the span of two vectors to be a plane. However, here we have the

same problem as we had with loci: we may have redundant information. For example, in $\mathbb{R}^2$, we could consider the span $\text{Span}\left\{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}\right\}$. We would hope the span of two vectors would be the entire plane, but this is just the line in the direction $\begin{pmatrix}1\\2\end{pmatrix}$. The vector $\begin{pmatrix}2\\4\end{pmatrix}$, since it is already a multiple of $\begin{pmatrix}1\\2\end{pmatrix}$, is redundant.

The problem is magnified in higher dimensions. If we have the span of a large number of vectors in $\mathbb{R}^n$, it is nearly impossible to tell, at a glance, whether any of the vectors are redundant. We would like to have tools to determine this redundancy. As with the tools for dimensions of loci, we have to wait until a later section of these notes.

7.2 Dimension and Basis

Definition 7.2.1. A set of vectors $\{v_1, v_2, \ldots, v_k\}$ in $\mathbb{R}^n$ is called *linearly independent* if the equation

$$a_1v_1 + a_2v_2 + a_3v_3 + \ldots + a_kv_k = 0$$

has only the trivial solution: for all i, $a_i = 0$. If a set of vectors isn't linearly independent, it is called *linearly dependant.*

This may seem like a strange definition, but it algebraically captures the idea of independent directions. A set of vectors is linearly independent if all of them point in fundamentally different directions. We could also say that a set of vectors is linearly independent if no vector is in the span of the other vectors. No vector is a redundant piece of information; if we remove any vectors, the span of the set gets smaller.

In order for a set like this to be linearly independent, we need $k \leq n$. $\mathbb{R}^n$ has only n independent directions, so it is impossible to have more than n linearly independent vectors in $\mathbb{R}^n$.

Definition 7.2.2. Let L be a linear subspace of $\mathbb{R}^n$. Then L has *dimension* k if L can be written as the span of k linearly independent vectors.

Definition 7.2.3. Let A be an affine subspace of $\mathbb{R}^n$ and write A as $A = u + L$ for L a linear subspace and u a offset vector. Then A has *dimension* k if L has dimension k.

This is the proper, complete definition of dimension for linear and affine spaces. It solves the problem of redundant information (either redundant equations for loci or redundant vectors for spans) by insisting on a linearly independent spanning set.

Definition 7.2.4. Let L be a linear subspace of $\mathbb{R}^n$. A *basis* for L is a minimal spanning set; that is, a set of linearly independent vectors which span L.

Since a span is the set of all linear combinations, we can think of a basis as as way of writing the vectors of L: every vector in L can be written as a linear combination of the basis vectors. A basis gives a nice way to account for all the vectors in L.

Linear subspaces have many (infinitely many) different bases. There are some standard choices.

Definition 7.2.5. The *standard basis* of $\mathbb{R}^2$ is composed of the two unit vectors in the positive x and y directions. We can write any vector as a linear combination of the basis vectors.

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} x \\ y \end{pmatrix} = xe_1 + ye_2$$

The *standard basis* of $\mathbb{R}^3$ is composed of the three unit vectors in the positive x, y and z directions. We can again write any vector as a linear combination of the basis vectors.

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = xe_1 + ye_2 + ze_3$$

The *standard basis* of $\mathbb{R}^n$ is composed of vectors $e_1, e_2, \ldots, e_n$ where e_i has a 1 in the ith component and zeroes in all other components. e_i is the unit vector in the positive ith axis direction.

Lecture 8

Normals to Planes and Hyperplanes

8.1 Dot Products and Loci

Having discussed spans, let's return to loci. By re-examing the linear equations, we can define loci via dot products. Consider, again, the general linear equation in $\mathbb{R}^n$.

$$a_1u_1 + a_2u_2 + \ldots + a_nu_n = c$$

Let's think of the variables u_i as the components of a vector $u \in \mathbb{R}^n$. We also have n scalars a_i which we likewise treat as a components of the vector $a \in \mathbb{R}^n$. The vector u is variables, the vector a is constant. Then we can re-write the linear equation using these two vectors.

$$a_1u_1 + a_2u_2 + \ldots + a_nu_n = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = a \cdot u = c$$

In this way, a linear equation specifies that the dot product result of a variables vector u with a fixed vector a must have the result c. In this light, an affine plane in $\mathbb{R}^3$ is given by the equation $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = c$. This plane is precisely all vectors whose dot product with the vector $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ is the fixed number c. If $c = 0$, then we get $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0$. A linear plane is the set of all vectors which are perpendicular to a fixed vector $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$.

Definition 8.1.1. Let P be a plane in $\mathbb{R}^3$ determined by the equation $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = c$. The vector $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ is called the *normal to the plane*. Let H be a hyperplane in $\mathbb{R}^n$ determined by the equivalent equation in $\mathbb{R}^n$.

$$u \cdot a = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = c$$

The vector a is called the *normal to the hyperplane*.

If $c = 0$, the plane or hyperplane is perpendicular to its normal. This notion of orthogonality still works when $c \neq 0$. In this case, the normal is a *local* perpendicular direction from any point on the affine plane. Treating any such point as a local origin, the normal points in a direction perpendicular to all the *local direction* vectors which lie on the plane.

8.2 An Algorithm for Equations of Planes

Now we can build a general process for finding the equation of a plane in $\mathbb{R}^3$. Any time we have a point p on the plane and two *local direction vectors* u and v which remain on the plane, we can find a normal to the plane by taking $u \times v$. Then we can find the equation of the plane by taking the dot product $p \cdot (u \times v)$ to find the constant c.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot (u \times v) = c$$

If we are given three points on a plane (p, q and r), then we can use p as the local origin and construct the local direction vectors as $q - p$ and $r - p$. The normal is $(q - p) \times (r - p)$. In this way, we can construct the equation of a plane given three points or a single point and two local directions.

Lecture 9

Systems of Linear Equations

9.1 Definitions

The previous eight chapters introduced the geometric side of this course. This chapter changes perspective to look at the algebraic side of linear algebra. A major problem in algebra is solving systems of equations: given several equations in several variables, can we find values for each variable which satisfy all the equations? In general, algegra considers any kind of equation.

Example 9.1.1.

$$\begin{aligned} x^2 + y^2 + z^2 &= 3 \\ xy + 2xz - 3yz &= 0 \\ x + y + z &= 3 \end{aligned}$$

This system is solved by $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. When substitued in the equations, these values satisfy all three. No other triples satisfies, so this is a unique solution.

Solving arbitrary systems of equations is a very difficult problem. There are two initial techniques, which are usually taught in high-school mathematics: isolating and replacing variables; or performing arithmetic with the equations themselves. Both are useful techniques. However, when the equations become quite complicated, both methods can fail. In that case, solutions to systerms can only be approached with approximation methods. The methodology of these approximations is a whole branch of mathematics in itself.

We are going to restrict ourselves to a class of systems which is more approchable: systems of linear equations. Recall the definition of a linear equation.

Definition 9.1.2. Let a_i, for $i \in \{1, \ldots n\}$ and c be real numbers. A *linear equation* in the variables $x_1, x_2 \ldots, x_n$ is a equation of the following form.

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = c$$

Given a fixed set of variables $x_1, x_2, \ldots, x_n$, we want to consider systems of linear equations (and only linear equations) in those variables. While much simpler than the general case, linear systems are a common and useful type of system to consider. Such systems also work well with arithmetic solution techniques, since we can add two linear equations together and still have a linear equation.

This turns out to be the key idea: we can do arithmetic with linear equations. Specifically, there are three things we can do with a system of linear equation. These three techniques *preserve the solutions*, that is, they don't alter the values of the variables that solve the system.

- Multiple a single equation by a (non-zero) constant. (Multiplying both sides of the equation, of course).
- Change the order of the equations.
- Add one equation to another.

If we combine the first and third, we could restate the third as: add a multiple of one equation to another. In practice, we often think of the third operation this way.

9.2 Matrices

If we worked with the equations directly, we would find, with careful application of the three techniques, we could always solve linear systems. However, the notation becomes cumbersome for large systems. Therefore, we introduce a method to encode the information in a nicely organized way. This method takes the coefficients of the linear equations in a system and puts them in a rectangular box called a matrix.

Definition 9.2.1. A *matrix* is a rectangular array of scalars. If the matrix has m rows and n columns, we say it is an $m \times n$ matrix. The scalars in the matrix are called the *entries*, *components* or *coefficients*.

Definition 9.2.2. The rows of a matrix are called the *row vectors* of the matrix. Likewise the columns are called *column vectors* of the matrix.

A matrix is a rectangular array of numbers, enclosed either in square or round brackets. Here are two ways of writing a particular 3×2 matrix with integer coefficients.

$$\begin{pmatrix} 5 & 6 \\ -4 & -4 \\ 0 & -3 \end{pmatrix} \qquad \begin{bmatrix} 5 & 6 \\ -4 & -4 \\ 0 & -3 \end{bmatrix}$$

The rows of this matrix are the following three vectors.

$$\begin{pmatrix} 5 \\ 6 \end{pmatrix} \qquad \begin{pmatrix} -4 \\ -4 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

The columns of this matrix are the following two vectors.

$$\begin{pmatrix} 5 \\ -4 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 6 \\ -4 \\ -3 \end{pmatrix}$$

In these notes, we'll use curved brackets for matrices; however, in many texts and books, square brackets are very common. Both notations are conventional and acceptable.

If we want to write a general matrix, we can use a double subscript.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

By convention, when we write an aribitrary matrix entry a_{ij} the first subscript tells us the row and the second subscript tells us the column. For example, a_{64} is the entry in the sixth row and the fourth column. In the rare occurence that we have matrices with more than 10 rows or columns, we can seperate the indices by commas: $a_{12,15}$ would be in the twelfth row and the fifteenth column. Sometime we write $A = a_{ij}$ as short-hand for the entire matrix when the size is understood or undetermined.

Definition 9.2.3. A *square matrix* is a matrix with the same number of rows as columns. Here are two examples.

$$\begin{pmatrix} 4 & -2 & 8 \\ -3 & -3 & -3 \\ 0 & 0 & -1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Definition 9.2.4. The *zero matrix* is the unique matrix (one for every size $m \times n$) where all the coefficients are all zero.

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

Definition 9.2.5. The *diagonal entries* of a matrix are all entries a_{ii} where the row and column indices are the same. A *diagonal matrix* is a matrix where all the non-diagonal entries are zero.

$$\begin{pmatrix} 5 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Definition 9.2.6. The *identity matrix* is the unique $n \times n$ matrix (one for each n) where the diagonal entries are all 1 and all other entries are 0. It is often written as I or Id.

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

Sometimes it is useful add separation to the organization of the matrix.

Definition 9.2.7. An *extended matrix* is a matrix with a vertical division seperating the columns into two groups.

$$\left(\begin{array}{ccc|c} -3 & 0 & 6 & 1 \\ 4 & -2 & 2 & -1 \\ 0 & 0 & 3 & -7 \end{array}\right)$$

Definition 9.2.8. The set of all $n \times m$ matrices with real coefficients is written $M_{n,m}(\mathbb{R})$. For square matrices ($n \times n$), we simply write $M_n(\mathbb{R})$. If we wanted to change the set of scalars to some other number set S, we would write $M_{n,m}(S)$ or $M_n(S)$.

9.3 Matrix Represetation of Systems of Linear Equations

Consider a general system with m equations in the variables $x_1, x_2, \ldots, x_n$, where a_{ij} and c_i are real numbers.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n &= c_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n &= c_m \end{aligned}$$

We encode this as an extended matrix by taking the a_{ij} and the c_i as the matrix coefficients. We drop the x_i, keeping track of them implicitly by their positions in the matrix. The result is a $m \times (n+1)$ extended matrix where the vertical line notes the position of the equals sign in the original equation.

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & c_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & c_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{11} & a_{12} & \ldots & a_{1n} & c_1 \end{array}\right)$$

Example 9.3.1.

$$\begin{aligned} -x + 3y + 6z &= 1 \\ 2x + 2y + 2z &= -6 \\ 5x - 5y + z &= 0 \end{aligned}$$

We transfer the coefficients into the matrix representation.

$$\left(\begin{array}{ccc|c} -1 & 3 & 6 & 1 \\ 2 & 2 & 2 & -6 \\ 5 & -5 & 1 & 0 \end{array}\right)$$

Example 9.3.2.

$$\begin{aligned} -3x - 10y + 15z &= -34 \\ 20x - y + 19z &= 25 \\ 32x + 51y - 31z &= 16 \end{aligned}$$

We transfer the coefficients into the matrix representation.

$$\left(\begin{array}{ccc|c} -3 & -10 & 15 & -34 \\ 20 & -1 & 19 & 25 \\ 32 & 51 & -31 & 16 \end{array}\right)$$

Example 9.3.3. Sometimes not every equations explicitly mentions each variable.

$$\begin{aligned} x - 2z &= 0 \\ 2y + 3z &= -1 \\ -3x - 4y &= 9 \end{aligned}$$

This system is clarified by adding the extra variables with coefficient 0.

$$\begin{aligned} x + 0y - 2z &= 0 \\ 0x + 2y + 3z &= -1 \\ -3x - 4y + 0z &= 9 \end{aligned}$$

Then it can be clearly encoded as a matrix.

$$\left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 2 & 3 & -1 \\ -3 & -4 & 0 & 9 \end{array}\right)$$

In this way, we can change any system of linear equations into a extended matrix (with one column after the vertical line), and any such extended matrix into a system of equations. The columns represent the hidden variables. In the examples above, we say that the first column is the x column, the second is the y column, the third if the z column, and the column after the vertical line is the column of constants.

Lecture 10

Solving by Row Reduction

10.1 Row Operations and Gaussian Elimination

We defined three operations on systems of equations. Now that we have encoded systems as matrices, we need to understand the equivalent operations for matrices. Each equation in the system gives a row in the matrix; therefore, equation operations become row operations.

- Multiply an equation by a non-zero constant $\implies$ multiply a row by a non-zero constant.
- Change the order of the equations $\implies$ exchange two rows of a matrix.
- Add (a multiple of) an equation to another equation $\implies$ add (a multiple of) one row to another row.

Since the original operations didn't change the solution of a system of equations, the row operations on matrices also preserve the solution of the associated system.

In addition, we need to know what a solution looks like in matrix form.

Example 10.1.1. This example shows the encoding of a direction solution.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 8 \end{array}\right)$$

The system has three columns before its vertical line, therefore it corresponds to a system of equation in x, y, z. If we translate this back into equations, we get the following system.

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

Removing the zero terms gives $x = -3$, $y = -2$ and $z = 8$. This equation delivers its solution directly: no work is necessary. We just read the solution off the page.

We have a name for the special form of a matrix where we can directly read the solutions of the associated linear system.

Definition 10.1.2. A matrix is a *reduced row-echelon* matrix, or is said to be in *reduced row-echelon form*, if the following things are true:

- The first non-zero entry in each row is one. (A row entirely of zeros is fine). These entries are called leading ones.
- Each leading one is in a column where all the other entries are zero.

We shall see that as long as an extended matrix is in reduced row-echelon form, we can always directly read off the solution to the system.

Knowing how to recognize solutions in matrix form and knowing the row operations, we can now state our matrix-based approach to solving linear systems. First, we to translate a system into a matrix. We then use row operations (which don't change the solution at all) to turn the matrix into reduced row-echelon form, where we can just read off the solution. Row operations will always be able to accomplish this.

Definition 10.1.3. The process of using row operations to change a matrix into reduced row echelon form is called *Guassian elimantion* or *row reduction*. This process proceeds in the following steps.

- Take a row which has a non-zero entry in the first column. (Often we exchange this row with the first, so that we are working on the first row).
- Multiply by a constant to get a leading one in this row.
- Now that we have a leading one, we want to clear all the other entries in the first column. Thererfore, we add multiples of the row with the leading one to the other rows, one by one, to make those first column entries zero.
- That produces a column with a leading one in the first colum and sets all other first column entries zero.
- Then we proceed to the next column and repeat the process all over again. Any column which is all zeros is skipped. Any row which is all zeros is also left alone.

10.2 Solution Spaces and Free Parameters

Definition 10.2.1. If we have a system of equations in the variables $x_1, x_2, \ldots, x_n$, the set of values for these variables which satisfy all the equations is called the *solutions space* of the system. Since each set of values is a vector, we think of the solution space as a subset of $\mathbb{R}^n$. For linear equations, the solution space will always be a affine subspace. If we have encoded the system of equations in an extended matrix, we can refer to the solution space of the matrix instead of the system.

Definition 10.2.2. Since solution spaces are affine, they can be written as $u + L$ where u is a fixed offset vector and L is a linear space. Since L is linear, it has a basis $v_1, v_2, \ldots v_k$ and any vector in L is a linear combination of the v_i. We can write any and all solutions to the system in the following fashion (where $a_i \in \mathbb{R}$).

$$u + a_1v_1 + a_2v_2 + a_3v_3 + \ldots + a_kv_k$$

In such a description of the solution space, we call the a_i *free parameters*. One of our goals in solving systems is to determine the number of free parameters for each solution space.

Example 10.2.3. Let's return to the example at the start of this section to understand how we read solutions from reduced row-echelon matrices.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 8 \end{array}\right)$$

This example is the easiest case: we have a leading one in each row and column left of the horizontal line, and we just read off the solution, $x = -3$, $y = -2$ and $z = 8$. Several other things can happen, though, which we will also explain by examples.

Example 10.2.4. We might have a row of zeros in the reduced row-echelon form.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This row of zeros doesn't actually change the solution at all: it is still $x = -3$, $y = -2$ and $z = 8$. The last line just corresponds to an equation which says $0x + 0y + 0z = 0$ or just $0 = 0$ which is always true. Rows of zeros arise from redundant equations in the original system.

Example 10.2.5. We might have a column of zeros. Consider something similar to the previous examples, but now in four variables w, x, y, z. (In that order, so that the first column is the w column, the second is the x column, and so on).

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 8 \end{array}\right)$$

The translation gives $w = -3$, $x = -2$ and $y = 8$ but the matrix says nothing about z. A column of zeros corresponds to a free variable: any value of z solves the system as long as the other variables are set.

Example 10.2.6. The columns of a free variables need not contain all zeros. The following matrix is still in reduced row-echelon form.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 & -3 \\ 0 & 1 & 0 & -1 & -2 \\ 0 & 0 & 1 & -1 & 8 \end{array}\right)$$

Reduced row-echelon form needs a leading one in each non-zero row, and all other entries in a column with a lead one must be zero. This matrix satisfies all the conditions. The fourth column corresponds to the z variable, but lacks a leading one. Any column lacking a leading one corresponds to a free variable. Look at the translation of this matrix back into a linear system.

$$\begin{aligned} w + 2z &= -3 \\ x - z &= -2 \\ y - z &= 8 \end{aligned}$$

Let's move the z terms to the right side.

$$\begin{aligned} w &= -2z - 3 \\ x &= z - 2 \\ y &= z + 8 \\ z &= z \end{aligned}$$

We've added the last equation, which is trivial to satisfy, to show how all the terms depend on z. This makes it clear that z is a free variable, since we can write the solution as follows:

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -3 \\ -2 \\ 8 \\ 0 \end{pmatrix} + z \begin{pmatrix} -2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Any choice of z will solve this system. If we want specific solutions, we take specific values of z. For example, if $z = 1$, these equations give $w = -5, x = -1$ and $y = 9$. Moreover, this solutions is explicity given as an offset spane, where the columns of constants is the offset and the free variable gives the span.

Example 10.2.7. We can go further. The following matrix is also in reduced row-echelon form (again, using w, x, y and z as variables in that order).

$$\left(\begin{array}{cccc|c} 1 & 0 & 3 & 2 & -3 \\ 0 & 1 & -1 & -1 & -2 \end{array}\right)$$

Here, both the third and the fourth columns have no leading one, therefore, both y and z are free variables. We translate the matrix back into a linear system.

$$\begin{aligned} w &= -3y - 2z - 3 \\ x &= y + z - 2 \\ y &= y \\ z &= z \end{aligned}$$

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

Any choice of y and z gives a solution. For example, if $y = 0$ and $z = 1$ then $w = -5$ and $x = -1$ completes a solution. Since we have all these choices, any system with a free variable has infinitely many solutions.

Example 10.2.8. Finally, we can also have a reduced row-echelon matrix of the following form.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

The first three rows translate well. However, the fourth row translates to $0x + 0y + 0z = 1$ or $0 = 1$. Obviously, this can never be satisfied. Any row which translates to $0 = 1$ leads to a contradiction: this means that the system has no solutions, since we can never satisfy an equation like $0 = 1$.

We summarize the three possible situations for reduced row-echelon matrices.

- If there is a row which translates to $0 = 1$, then there is a contradiction and the system has no solutions. This overrides any other information: whether or not there are free variables elsewhere, there are still no solutions.
- If there are no $0 = 1$ rows and all columns left of the vertical line have leading ones, then there is a unique solution. There are no columns representing free variables, and each variable has a specific value which we can read directly from the matrix.
- If there are no $0 = 1$ rows and there is at least one column left of the vertical line without a leading one, then each such column represents a free variable. The remaining variables can be expressed in terms of the free variables and any choices of the free variables leads to a solution. There are infinitely many solutions. The solutions space can always be expressed as a offset span.

In particular, note that there are only three basic cases for the number of solutions: none, one, or infinitely many. No linear system has exactly two or exactly three solutions. In the case of infinitely many solutions, the interesting question is the dimension of the solution space. This question, however, is easily answered.

Proposition 10.2.9. *The dimension of a solution space of a linear system of equations is the number of free variables.*

Lecture 11

Dimensions of Spans and Loci

11.1 Rank

Definition 11.1.1. Let A be a $m \times n$ matrix. The *rank* of A is the number of leading ones in its reduced row-echelon form.

11.2 Dimensions of Spans

We can now define the desired techniques to solve the dimension problems that we presented earlier. For loci and spans, we didn't have a way of determining what information was redundant. Row-reduction of matrices gives us the tool we need.

Given some vectors $\{v_1, v_2, \ldots, v_k\}$, what is the dimension of Span$\{v_1, v_2, \ldots, v_k\}$? By definition, it is the number of linearly independent vectors in the set. Now we can test for linearly independent vectors.

Proposition 11.2.1. *Let $\{v_1, v_2, \ldots, v_k\}$ be a set of vectors in $\mathbb{R}^n$. If we make these vectors the rows of a matrix A, then the rank of the matrix A is the number of linearly independent vectors in the set. Moreover, if we do the row reduction without exchanging rows (which is always possible), then the vectors which correspond to rows with leading ones in the reduced row-echelon form form a maximal linearly independent set, i.e., a basis. Any vector corresponding to a row without a leading one is a redundant vector in the span. The set is linearly independent if and only if the rank of A is k.*

11.3 Bases for Spans

If Span$\{v_1, v_2, v_3\}$ has dimension two and the first two vectors form a basis, then v_3 is a redundant part of the span. This means that we can write v_3 as a linear combination of the first two vectors. That is, there are two constants, a and b, such that $v_3 = av_1 + bv_2$. We would like to be able to determine these constants. Matrices and row-reduction is again the tool we turn turn.

Example 11.3.1. Let's work by example and consider Span$\left\{ \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 3 \end{pmatrix}, \begin{pmatrix} -4 \\ -4 \\ 7 \end{pmatrix} \right\}$. If we make these rows of a matrix and row-reduce, we find only the first two rows have leading one; therefore, the first vector is redundant in the spane. Therefore, we should be able to find number a and b in the following equation.

$$\begin{pmatrix} -4 \\ -4 \\ 7 \end{pmatrix} = a \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix} + b \begin{pmatrix} 0 \\ -2 \\ 3 \end{pmatrix}$$

Let's write each components of this vector equation seperately.

$$\begin{aligned} -4 &= -2a + 0b \\ -4 &= a + (-2)b \\ 7 &= -1a + 3b \end{aligned}$$

This is a just a new linear system, with three equations and two variables, a and b. We can solve it as we did before, either directly or using a matrix and row reduction. In this case, the linear system is solved by $a = 2$ and $b = 3$.

$$\begin{pmatrix} -4 \\ -4 \\ 7 \end{pmatrix} = 2 \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ -2 \\ 3 \end{pmatrix}$$

This example can be generalized. Any time some vector $u \in \$RR^n$ is in Span$\{v_1, v_2, \ldots, v_k\}$, there are constants a_i such that $u = a_1v_1 + a_2v_2 + \ldots + a_kv_k$. Writing each component of the vector equation seperately gives a system of n different linear equation, which we solve as before.

This gives a general algorithm for expressing vector in terms of a new basis. As we discussed before, there are many different bases for linear spaces. The standard basis for $\mathbb{R}^3$ is are the axis vectors e_1, e_2 and e_3. However, sometimes we might want to use a different basis. Any three linearly independent vectors in $\mathbb{R}^3$ form a basis.

Example 11.3.2. Take the basis $\left\{ \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 7 \end{pmatrix} \right\}$. Let's express the vector $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ in terms of this basis. We are looking for constants a, b and c that solve this vector equation.

$$\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = a \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix} + b \begin{pmatrix} -2 \\ 0 \\ -3 \end{pmatrix} + c \begin{pmatrix} 0 \\ 2 \\ 7 \end{pmatrix}$$

Writen in components, this gives the following linear system.

$$\begin{aligned} 3a - 2b + 0c &= 1 \\ -a + 0b + 2c &= 1 \\ -a + 3b + 7c &= 2 \end{aligned}$$

We translate this into a matrix.

$$\left(\begin{array}{ccc|c} 3 & -2 & 0 & 1 \\ -1 & 0 & 2 & 1 \\ -1 & 3 & 7 & 2 \end{array}\right)$$

We row reduce this matrix to get the reduced row-echelon form.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} \end{array}\right)$$

We read that $a = 0$, $b = \frac{1}{2}$ and $c = \frac{1}{2}$.

$$\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = 0 \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -2 \\ 0 \\ -3 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 2 \\ 7 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -2 \\ 0 \\ -3 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 2 \\ 7 \end{pmatrix}$$

11.4 Dimensions of Loci

Similarly, we can use row-reduction of matrices to find the dimensions of a locus. In particular, we will be able to determine if any of the equations were redundant. Recall that we described any affine or linear subspace of $\mathbb{R}^n$ as the locus of some finite number of linear equations: the set of points that satisfy a list of equations.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n &= c_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n &= c_m \end{aligned}$$

This is just a system of linear equations, which has a corresponding matrix.

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & c_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & c_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{11} & a_{12} & \ldots & a_{1n} & c_1 \end{array}\right)$$

So, by definition, the locus of a set of linear equation is just the geometric version of the solution space of the system of linear equations. Loci and solutions spaces are exactly the same thing; only loci are geometry and solution spaces are algebra. The dimension of a locus is same dimensions of a solution space of a system. Fortunately, we already know how to figure that out.

Proposition 11.4.1. *Consider a locus defined by a set of linear equations. Then the dimension of the locus is the number of free variables in the solution space of the matrix corresponding to the system of the equations. If the matrix contains a row that leads to a contradction $0 = 1$, then the locus is empty.*

To understand a locus, we just solve the associated system. If it has solutions, we count the free variables: that's the dimenion of the locus. We can get even more information from this process. When we row reduce the matrix A corresponding to the linear system, the equations corresponding to rows with leading ones (keeping track of exchanging rows, if necessary) are equations which are necessary to define the locus. Those which end up a rows without leading ones are redundant and the locus is unchanged if those equations are removed.

If the ambient space is $\mathbb{R}^n$, then our equations have n variables. That is, there are n columns to the right of the vertical line in the extended matrix. If the rank of A is k, and there are no rows that reduce to $0 = 1$, then there will be $n - k$ columns without leading ones, so $n - k$ free variables. The dimension of the locus is the ambient dimension n minus the rank of the matrix k.

This fits our intuition for spans and loci. Spans are built up: their dimension is equal to the rank of an associated matrix, since each leading one corresponds to a unique direction in the span, adding to the dimension by one. Loci are restictions down from the total space: their dimension is the ambient dimension minus the rank of an associated matrix, since each leading one corresponds to a real restriction on the locus, dropping the dimensions by one. In either case, the question of calculating dimension boils down to the rank of an associated matrix.

11.5 Bases for Loci

If we have a solution space for a linear system expressed in free variables, we can write it as a vector sum using those free variables.

Example 11.5.1. The following is an example of a solution space in $\mathbb{R}^4$, with variables w, x, y, and z, where y and z are free variables.

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

This is an offset span. The span is all combinations of the vectors $\begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ -3 \\ 0 \\ 1 \end{pmatrix}$. Therefore, those two vectors are the basis for the span. An offset space itself doesn't have a basis, since basis is only defined for linear subspaces.

This gives us a way to describe the basis for any locus: write it as a system, find the solution space by free parameter and idenfity the basis of the linear part.

Lecture 12

Linear Transformations

12.1 Definitions

After a good definition of the environment ($\mathbb{R}^n$) and its objects (lines, planes, hyperplanes, etc), the next mathematical step is to understand the functions that live in the environment and affect its objects. First, we need to generalize the simple notion of a function to linear spaces. In algebra and calculus, we worked with functions of real numbers. These functions are rules $f : A \to B$ which go between subsets of real numbers. The function f assigns to each *number* in A a unique *number* in B. They include the very familiar $f(x) = x^2$, $f(x) = \sin(x)$, $f(x) = e^x$ and many others.

Definition 12.1.1. Let A and B be subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. A *function* between linear spaces is a rule $f : A \to B$ which assigns to each *vector* in A a unique *vector* in B.

Example 12.1.2. We can define a function $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x^2 \\ y^2 \\ z^2 \end{pmatrix}$.

Example 12.1.3. Another function $f : \mathbb{R}^3 \to \mathbb{R}^2$ could be $f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x - y \\ z - y \end{pmatrix}$.

Definition 12.1.4. A *linear function* or *linear transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function $f : \mathbb{R}^n \to \mathbb{R}^m$ such that for two vectors $u, v \in \mathbb{R}^n$ and any scalar $a \in \mathbb{R}$, the function must obey two rules.

$$f(u + v) = f(u) + f(v)$$
$$f(au) = af(u)$$

Informally, we say that the function *respects* the two main operations on linear spaces: addition of vectors and multiplication by scalars. If we perform addition before or after the function, we get the same result. Likewise for scalar multiplication.

By inspection and testing, one could determined that the first example above fails these two rules, but the second example satisfies them.

This definition creates to the restrictive but important class of linear functions. We could easily define linear algebra as a study of these transformations. There is an alternative and equivalent definition of linear transformation.

Proposition 12.1.5. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if it sends linear objects to linear objects.*

Under a linear transformation points, lines, planes are changed to other points, lines, planes, etc. A line can't be bent into an arc or broken into two different lines. Hopefully, some of the major ideas of the course are starting to fit together: the two basic operations of addition and scalar multiplication give rise to spans, which are flat objects. Linear transformation preserve those operations, so they preserve flat objects. Exactly *how* they change these objects can be tricky to determine.

Lastly, because of scalar multiplication, if we take $a = 0$ we get that $f(0) = 0$. Under a linear transformation, the origin is always sent to the origin. So, in addition to preserving flat objects, linear transformation can't move the origin. We could drop this condition of preserving the origin to get another class of functions.

Definition 12.1.6. A *affine transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a transformation that preserves affine subspaces. These transformations preserve flat objects but may move the origin.

Though they are interesting, we don't spend much time with affine transformations. They can always be realized as a linear transformation combined with a shift or displacement of the entire space by a fixed vector. Since shifts are relatively simple, we can usually reduce problems of affine transformations to problems of linear transformations.

12.2 Composition

Once we understood functions of real numbers, we learned to compose them. We do the same for linear transformations.

Definition 12.2.1. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^l$ be linear transformations. Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^l$ is the linear transformation formed by first applying f and then g. Note that the $\mathbb{R}^m$ has to match: f outputs to $\mathbb{R}^m$, which is the input for g. Also note that the notation is written right-to-left: In $g \circ f$, the transformation f happens first, followed by g. This new transformation is called the *composition* of f and g.

12.3 Linear Transformations of $\mathbb{R}^2$

It is useful for us to specialize to transformations $\mathbb{R}^2 \to \mathbb{R}^2$, in order to build experience and intuition. As we noted above, linear transformations preserve flat objects. What can we do to $\mathbb{R}^2$ that preserves lines and preserves the origin?

Proposition 12.3.1. *There are five basic types of linear transformation of $\mathbb{R}^2$.*

- *Rotations* about the origin (either clockwise or counter-clockwise) preserve lines. We can't rotate around any other point, since that would move the origin. Since we can choose any angle, there are infinitely many such rotations. Also, since rotating by θ radians clockwise is the same as $2\pi - \theta$ radians counter-clockwise, we typically choose to only deal with counter-clockwise rotations. Counter-clockwise rotations are considered positive and clockwise rotations are considered negative.
- *Reflections* over lines through the origin preserve lines. The line of refleciton must passthrough the origin or else the reflection will move the origin. Since there are infinitely many different lines through the origin, there are infinitely many such reflections.
- *Skews* are a little tricker to visualize. A skew is a transformation that takes either veritcal or horizontal lines (but not both) and tilts them diagonally. It changes squares into parallelograms. The tilted lines are still lines, so it is a linear transformation.
- *Dialations* are transformation which stretch or shink in various directions.
- *Projections* are transformations which collapse $\mathbb{R}^2$ down to a line through the origin. Two important examples are projection onto either axis. Projection onto the x axis sends a point (a, b) to $(a, 0)$, removing the y component. Likewise, projection onto the y axis sends a point (a, b) to $(0, b)$, removing the x component. In a similar manner, we can project onto any line though the origin by sending each point to the closest point on the line. Finally, there is the projection to the origin which sends all points to the origin.

We present the following interesting theorem without proof.

Theorem 12.3.2. *All linear transformation of $\mathbb{R}^2$ are generated by composition of transformations of the previous five types.*

12.4 Linear Transformations of $\mathbb{R}^3$

We can also specialize to transformations of $\mathbb{R}^3$, though it is not as easy to give a complete account of the basic types. However, all of the types listed for $\mathbb{R}^2$ generalize.

- Rotations in $\mathbb{R}^3$ are no longer about the origin. Instead, we have to choose an axis. Any line through the origin will do for an axis of rotation. Any rotation in $\mathbb{R}^3$ is determined by an axis of rotation and an angle of rotation about that axis.
- Reflections are also altered: instead of reflecting over a line, we have to reflect over a plane through the origin. Any plane through the origin determines a reflection.
- Skews are similarly defined: one or two directions are fixed and the remaining directions are tilted.
- Dialations are also similar, though we have three possible axis directions in which to stretch or compress.
- Like $\mathbb{R}^2$, we can project onto the origin, sending everything to zero, or onto a line, sending every point to the closest point on a line. Examples include projection onto the axes. However, we can also project onto planes. Sending (a, b, c) to $(a, b, 0)$, for example, removes the z coordinate; this is projection onto the xy plane.

12.5 Symmetry

Notice that we've defined linear functions by the objects and/or properties they preserve. This is a very general technique in mathematics. Very frequently, functions are classified by what they preserve. As discussed in the very first chapter, we use the word 'symmetry' to describe this perspective: the symmetries of a function are the objects or algebraic properties preserved by the function. A function exhibits more symmetry if it preserves more objects or more properties. The conventional use of symmetry in English relates more to a shape than a function: what are the symmetries of a hexagon? We can connect the two ideas: asking for the symmetries of the hexagon can be thought of as asking for the transformations of $\mathbb{R}^2$ that preserve a hexagon. This is a bit of a reverse: the standard usage of the word talks about transformation as the symmetries of a shape. Here we start with a transformation and talk about the shape as a symmetry of the transformation: the hexagon is a symmetry of rotation by one sixth of a full turn.

12.6 Dihedral Groups

The case of regular polygons, such as the hexagon, is a very useful place for us to start. Transformations which do not preserve flat objects may not going to preserve the edges of the polygons. Therefore, we want to know which linear transformations preserve the regular polygons.

Definition 12.6.1. A *Dihedral Group* is the group of linear symmetries of a regular polygon. If the polygon has n edges, then we write D_n for the associated dihedral group. Thus, D_3 is the group of symmetries of the triangle, D_4 is the group of symmetries of the square, D_5 is the group of symmetries of the pentagon, and so on. (Some texts use D_{2n} instead of D_n.)

Example 12.6.2. Let's first consider the square. For all the polygons we are considering, we assume the origin is at the centre of the polygon. We can assume the square has vertices $(1,1)$, $(1,-1)$, $(-1,-1)$ and $(-1,1)$. Conveniently, the vertices are all we need to think about: since linear transformations preserve lines, if we know where the vertices go, then we know that the lines connecting them are preserved and hence the shape is preseved. Which transformations preserve the vertices?

There are two relatively obvious classes that might occur to us. First, there are rotation about the origin. If we rotate by $\pi/2$ radians (a quarter turn), the vertices are preserved. Likewise for rotations by π and $3\pi/2$ radians. If we are really paying attention, we might also think of rotation by 2π radians, a full turn. This might seem silly, but the identity transformation (the transformation that leaves everything fixed) is a very important transformation and should be included in the list. So, for now, we have the identity and three rotations.

We might also think about reflections. Looking at the square, we can find four lines of reflection: the x-axs, the y-axis, the line $x = y$ and the line $x = -y$. That brings the total number of transformation to eight: the identity, three rotations, and four reflections. These eight are, in face, the complete group.

You may have noticed that we used the work 'group' instead of 'set' to describe this collection. 'Group' is a technical word in mathematics, refering to a set with some extra properties. We won't go into the formal details, but we'll use the dihedral groups as archetypical examples. There are three properties of the dihedral groups which hold for all groups.

- The composition of two elements of the group is still in the group.
- The identity transformation is in the group.
- The inverse operation of any group element is still a group element.

Though we haven't formally defined inverses, the third property is not to hard to see for the dihedral groups. For the reflections, if we perform the reflection again we get back to where we started. The reflections are their own inverses. For rotations, we add another rotation so that we have completed a full turn. The inverse of rotation by $\pi/2$ radians is rotation by $3\pi/2$ radians. Equivalently, the inverse is rotation by $-\pi/2$ radians.

The first property is a little trickier. We leave it to you to try out some compositions of rotations and reflections to see how the result is still one of the eight symmetries. Likewise, we leave it to you to extend this analysis of D_4 to other diheadral groups: D_3, D_5, D_6 and so on.

All dihedral groups are composed of the identity, some rotations, and some reflections. D_4 is a useful enough example that we should fix notation. I or Id is the identity transformation, R_1 is rotation by $\pi/2$ radians, R_2 is rotation by πradians and R_3 is rotation by $3\pi/2$ radians. F_1 is reflection over the x axis, F_2 over $x = y$, F_3 over the y axis and F_4 over $y = -x$.

Lecture 13

Matrix Representation of Linear Transformations

Now we come to the second eajor application of matrices. In addition to succintly encoding linear systems, matrices can also be used very efficiently to encode linear transformations. This is done by defining how a matrix can act on a vector.

Definition 13.0.1. Let $A = a_{ij}$ be a $m \times n$ matrix and let v be a vector in $\mathbb{R}^n$. There is an action of A on v, written Av, which defines a new vector in $\mathbb{R}^m$. That action is given in the following formula.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \ldots + a_{2n}v_n \\ \vdots \\ a_{n1}v_1 + a_{n2}v_2 + \ldots + a_{nn}v_n \end{pmatrix}$$

This is a bit troubling to work out in general. Let's see what it looks like slightly more concretely in $\mathbb{R}^2$ and $\mathbb{R}^3$.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + yb \\ cx + dy \end{pmatrix}$$

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{pmatrix}$$

In this way, all $m \times n$ matrices determine a method of sending vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$: a function $\mathbb{R}^n \to \mathbb{R}^m$. It is not at all obvious from the definition, but matrices completely describe all linear transformations.

Proposition 13.0.2. *If A is a $m \times n$ matrix, then the associated function defined by the matrix action is a linear function $\mathbb{R}^n \to \mathbb{R}^m$. Moreover, all linear functions $\mathbb{R}^n \to \mathbb{R}^m$ can be encoded this way. Finally, each linear function is encoded uniquely, i.e., each $m \times n$ matrix corresponds to a different transformation.*

In this way, the set of linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ is *exactly* the same as the set of $m \times n$ matrices. This is a very powerful result: in order to understand linear transformations of Euclidean space, we only have to understand matrices and their properties.

Let's see what certain special matrices mean as transformations. We'll look at 3×3 matrices for these examples.

Example 13.0.3. First, we had the zero matrix: all coefficient are zero.

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0x + 0y + 0z \\ 0x + 0y + 0z \\ 0x + 0y + 0z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The zero matrix corresponds to the projection that sends all vectors to the origin.

Example 13.0.4. We also had the identity matrix: ones on the diagonal and zeros elsewhere.

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

The identity matrix corresponds to the transformation which doesn't change anything. Appropriately, we called this the identity transformation.

Example 13.0.5. Diagonal matrices only have non-zero entires on the diagonal.

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + 0y + 0z \\ 0x + by + 0z \\ 0x + 0y + cz \end{pmatrix} = \begin{pmatrix} ax \\ by \\ cz \end{pmatrix}$$

This is a dialation: the x direction is stretched by the factor a, the y direction by the factor b and the z direction by the factor c. Diagonal matrices are dialations.

13.1 Composition and Matrix Multiplication

We previously defined the composition of linear transformation. Compositon allows us to combine transformations: we can ask what happens if we first rotate and then reflect in $\mathbb{R}^2$. However, since matrices represent transformations, this compsotion should somehow be accounted for in the matrix representation. If A is the matrix of S and B is the matrix of T, what is the matrix of $S \circ T$? The answer is given by matrix multiplication.

Definition 13.1.1. Let A be a $k \times m$ matrix and B a $m \times n$ matrix. We can think of the *rows* of A as vectors in $\mathbb{R}^m$, and the *columns* of B as vectors in $\mathbb{R}^m$ as well. To emphasise this perspective, we write the following, using u_i for the rows of A and v_i for the columns of B.

$$A = \begin{pmatrix} \to & u_1 & \to \\ \to & u_2 & \to \\ \to & u_3 & \to \\ \vdots & \vdots & \vdots \\ \to & u_k & \to \end{pmatrix} \qquad B = \begin{pmatrix} \downarrow & \downarrow & \downarrow & \dots & \downarrow \\ v_1 & v_2 & v_3 & \dots & v_n \\ \downarrow & \downarrow & \downarrow & \dots & \downarrow \end{pmatrix}$$

With this notation, the *matrix multiplication* of A and B is the $k \times n$ matrix where the entires are the dot products of rows and columns.

$$AB = \begin{pmatrix} u_1 \cdot v_1 & u_1 \cdot v_2 & u_1 \cdot v_3 & \dots & u_1 \cdot v_n \\ u_2 \cdot v_1 & u_2 \cdot v_2 & u_2 \cdot v_3 & \dots & u_2 \cdot v_n \\ u_3 \cdot v_1 & u_3 \cdot v_2 & u_3 \cdot v_3 & \dots & u_3 \cdot v_n \\ \vdots & \vdots & \vdots & \dots & \vdots \\ u_k \cdot v_1 & u_k \cdot v_2 & u_k \cdot v_3 & \dots & u_k \cdot v_n \end{pmatrix}$$

This operation has the desired property: the product of the matrices repesents the composition of the transformations. (This remarkable fact is presented here without proof; I'll leave it to you to wonder why this wierd combinations of dot products has the desired geometric interpretation.) Remember that the composition still works from right to left, so that the matrix multiplication AB represents the transformation associated to B first, followed by the transformation associated to A. When we think of matrices acting on vectors, we wrote the action on the right: Av. Now when when a composition acts, as in ABv, the closest matrix gets to act first.

We have defined a new algebraic operation. As with the new products for vectors (dot and cross), we want to know the properties of this new operation.

Proposition 13.1.2. *Let A be a $k \times l$ matrix, B and C be $l \times m$ matrices, and D a $m \times n$ matrix. Also let I_i be the identity matrix in $\mathbb{R}^i$ for any $i \in \mathbb{N}$. Then there are three important properties of matrix multiplication.*

$$\begin{array}{rcll} A(BD) & = & (AB)D & \textit{Associative} \\ A(B+C) & = & AB + AC & \textit{Distributive} \\ AI_l & = & I_k A = A & \textit{Identity} \end{array}$$

Note the lack of commutativity: $AB \neq BA$. In fact, if A and B are not square matrices, if AB is defined and BA will not be; the indices will not match. Not only are we unable to exchange the order of matrix multiplication; sometimes that multiplication doesn't even make sense as an operation. Matrix multiplication is a very important example of a non-commutative product.

Lecture 14

Inverse Linear Transforms and Matrix Inversion

14.1 Inverse Transforms

When we defined Dihedral Groups, we already talked briefly about inverse linear transformation. Let's give a general definition.

Definition 14.1.1. Let $T : \mathbb{R}^n \to \mathbb{R}^n$. The inverse transformation of T is a transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ which undoes what T did. Equivalent, it is a transformation S such that the compositions $T \circ S$ and $S \circ T$ are the identity. For an arbitrary transformation, there is no guarantee that an inverse exists; if it does, it is written T^{-1}.

This is the geometric definition. Now that we've encoded transformations as matrices, we can use matrices and ask for an algebriac definition of the inverse as well.

14.2 Inverse Matrices

The identity property: $AI_l = I_k A = A$ reminds us of multiplication by one in $\mathbb{R}$. This is the multiplication that doesn't accomplish anything. In number systems, when we have an identity, we usually have a way of getting back to that identity. Zero is the identity in addition. For any number a, we have $(-a)$ so that $a + (-a) = 0$ gets us back to the identity. One is the identity in multiplication. For any *non-zero* number a, we have $\frac{1}{a}$ so that $a\frac{1}{a} = 1$ gets us back to the identity.

For matrices, we have the same question. For any matrix M, is there another matrix N such that $MN = I$? Multiplication of numbers alreay shows us that we need to take care: for the number zero, there is no such number. For matrices, we have to be even more cautious.

Definition 14.2.1. Let M be a $n \times n$ (square) matrix. The the *inverse* of M is the *unique* matrix M^{-1} (if it exists) such that $MM^{-1} = M^{-1}M = I_n$.

We should note a couple of thing about this definition. First, it only applies to square matrices. We don't even try to invert non-square matrices. Second, we need to have both orders of multiplication MM^{-1} and $M^{-1}M$. This is due to the previous observation that matrix multiplication is non-commutative. In general, these two orders could result in different products. In this case, we insist that both orders get us back to the identity.

14.3 Calculating Inverse Matrices

The definition is good, but we are left with the problem of determining which matrices have inverses and calculating those inverses. It turns out there is a convenient algorithm using techniques we already know. We write an extended matrix $(M|I_n)$ with our original matrix and the identity next to each other, separated by a vertical line. Then we row reduce, treating the conglomeration as one matrix with long rows. If our row reduction of M results in the identity on the left, then the matrix has the form $(I|M^{-1})$; the right hand side will be the inverse.

This algorithm gives us two observations. First, obviously, it is a way to calculate inverses. But it also is a condition: an $n \times n$ matrix is invertible, in this algorithm, only if we get the identity matrix on the left after row reduction. This is part of a general result, which is the first of several conditions for intertible matrices.

Proposition 14.3.1. *A $n \times n$ matrix is invertible if and only if it row reduces to the identity matrix I_n. Since the identity has n leading ones, this is equivalent to the matrix having rank n.*

Definition 14.3.2. The set of all intervible $n \times n$ matrices with real coefficient forms a group. It is called the *General Linear Group* and written $GL_n(\mathbb{R})$. If we wanted to change the coefficients to another set of scalars S, we would write $GL_n(S)$. When the coefficients are understood, we sometime write GL_n or $GL(n)$.

Lecture 15

Transformations of Spans and Loci

15.1 Transformations of Spans

Now that we have matrices to describe transformations, we want to know how transformations affect to linear subspaces. We know that they preserve them, but we want to be more specific: given a particular linear subspace, can we determine where it goes under a transformation?

Definition 15.1.1. If $M : \mathbb{R}^n \to \mathbb{R}^m$ is a transformation (represented by a $m \times n$and L is a linear or affine subspace of $\mathbb{R}^n$, then the result of the matrix M acting on all vectors in L will be a linear or affine subspace of $\mathbb{R}^m$ called the *image of L under M*. The image of all of $\mathbb{R}^n$ under M is simply called the *image of M*.

The most important property of linear transformations is that they preserve addition and scalar multiplication. That means they preserve the linear combintaions. Let a_i be scalars and v_i be vectors in $\mathbb{R}^n$, with M a $m \times n$ matrix.

$$\begin{aligned} M(a_1v_1 + a_2v_2 + \ldots + a_kv_k) &= M(a_1v_1) + M(a_2v_2) + \ldots + M(a_kv_k) \\ &= a_1(Mv_1) + a_2(Mv_2) + \ldots + a_k(Mv_k) \end{aligned}$$

We see that the image of a linear combination is still a linear combination, just in the vectors Mv_i instead of v_i. Since all elements of a span are linear combinations, we can say that everything in the span of the v_i is sent to the span of the vectors Mv_i. We can summarize these observations in a proposition.

Proposition 15.1.2. *A linear transformation represented by a $m \times n$ matrix M sends spans to spans. In particular, the span* $\text{Span}\{v_1, v_2, \ldots, v_k\}$ *is sent to* $\text{Span}\{Mv_1, Mv_2, \ldots, Mv_k\}$.

Matrices acting on spans are easy: we just figure out where the individual vectors go and the span of those vectors will be transformed into the span of their images. Offset spans are almost as easy: consider an affine subspace $u + \text{Span}\{v_1, v_2, \ldots, v_k\}$. Any element of this looks like $u + a_1v_1 + a_2v_2 + \ldots a_kv_k$. Under M, we have:

$$M(u + a_1v_1 + a_2v_2 + \ldots + a_kv_k) = Mu + a_1Mv_1 + a_2Mv_2 + \ldots + a_kMv_k$$

The offset span is still sent to an offset span, with offset Mu and spanning vectors Mv_i.

Be careful that the matrix need not preserve the dimension of the span. Even if the v_i are linearly independent and form a basis for the span, the vectors Mv_i need not be linearly independent. The dimension of the new span might be smaller.

15.2 Transformation of Loci

For loci, the picture is much more complicated. Equations do not transform nearly as pleasantly as spans. Planes in $\mathbb{R}^3$ are defined by a normal; we might hope that the new plane is defined by the image of the normal. Unfortunatly, since the matrix may not preserve orthogonality, this will usually not happen. To determine the image of a locus, the best solution is to describe the locus as a span or offset span find the image of the that span.

Lecture 16

Kernels and Images

16.1 Row and Column Spaces

Definition 16.1.1. Let M be a $m \times n$ matrix. The *rowspace* of the matrix is the span of the row of the matrix, thought of as vectors in $\mathbb{R}^n$. It is a subspace of $\mathbb{R}^n$. The *columnspace* of the matrix is the span of the columns of the matrix, thought of as vectors in $\mathbb{R}^m$. It is a subspace of $\mathbb{R}^m$.

Proposition 16.1.2. *Let M be a $m \times n$ matrix. The columnspace and the row spaces of M are linear subspaces (of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively) with dimension equal to the rank of M.*

Proposition 16.1.3. *The image of a matrix M is the same as its columnspace.*

Proof. The image is the set of all outputs of M, acting on the whole space $\mathbb{R}^n$. We can think of $\mathbb{R}^n$ as the span of its standard basis $e_1, e_2, \ldots e_n$. The image, then, is the span of $Me_1, Me_2, \ldots Me_n$. These vectors are precisely the columns of the matrix, so the their span is the columnspace. □

Example 16.1.4. Look at a 3×3 vector acting on the standard basis of $\mathbb{R}^3$ to see how the images of the basis vectors are the columns of the matrix.

$$\begin{pmatrix} -2 & -2 & 0 \\ 7 & -3 & 2 \\ -4 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 7 \\ -4 \end{pmatrix}$$
$$\begin{pmatrix} -2 & -2 & 0 \\ 7 & -3 & 2 \\ -4 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \\ -1 \end{pmatrix}$$
$$\begin{pmatrix} -2 & -2 & 0 \\ 7 & -3 & 2 \\ -4 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$$

Definition 16.1.5. Let M be a $m \times n$ matrix representing a transformation $\mathbb{R}^n \to \mathbb{R}^m$. For any $u \in \mathbb{R}^m$, the *preimage* of u is the set of all vectors in $\mathbb{R}^m$ which are mapped to u by M. It is writtem $M^{-1}\{u\}$. The *kernel* or *nullspace* of M is all vectors in $\mathbb{R}^n$ which are sent to the zero vector under the transformation associated to M, i.e., the preimage of the origin.

Proposition 16.1.6. *The preimage of u is the solution space of the system of equations associated to the extended matrix $(M|u)$ where we add the vector u as a column to the matrix M.*

Proof. The preimage is all vectors v in $\mathbb{R}^n$ that are sent to u. If we write v in coordinates $x_1, \ldots, x_n$, and apply the matrix action, we get a system of m equations with the entries of u as the constants. The matrix encoding of this system is precisely the extended matrix $(M|u)$. □

Example 16.1.7. We can look at $\mathbb{R}^2$ to build our intuition.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

If we write this out in terms of coordinates, we get

$$ax + by = x_0$$
$$cd + dy = y_0$$

This is the system associated to the extended matrix with $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ as constants.

The connection between kernels, preimages and solutions spaces is a deep and important connection between the two pieces of this course so far: vector geometry and solving linear system. When we started with a linear system, we wrote it as an extended matrix. The solution space of that extended matrix ($(M|u)$, where u is the column vector as constants) is the same as the vectors $v \in \mathbb{R}^n$ such that $Mv = u$. Solutions spaces are preimages and preimages are solutions spaces.

Specifically, the kernel is the linear space of all vector v with $Mv = 0$. This is a solution space for the linear system with matrix $(M|0)$, where the column after the dividing line is all zeroes. This gives us a way of calcluation kernels and expressing them as spans.

Lecture 17

Determinants

17.1 Definition

Linear transformations are defined by their symmetries: they preserve linear subspaces and linear operations. We use associated matrices to understand these transformations. Matrices are an algebraic description, so that questions of transformations can be turned into algebraic problems. To figure out which points went to zero under a transformation, we had an algebraic method of calculating the kernel as a solution space. Now that we have the algebraic description of transformations as matrices, we want to investigate what else the matrix can tell us about the transformation. In this section, we are asking two questions in particular.

First, what does the transformation do to the size of objects? I choose the vague word 'size' intentionally because we work in various dimensions. In one dimension, size is length. In two dimensions, size is area. In three dimensions, size is volume. And in higher dimensions, we have some extended notion of volume that fits that dimension; we can call this hyper-volume. It is important to note that size depends on the ambient dimension. A square in $\mathbb{R}^2$ has some area, some non-zero size. But if we have a flat square somewhere in $\mathbb{R}^3$, it has no volume, therefore no substantial size in that space.

Second, what does the transformation do to the orientation of an object? We defined orientation as a choice of axis directions relatively to each other, but that isn't the most enlightening direction. To help, we will describe orientation for each dimension. In $\mathbb{R}$, orientation is direction: moving in the positive or negative direction along the numberline. There are only two directions of movement, so two orientations. If we have a transformation of $\mathbb{R}$, we can ask if it changes or preserves these directions. In $\mathbb{R}^2$, instead of moving in line, we think of moving in closed loops or paths. These paths can be clockwise or counter-clockwise. Then we can ask if a transformation changes clockwise loops into other clockwise loops or into counter-clockwise loops. The axis system in $\mathbb{R}^3$ is, conventionally, given by a right-hand-rule. If we know the x and y directions, the right-hand-rule indicates the positive

z direction. Then we can ask where these three directions go under a transformation and if a right-hand-rule still applies. If it does, we preserve the origientation. If it doesn't, and a left-hand-rule would work instead, the transformation reverses orientation. In higher dimension, there are other extentions of the notion of orientation. In each case, the question is binary: a transformation either preserves or reverses orientation.

Definition 17.1.1. Let M be a square $n \times n$ matrix. The *determinant* of M is a real number, written $\det M$, with two properties. Its absolute value, $|\det M|$, measures the effect that M has on size (length, area, volume, hyper-volume). Its sign (positive or negative) measures the effect on orientation; if the sign is positive, orientation is preserved, and if the sign is negative, orientation is reversed. For a succinct notation, we replace the curved braces with straight lines to indicate the determinant of a matrix.

That definition is all well and good, but we need to show that such a thing can be constructed. The next section gives an algorithm for building determinants.

17.2 Cofactor Expansion

The determinant of a (square) matrix is constructed recursively, starting with 2×2 matrices.

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

Example 17.2.1. For the matrix $\begin{pmatrix} 1 & -3 \\ 2 & -1 \end{pmatrix}$, the determinant is $(1)(-1) - (2)(-3) = -1 + 6 = 5$. Therefore, we know that this matrix multiplies all areas by a factor of 5 and preserves orientation.

Definition 17.2.2. The algorithm for calculating determinants in general is call *Co-Factor Expansion*. It is a recursive algorithm that reduces determinant calculation to determinants of smaller square matrices, eventually down to 2×2 matrices.

Co-factor expansion procceds in this way: we choose *any* column or row of the matrix. We take the coefficients from that column or row. For each of the coefficients, we multiply that coefficient by the detminant of the matrix formed by removing both the columnn and row containig that coefficient. Then we add up the determinants of these small matrices multiplied by matching coefficients, with a pattern of $\pm$ signs. That pattern of $\pm$ signs is a checkerboard pattern, as in the following 5×5 matrix.

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

That's a hard algorithm to intuit from the formal description. Let's do a number of examples.

Example 17.2.3. Here is a 3×3 example where we choose the first row for cofactor expansion.

$$\begin{vmatrix} 5 & -2 & 0 \\ -3 & 3 & -2 \\ 1 & -5 & 3 \end{vmatrix} = (+1)5 \begin{vmatrix} 3 & -2 \\ -5 & 3 \end{vmatrix} + (-1)(-2) \begin{vmatrix} -3 & -2 \\ 1 & 3 \end{vmatrix} + (+1)(0) \begin{vmatrix} -3 & 3 \\ 1 & -5 \end{vmatrix}$$
$$= 5((3)(3) - (-2)(-5)) + 2((-3)(3) - (-2)(1)) + (0)(-3)(-5) - (3)(1)$$
$$= 5(-1) + 2(-7) = -19$$

If we did cofactor expansion of an arbitrary 3×3 matrix, we would get a direct expression for the determinant.

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei - ahf - bdi + bfg + cdh - ceg$$

We could use this formula for 3×3 matrices, if we wished. We could do the same for larger matrices, but it starts to get quite complicated. The computational complexity of the recursive determinant algorithm grows very quickly. For a 3×3 matrix, we had 3 recursions to 2×2 matrices, each with 2 multiplication terms in the determinant, giving 6 terms. Each term was the multiple of three coefficients, giving 12 total multiplications. For a 4×4, the recursions gives 24 terms, each with 4 coefficients, so $24 \cdot 3 = 72$ multiplications. For a 5×5 matrix, we recurse five times to a 4×4 matrix, for 120 terms each with 5 coefficients, which is $120 \cdot 4 = 600$ multiplications. The pattern continues: there are $n!$ terms in the determinant of a $n \times n$ matrix, each with n coefficient, for $n!(n-1)$ multiplications. This is computationally terrifying, making determinants of large matrices computationally very difficult.

17.3 Triangular Matrices

The following defintion will be used in the activity associated to this lecture.

Definition 17.3.1. An *upper triangular* matrix is a matrix where all entries below the diagonal are zero. A *lower triangular* matrix is a matrix where all entries above the diagonal are zero. Below are two examples of upper triangular matrices.

$$\begin{pmatrix} -1 & 3 & 5 \\ 0 & 2 & -9 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & -7 & 4 & 0 \\ 0 & 0 & 9 & -4 \\ 0 & 0 & 0 & -4 \end{pmatrix}$$

17.4 Properties of Determinants

Algebra considers sets with structre. On the sets $M_n(\mathbb{R})$ of $n \times n$ matrices, the determinant is a new algebraic structre. We would like to investigate how it interacts with existing structures, starting with matrix multiplication.

To compose two transformations, we multiply the matrices. In this composition, the effects on size should compound. If the first transformation doubles size and the second triples it, the composition should multiply size by a factor of 6. Similarly with orientation: if both preserve orientation, the composition should as well. If both reverse orientation, reversing twice should return us to the normal orientation. So, two positive or two negatives should result in a positive, and one positive and one negative should result in a negative.

Proposition 17.4.1. *Let A and B be two $n \times n$ matrices. Then* $\det(AB) = \det(A)\det(B)$.

Proposition 17.4.2. *If A is an invertible $n \times n$ matrix, then* $\det A^{-1} = \frac{1}{\det A}$.

In particular, A invertible means that $\det A$ must be non-zero. For A to be invertible, it must have a non-zero determinant. It turns out this property is sufficient as well as necessary. A non-zero determinant preserves the dimension: it may increase or decrease size but it doesn't entirely destroy it. Therefore, there is always a way to go back, to reverse the process.

Proposition 17.4.3. *A $n \times n$ matrix A is invertible if and only if* $\det A \neq 0$.

We can make a list of criteria for invertibility, including this new determinant condition.

Proposition 17.4.4. *Let A be an $n \times n$ matrix. All of the following properties are equivalent.*

- *A is invertible.*
- *A has rank n.*
- *A row reduces to the identity matrix.*
- $\mathrm{Ker}(A) = \{0\}$.
- $\mathrm{Im}(A) = \mathbb{R}^n$.
- *The columnspace of A is $\mathbb{R}^n$.*
- *The columns of A are linearly independent.*
- *The rowspace of A is $\mathbb{R}^n$.*
- *The rows of A are linearly independent.*
- *$Au = v$ has a unique solution u for any choice of $v \in \mathbb{R}^n$.*
- $\det A \neq 0$.

Here are two more properties of determinants.

Proposition 17.4.5. *Let A and B be $n \times n$ matrices and let $\alpha \in \mathbb{R}$ be a scalar.*

$$\det(AB) = \det(BA)$$
$$\det \alpha A = \alpha^n \det A$$

Row reduction is an important operations on matrices. It is convenient to know how the determinant changes under row operations.

Proposition 17.4.6. *Let A be a $n \times n$ matrix. If we exchange two rows of A, the determinant changes sign (positive to negative, negative to positive). If we multiply a row by a constant α, the determinant is multiplied by the same constant α. If we add a multiple of one row to another, the determinant is unchanged.*

This last property is perhaps surprising: adding one row to another, even with a multiple, doesn't change the determinant. This also gives us another algorithm for calculating determinants: we can row reduce the matrix to a triangular matrix and calculate the determinant of the triangular matrix by multiplying the diagonal elements. If we kept track of the multiplication by constants and the row exchanges, we can use these rules to calculate the determinant of the original matrix.

All of the properties of determinant listed so far have been multiplicative. The situation for matrix addition and determinants is less elegant: $\det(A + B)$ has no pleasant identity. This is an interesting contrast from many of the thing in this course: determinants are not linear functions $M_n(\mathbb{R}) \to \mathbb{R}$ since they do not act nicely with addition. Instead, they act nicely with multiplication. On sets with various algebraic structure, it is common for an operation to interact well with one structure but poorly with the others.

Lecture 18

Orthogonality and Symmetry

18.1 Symmetry, Again

Throughout the course, we've talked about symmetry and defined types of transformation by their symmetries, that is, by what they preserve. We started the entire section of transformation with a symmetry property: linear transformations have the symmetry of preserving flat object (linear subspaces). We defined the dihedral groups as the linear transformation which preserved certain polygons.

This persepctive is only strengthened as we get further into the discipline of linear algebra. Our classification of matrices, at least geometrically, is almost always a question of symmetry. In this section, we're going to define some new classes of matrices in terms of their symmetries.

Recall the notation we're already established for matrix sets and groups (with real coefficients).

- The set of $n \times m$ matrices is written $M_{n,m}(\mathbb{R})$.
- The set of square $(n \times n)$ matrices is written $M_n(\mathbb{R})$.
- The group of invertible matrices is called the general linear group and written $GL_n(\mathbb{R})$ or $GL(n)$.

18.2 Matrices of Determinant One

Determinants measure two geometric properites: size and orientation. These are both symmetries. If we have matrices of determinant ± 1, we know that they preserve size. If we have matrices of positive determinant, we know that they preserve orientation. If we have matrices of determinant one, they preserve both.

Definition 18.2.1. The group of matrices of determinant one with real coefficients is called the *Special Linear Group* and is written $SL_n(\mathbb{R})$ or $SL(n)$.

18.3 Transposes

The following definition will be useful in this chapter.

Definition 18.3.1. The *transpose* of a matrix is the matrix formed by switching the rows and columns. Alternatively, it is the mirror of the matrix over the diagonal. If $A = a_{ij}$ is a matrix expressed in indices, then the transpose is written A^T and has entries a_{ji} with indices switched.

$$A = \begin{pmatrix} 3 & 0 & 6 & -1 \\ 4 & -2 & -7 & 0 \\ 6 & -6 & -1 & -2 \end{pmatrix} \qquad A^T = \begin{pmatrix} 3 & 4 & 6 \\ 0 & -2 & -6 \\ 6 & -7 & -1 \\ -1 & 0 & -2 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 0 & 8 & -5 \\ 0 & 1 & 1 & -2 \\ 0 & 6 & 6 & -1 \\ 0 & 0 & -4 & -4 \end{pmatrix} \qquad A^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 6 & 0 \\ 8 & 1 & 6 & -4 \\ -5 & -2 & -1 & -4 \end{pmatrix}$$

18.4 Orthogonal and Special Orthogonal Matrices

Definition 18.4.1. An $n \times n$ matrix A is called *orthogonal* if it preserves lengths. That is, for all $v \in \mathbb{R}^n$, $|v| = |Av|$.

Preservation of lengths is not the only way to define orthogonal matrices. The following proposition shows the rich structure of this group of matrices. Before stating the proposition, here is a useful definition.

Definition 18.4.2. Let i, j be two indices which range from 1 to n. The *Kronecker delta* is a twice-indexed set of numbers δ_{ij} which evaluates to 0 whenever $i \neq j$ and 1 whenever $i = j$.

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Proposition 18.4.3. *Let A be an $n \times n$ matrix. Write $A = (a_1, a_2, \ldots, a_n)$ where the a_i are column vectors. Then the following seven statements are all equivalent and all can be taken as the definition of an orthogonal matrix.*

(a) A preserves lengths: $|v| = |Av|$ for all $v \in \mathbb{R}^n$.
(b) A preserves dot products: $u \cdot v = (Au) \cdot (Av)$ for all u, $v \in \mathbb{R}^n$.
(c) $A^T = A^{-1}$.
(d) $A^T A = \mathrm{Id}_n$.
(e) $a_i \perp a_j$ for all $i \neq j$ and $|a_i| = 1$.
(f) $a_i \cdot a_j = \delta_{ij}$

(g) *A preserves angles (the angle between u and v is the same as the angle between Au and Av for all u, $v \in \mathbb{R}^n$)* amd $\det A = \pm 1$.

There are several very surprising statements here. First, preserving angles and preserving lengths are nearly equivalent (preserving angles needs the additional determinant condition). This is a little odd; there isn't a strong intuition that exactly the same transformations should preserve both properties. Second, we have a convenient algebraic method of identifying orthogonal matrices in property d). We can multiply by the transpose: if we get the identity, we have an orthogonal matrix. The equivalence of property d) and property f) is seen in simply calculating the matrix multiplication; it involves the dot products of all the columns, which must evaluate to 0 or 1 exactly matching the Kronecker delta.

Orthogonal matrices can be though of as rigid-body transformations. Since they preserve both lengths and angles, they preserve the shape of anything formed of vertices and lines: any polygon, polyhedron, of higher dimensional analogue. They may not preserve the position (they may be moved around, rotated, reflected, etc) but they will preserve the shape. This makes orthogonal matrices extremely useful, since many applications want to use transformations that don't destroy polygons or polyhedra.

Proposition 18.4.4. *Here are some other properties of orthogonal matrices.*

- *In $\mathbb{R}^2$, the only orthogonal transformations are the identity, the rotations and the reflections.*
- *If A is orthogonal, so if A^{-1}.*
- *All orthogonal matrices have determinant ± 1.*
- *If A and B are orthogonal, then so are AB and BA.*

Even in $\mathbb{R}^3$, orthogonality already gets a bit trickier than just the rotations/reflection of $\mathbb{R}^2$. Consider the following matrix.

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

This matrix is orthogonal, but it isn't a rotation or relection in $\mathbb{R}^3$. It is some kind of 'reflection through the origin' where every point is sent to the opposite point with respect to the origin. (The equivalent in $\mathbb{R}^2$ is rotation by π radians). It isn't even a physical transformation: it's not something we can do with a physical object without destroying it. However, it satisfies the condition of orthogonality.

Definition 18.4.5. The group of orthogonal $n \times n$ matrices is called the *Orthogonal Group* and is written $O_n(\mathbb{R})$ or $O(n)$. The group of orthogonal matrices which also preserve orientation (have determinant one) is called the *Special Orthogonal Group* and is written $SO_n(\mathbb{R})$ or $SO(n)$.

18.5 Orthogonal Bases and the Gram-Schmidt Process

Definition 18.5.1. A set of linearly independent non-zero vectors $\{v_1, \ldots, v_k\}$ is called an *orthogonal set* if $v_i \cdot v_j = 0$ whenever $i \neq j$. The set is called an *orthonormal set* if $v_i \cdot v_j = \delta_{ij}$. If the set is a basis for a linear subspace, we use the terms *orthogonal basis* and *orthonormal basis*.

In many ways, an orthonomal basis is the best kind of basis for a linear subspace: all the vectors are perpendicular to each other and all the vectors are unit vectors. The standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$ is the classic example of a orthonormal basis; all other orthonomal bases are trying to mimic the standard basis.

Given a set of linearly independent vectors $\{v_1, \ldots, v_k\}$ (usually a basis for a linear subspace), there is a general algorithm for adjusting them into an orthogonal or orthonomal basis. This algorithm is called the Gram-Schmidt algorithm. We define new vectors u_i as follows.

$$
\begin{aligned}
u_1 &= v_1 \\
u_2 &= v_2 - \text{Proj}_{u_1} v_2 \\
u_3 &= v_3 - \text{Proj}_{u_1} v_3 - \text{Proj}_{u_2} v_3 \\
u_4 &= v_4 - \text{Proj}_{u_1} v_4 - \text{Proj}_{u_2} v_4 - \text{Proj}_{u_2} v_4 \\
&\vdots \\
u_k &= v_k - \text{Proj}_{u_1} v_k - \text{Proj}_{u_2} v_k - \ldots - \text{Proj}_{u_{k-1}} v_4
\end{aligned}
$$

Conceptually, this algorithm takes the first vector v_1 as the first direction. Then in the second step, we take v_2 and subtract any portion with lies in the u_1 direction. Removing that commonality gives a perpendicular vector. In the third step, we take v_3 and remove the projection onto the first two pieces. This again removes any commonality that we had so far, giving a perpendicular direction. We proceed in the same way through the list. The result is an orthogonal basis $\{u_1, \ldots, u_k\}$.

The result is not, however, orthonormal. To get an orthonormal set, we divide each u_i by its length.

$$\left\{ \frac{u_1}{|u_i|}, \frac{u_2}{|u_2|}, \ldots, \frac{u_k}{|u_k|} \right\}$$

In this way, we can take any basis and turn it into an orthogonal or othonormal basis.

Example 18.5.2.

$$\begin{aligned}
&\text{Span}\left\{\begin{pmatrix}1\\1\\3\end{pmatrix},\begin{pmatrix}0\\2\\2\end{pmatrix}\right\}\\
u_1 &= \begin{pmatrix}1\\1\\3\end{pmatrix}\\
u_2 &= \begin{pmatrix}0\\2\\2\end{pmatrix} - \frac{\begin{pmatrix}0\\2\\2\end{pmatrix}\cdot\begin{pmatrix}1\\1\\3\end{pmatrix}}{\left|\begin{pmatrix}1\\1\\3\end{pmatrix}\right|^2}\begin{pmatrix}1\\1\\3\end{pmatrix}\\
&= \begin{pmatrix}0\\2\\2\end{pmatrix} - \frac{8}{11}\begin{pmatrix}1\\1\\3\end{pmatrix}\\
&= \frac{1}{11}\begin{pmatrix}-8\\14\\-2\end{pmatrix}\\
|u_1| &= \sqrt{11}\\
\frac{u_1}{|u_1|} &= \frac{1}{\sqrt{11}}\begin{pmatrix}1\\1\\3\end{pmatrix}\\
|u_2| &= \frac{1}{11}\sqrt{264}\\
\frac{u_2}{|u_2|} &= \frac{1}{\sqrt{264}}\begin{pmatrix}-8\\14\\-2\end{pmatrix}
\end{aligned}$$

Example 18.5.3. Sometimes, not all the work is required. Say we are asked to find an orthonormal basis for the following linearly independent set.

$$\left\{\begin{pmatrix}3\\1\\-2\\-2\end{pmatrix},\begin{pmatrix}-3\\0\\0\\1\end{pmatrix},\begin{pmatrix}2\\0\\2\\0\end{pmatrix},\begin{pmatrix}0\\3\\1\\1\end{pmatrix}\right\}$$

We could go through the whole process, but since these are four linearly independent vectors in $\mathbb{R}^4$, they must span the whole space. Therefore, we can take e_1, e_2, e_3 and e_4, the standard basis, as the orthonormal basis. No calculation was required.

Lecture 19

Conjugation

19.1 Definitions

Definition 19.1.1. Let A be an $n \times n$ matrix. The *order* of A is the smallest positive integer n such that $A^n = \mathrm{Id}$. If no such integer exists, the order of A is ∞.

Definition 19.1.2. Let A be an $n \times n$ matrix and $v \in \mathbb{R}^n$. The v is a *fixed point* of A if $Av = v$.

Definition 19.1.3. Let A and B be $n \times n$ matrices with B invertible. The *conjugation* of A by B is the matrix $B^{-1}AB$. (In some contexts, conjugation is BAB^{-1}. We will always write the inverse on the left.)

To define conjugation (and have it mean anything interesting), we require the non-commutativity of matrix multiplication. If matrix multiplication were commutative, we could calculate $B^{-1}AB = B^{-1}BA = A$, and conjugation would have no effect. Conjugation works because of the non-commutativity of matrix multiplication.

Conjugation changes the matrix A, but tends to keep some similarity to the original matrix. It also divides matrix groups into classes which share certain similarities or properties.

Definition 19.1.4. Let G be a group of linear transformations. Let $A \in G$. The *conjugacy class* of A is the set of all matrices which are conjugates of A by any other $B \in G$.

19.2 Properties of Conjugation

Proposition 19.2.1. *Let A and B be $n \times n$ matrices with B invertible. Conjugation preserves determinants:* $\det A = \det B^{-1}AB$.

Proposition 19.2.2. *Let A and B be invertible $n \times n$ matrices in a finite matrix group G. Then the order of A is the same as the order of $B^{-1}AB$.*

Proposition 19.2.3. *Let A and B be $n \times n$ matrices with B invertible. The number of fixed points of A is the same as the number of fixed points of $B^{-1}AB$.*

Lecture 20

Eigenvectors, Eigenvalues and Spectra

20.1 Definitions

Definition 20.1.1. Let A be an $n \times n$ matrix. A non-zero vector $v \in \mathbb{R}^n$ is an *eigenvector* for A with *eigenvalue* λ if $Av = \lambda v$. The set of all eigenvalues for the matrix A is called the *spectrum* of A.

Eigenvectors for a matrix are vectors which do not change direction. They may be dialated by a factor λ (including a flip if λ is negative), but they still point in the same direction. Projections are also included, since we allow $\lambda = 0$; such an eigenvector would represent a direction sent entirely to zero under the transformation. Note that if v is an eigenvector, then so is αv for any $\alpha \neq 0$, $\alpha \in \mathbb{R}$.

At first glance, this definintion may seem insignificant; interesting, perhaps, but not necessarily central. It turns out that eigenvectors and eigenvalues are one of the most useful and important definitions in linear algebra. Many problems in applied mathematics depend upon finding the eigenvalues of a particular matrix.

Before we learn how to calculate eigenvectors and eigenvalues, we can look at some obvious examples.

Example 20.1.2. The identity is the easiest: for any vector v we have $\mathrm{Id}v = v$. Therefore, all vectors are eigenvectors of the identity with eigenvalue $\lambda = 1$.

Example 20.1.3. The zero matrix is similarly easy: it sends all vectors to the origin. All vectors are eigenvectors with eigenvalue $\lambda = 0$.

Example 20.1.4. Consider a dialation matrix in $\mathbb{R}^3$:

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

Here, all three of the standard basis vectors are eigenvectors. e_1 has eigenvalue a, e_2 has eigenvalue b and e_3 has eigenvalue c. If a, b and c are all distinct numbers, there are no other eigenvectors (up to scale).

Example 20.1.5. Consider a rotation in $\mathbb{R}^2$. Assuming the rotation is not trivial ($\theta \neq 0$), then there are no preserved directions. This rotation has no eigenvectors. However, if we had a rotation in $\mathbb{R}^3$, we have to choose an axis. The axis direction is fixed, so a vector in that direction would be an eigenvector with eigenvalue one.

Example 20.1.6. Consider a reflection in $\mathbb{R}^2$. A vector along the line of reflection is unchanged, so it would be an eigenvector with eigenvalue one. A vector perpendicular to the line of reflection is fliped exactly, so it would be an eigenvector with eigenvalue -1. In $\mathbb{R}^3$, we reflect over a plane through the origin. Any vector in the plane of reflection is unchanged, so that vector is an eigenvectors with eigenvalue one. Any vector perpendicular to the plane of reflection is precisely fliped, so that vector is an eigenvector with eigenvalue -1.

Example 20.1.7. Finally, consider projections in $\mathbb{R}^2$. Projection onto the x-axis preserves the x axis, so any vector $\begin{pmatrix} a \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue one. However, it sends the y-axis direction to the origin, so any vector $\begin{pmatrix} 0 \\ b \end{pmatrix}$ is an eigenvector with eigenvalue zero.

20.2 Calculation of Eigenvalue and Eigenvectors

We would like to have an algorithm for finding eigenvectors and eigenvalues. Remember the definition: v is an eigenvector of A if $Av = \lambda v$ for some real number λ. We can write the scalar multiplication λv as $\lambda \mathrm{Id} v$, since the matrix $\lambda \mathrm{Id}$ multiplies each entry by λ. Therefore, the equation becomes $Av = \lambda \mathrm{Id} v$ or $Av - \lambda \mathrm{Id} v = 0$, where the right hand side is the zero vector. Since matrix action is linear, this is the same as $(A - \lambda \mathrm{Id})v = 0$.

We assumed that v was non-zero in the definition of eigenvectors. Therefore, the matrix $(A - \lambda \mathrm{Id})$ sends a non-zero vector v to the zero vector. This implies $(A - \lambda \mathrm{Id})$ has a non-trivial kernel, hence

must have zero determinant. Thererfore, a necessary condition for the existence of a eigenvector with eigenvalue λ is that $\det(A - \lambda \mathrm{Id}) = 0$. We start our algorithm by computing this determinant.

This determinant has terms which involve the entries of A and λ. Since A is an $n \times n$ matrix, each term can be the product of n entries. That means that the determinant will be a degree n polynomial in the variable λ.

Definition 20.2.1. Let A be an $n \times n$ matrix The polynomial $\det(A - \lambda \mathrm{Id})$ in the variable λ is called the *characteristic polynomial* of the matrix.

We need an important definition about the roots of polynomials. Recall that α is a root of a polynomial $p(\lambda)$ if and only if $(\lambda - \alpha)$ is a factor of the polynomial. However, it may be true that $(\lambda - \alpha)^m$ is a factor for some integer $m > 1$.

Definition 20.2.2. Let $p(\lambda)$ be a polynomial in the variable λ and α a root. Let m be the largest positive integer such that $(\lambda - \alpha)^m$ is a factor of $p(\lambda)$. Then m is called the *multiplicity* of the root α.

A degree n polynomial has at most n real roots, so we can find at most n values for λ where the determinant vanishes. These will be the possible eigenvalues. However, the polynomial also have at most n factors, so the sum of all the multiplicities of the roots is also at most n. That implies that if we have roots with higher multiplicities, we will certainly have fewer than n roots.

Now we return to the eigenvalue/eigenvector algorithm. When we have an eigenvalue λ, we then want to find the matching eigenvectors. These vectors will be in the kernel of $(A - \lambda \mathrm{Id})$, so we just have to calcluate this kernel.

The second step will always find eigenvectors: the determinant of the matrix is zero, so it must have a non-trivial kernel. However, we don't know now many we will find (we don't know the dimension of the kernel). In the first step, we are looking for roots of a polynomials of degree n. It may have as many as n distinct real roots, but it may have far fewer or even none. Recall the identity and the zero matrix, where all vectors in $\mathbb{R}^n$ are eigenvectors for the appropriate eigenvalues, and also the rotations in $\mathbb{R}^2$ where there were no eigenvectors for any eigenvalue. Also note that solving for roots of polynomials, particular in high degrees, is a very hard problem.

We have a definition that helps us keep track of this information.

Definition 20.2.3. Let A be a $n \times n$ matrix and let λ be an eigenvalue of A. Then the *eigenspace* of λ, written $E(\lambda)$, is the set of all eigenvectors for the eigenvalue λ. Equivalently, it is the kernel of the matrix $A - \lambda \mathrm{Id}$. The dimension of the eigenspace associated to λ is bounded by the multiplicity of λ as a root of the characteristic polynomial.

Example 20.2.4.

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \implies A - \lambda \mathrm{Id} = \begin{pmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{pmatrix}$$

The determinant of this is $9 - 6\lambda + \lambda^2 - 1 = 8 - 6\lambda + \lambda^2 = (\lambda - 4)(\lambda - 2)$. So the eigenvalues are $\lambda = 4$ and $\lambda = 2$. We calculate the kernel of $A - \lambda\mathrm{Id}$ first for $\lambda = 2$.

$$A - 2\mathrm{Id} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

The kernel of this matrix is $\mathrm{Span}\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$. All multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ are eigenvectors for $\lambda = 2$. Next we move on to $\lambda = 4$.

$$A - 4\mathrm{Id} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

The kernel of this matrix is $\mathrm{Span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$. All multiples of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are eigenvectors for $\lambda = 4$.

Example 20.2.5.

$$A = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \implies A - \lambda\mathrm{Id} = \begin{pmatrix} -\lambda & -2 \\ 2 & -\lambda \end{pmatrix}$$

The characteristic polynomial is $\lambda^2 + 4 = 0$, which has no roots. Therefore, there are no eigenvalues.

Example 20.2.6. Consider a diagonal matrix.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \implies A - \lambda\mathrm{Id} = \begin{pmatrix} 1-\lambda & 0 & 0 & \\ 0 & -3-\lambda & 0 & 0 \\ 0 & 0 & 4-\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{pmatrix}$$

The characteristic polynomial is $(1 - \lambda)(-3 - \lambda)(4 - \lambda)(-\lambda)$, which clearly has solutions $\lambda = 1$, -3, 4 and 0. The eigenvectors are just the axis vectors e_1, e_2, e_3 and e_4, respectively. As we said before, diagonal matrices have axis vectors as eigenvectors and diagonal entries as eigenvalues.

20.3 Symmetric Matrices

We defined the transpose of a matrix in the chapter on orthogonal matrices. The transpose leads to a new definition in this chapter.

Definition 20.3.1. A $n \times n$ matrix A is *symmetric* if $A = A^T$.

Example 20.3.2. Here are two symmetric matrices.

$$\begin{pmatrix} 5 & 1 & -3 \\ 1 & 2 & 0 \\ -3 & 0 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 4 & -2 & -9 \\ 4 & -4 & 3 & 3 \\ -2 & 3 & 8 & -1 \\ -9 & 3 & -1 & 7 \end{pmatrix}$$

We define symmetric matrices in this section because they behave very well with respect to eigenvalue and eigenvectors.

Proposition 20.3.3. *A symmetric $n \times n$ matrix always has n real eigenvalues, counted with multiplicity. Moreover, the eigenspaces all have the maximum dimension, which is the multiplicity of the eigenvalue.*

Lecture 21

Diagonalization

Definition 21.0.1. An $n \times n$ matrix is called *diagonalizable* if if has n eigenvalues, counted with multiplicity, and if its eigenspaces all have the maximum dimension (their dimension equals their multiplicity).

All symmetric matrices are diagonalizable by the proposition at the end of the previous section. The reason for the term 'diagonalizable' is the following proposition.

Proposition 21.0.2. *Let A be a diagonalizable $n \times n$ matrix. Then there exists a diagonal $n \times n$ matrix D and an intertible $n \times n$ matrix S such that $D = S^{-1}AS$. (Equivalently, A is* conjugate *to a diagonal matrix). Moreover, the entries on the diagonal of D are precisely the eigenvalues of A. S can then be construsted such that the ith column of S is a eigenvector for the ith diagonal entry of D.*

The proposition already provides an algorithm. We find the eigenvalues and eigenvectors of A, construct D with the eigenvalues on the diagonal, and construct S with the eigenvectors as columns, matching in place with D. Then we just need to invert S to have the complete formula. Often we write the diagonalization as $A = SDS^{-1}$.

Our symmetric matrices are even more pleasant than we noted before.

Proposition 21.0.3. *A matrix is diagonalizable with S an orthogonal matrix if and only if the matrix is symmetric. Equivalently, a matrix is symmetric if and only if there is an orthonormal basis of $\mathbb{R}^n$ consisting entirely of eigenvectors. To construct this S we simply take unit eigenvectors for its columns.*

Example 21.0.4.

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

The characteristic polynomial is $\lambda^2 - 2\lambda - 8$. The roots are $\lambda = 4$ and $\lambda = -2$. We start with $\lambda = 4$.

$$A - 3\mathrm{Id} = \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix}$$

The kernel is the span of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The unit vector is $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Next we take $\lambda = -2$.

$$A + 2\mathrm{Id} = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$$

The kernel is the span of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. The unit vector is $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then we can construct the diagonal matrix D and the invertible matrix S.

$$D = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix} \qquad\qquad S = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

Since S is orthogonal, its inverse is its transpose.

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

This completes the diagonalization.

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = SDS^{-1}$$

Lecture 22

Linear Dynamical Systems

22.1 General Theory

Let's consider an applied mathematics problem where a vector $v \in \mathbb{R}^n$ somehow represents the state of the situation under consideration. This state could be a position, keeping with the geometry of $\mathbb{R}^n$, but typically it is just a list of numbers with some other interpretation. Once we have states, which are vectors in $\mathbb{R}^n$, assume we can move from one state to the next over a period of time called a timestep. If we are in state v_k, then after the next second/minute/day/year, we are in the next state v_{k+1}. Thus we have a sequence $\{v_k\}_{k=1}^{\infty}$ of states in $\mathbb{R}^n$.

How do we move from one state to the next? If the process is linear, then we can describe the progression as a matrix multiplication. That it, there exists a matrix A such that $v_k = Av_{k+1}$ for all $k \in \mathbb{N}$. Notice that this is only a single matrix: the matrix A goes from one state to the next, no matter where we are in the sequence. In this way, these processes have no memory: each state only depends on the previous state, not on any of the states before the previous.

Definition 22.1.1. A sequence of vectors $\{v_k\}_{k=1}^{\infty}$ which satisfy the matrix equation $v_{k+1} = Av_k$ is called a *(discrete) linear dynamical system.*

As a first observation, we need to calculate A^n for large values of n. In general, this is computationally difficult. However, if A is diagonalizable, then $A = SDS^{-1}$ and $A^n = SD^nS^{-1}$. Higher powers of a diagonal matrix are computationally easy, so diagonalizaing the matrix is often a necessary and natural step.

We can also use the eigenvalue/eigenvector information in our interpretation. A natural question is: what is the long term behaviour of the sequence? For dynamical systems, we can asnwer this question with eigenvectors. If v is an eigenvector of A with eigenvalue λ, then $A^nv = \lambda^n v$, so the long term behaviour of this state is expressed as multiples of the original state. This behaviour can be broken down into cases depending on the eigenvalue λ.

- If $\lambda = 0$, then this is a collapsing state and all future states are simply the zero vector.
- If $|\lambda| < 1$, then we have exponential decay. The long term behaviour of $A^n v$ is exponential decay of the original vector.
- If $\lambda = -1$, then we have a 2-period oscilating state. The sequence begins with the state v and jumps back and forth between v and $-v$.
- If $\lambda = 1$, then v is a steady state: the sequence never changes. These steady states are often very important in modelling.
- If $|\lambda| > 1$, then we have exponential growth of the original vector. If $\lambda < -1$, this growth comes with $\pm$ oscilations as well.

We could ask what happens for complex eigenvalues, since those will naturally also occur. Since the characteristic polynomial will have real coefficients, these complex eigenvalues will come in conjugate pairs. As we might learn in other courses, such pairs and the exponential behaviour (λ^n) give rise to sinusoidal behaviour after clever linear combinations. The reasons will have to wait for another course.

In many models, the coefficients of the matrix will be probabilities, transition terms, growth rates or other *positive* real numbers. A will very often be a matrix with all non-negative entries. There is a powerful theorem that tells us what to expect for the eigenvalues/eigenvectors of such a matirx: the Perron-Frobenius theorem. There are some technical details in the assumptions for the theorem; we'll state a weak version first.

Theorem 22.1.2. *Let A be a $n \times n$ matrix with non-negative coefficients. Then there is a largest non-negative eigenvalue λ_1 with an eigenvector which has all non-negative entries. All other eigenvalues λ satisfy $|\lambda| \leq |\lambda_1|$.*

The stronger version needs an new definition.

Definition 22.1.3. Let A be an $n \times n$ matrix. Then A is called *irreducible* if for all i and j the exists a positive integer m such that the ijth entry of A^m is non-zero.

This definition roughly captures the idea that all the coefficients in the states are somehow related. Using the defitinion, here is a stronger version of the Perron-Frobenius theorem.

Theorem 22.1.4. *Let A be a $n \times n$ irreducible matrix with non-negative coefficients. There there is a* unique *largest* positive *eigenvalue λ_1 with a 1-dimensional eigenspace and an eignevector which has all* positive *entires. All other eigenvalues λ satisfy $|\lambda| < |\lambda_1|$. Moreover, if r_i is the sum of the entries of A in the ith row, then λ is bounded above and below by the largest and the smallest of the $|r_i|$.*

22.2 Linear Recurrence Relations

The first use of iterated processes is understanding linear recurrence relations.

Definition 22.2.1. A *linear recurrence relation* is a sequence of numbers $\{a_n\}$ for $n \in \mathbb{N}$ such that there is a linear equation for a_i in terms of the previous k terms. The number k is called the *order* of the recurrence relation. The first k terms must be explicitly given.

If we write $a_n = c_1 a_{n-1} + c_2 a_{n-2} + \ldots + c_k a_{n-k}$, then we can re-interpret this as a matrix and a dynamical system. We do this by writing a k-vector with entries $a_{n+k}, \ldots, a_{n+1}$. Then the recurrence relation is given by a matrix equation.

$$\begin{pmatrix} a_{n+k} \\ a_{n+k-1} \\ \vdots \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & \ldots c_{k-1} & & c_k \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \ldots 1 & 0 & \end{pmatrix} \begin{pmatrix} a_{n+k-1} \\ a_{n+k-2} \\ \vdots \\ a_n \end{pmatrix}$$

As long as the c_i are all non-zero, this is an irreducible matrix.

A recurrence relation is a recursive way of creating a sequence, but it is difficult to calculate with. If we wanted the 400th term, we would have to calculate all the intermediate terms as well. That leads to a natural question: is there a closed-form equation that calculates the kth term or a recurrence relation directly? In some circumstances, eigenvalues and diagonalization can help solve this problem.

Let C be the matrix of coefficients and let $u = \begin{pmatrix} a_k \\ a_{k-1} \\ \vdots \\ a_1 \end{pmatrix}$. Assume that C is diagonalizable and that $\{v_1 \ldots, v_k\}$ are a complete set of eigenvalues. These k vectors will span $\mathbb{R}^k$, so we can take the vector v, formed above of k consecutive terms of the recurrence relation, and express it in terms of the eigenvectors.

$$v = b_1 v_1 + b_2 v_2 + \ldots + b_k v_k$$

The matrix C acts on v by iterating the terms of the series. The term a_{n+k} of the series is the first term of the vector given by the matrix C acting n times on v: $C^n v$. So the matrix action $C^n v$ entirely determines the behaviour of the series. Then we can use the eigenvalue decomposition.

$$C^n v = C^n(b_1 v_1 + b_2 v_2 + v + b_k v_k) = b_1 \lambda_1^n v_1 + b_2 \lambda_2^n v_2 + \ldots + b_k \lambda_k^n)$$

Since all we are doing is taking higher powers of the eigenvalue, this is a reasonably easy calculation. Isolating the first coefficient of the vector will give the desired closed form expression.

22.3 Leslie Matrices

Let's conisder a population broken down into finitely many age categories. Leslie matrices are a dynamical system which model the age structure of this population. To demonstrate, we'll assume there

are four age categories, so the state vector v_k is in $\mathbb{R}^4$ and the matrix C is a 4×4 matrix. If we use coordinates w, x, y, z in $\mathbb{R}^4$, we can write the matrix equation for this dynamical system.

$$\begin{pmatrix} w_{k+1} \\ x_{k+1} \\ y_{k+1} \\ z_{k+1} \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & f_3 & f_4 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \end{pmatrix} \begin{pmatrix} w_k \\ x_k \\ y_k \\ z_k \end{pmatrix}$$

Nine of the entries are set to zero intentionally, but we can interpret the remaining entries. The first row outputs to the first category, representing the creation of new members of the population. Each f_i is the *fecundity* of each population state: its rate of producing offspring. Of course, some of these might be zero, if only certain age categories produce offspring. The s_i, on the other hand, are transitions from one age category to the next. These are *survival rates*: s_1 is the rate of survival from cateogry 1 to category 2, s_2 from category 2 to 3 and s_3 from category 3 to 4. The remaining elemnets of the second, third and fourth rows are zero since they represent unnatural jumps in age categories: i.e., no one goes from the third back down to the second age category.

These Leslie matrices are irreducible (recall that ireeducible meant that each state connects to the other; by survival and fertility, it is possible to pass from any age category to another over several years). This means that there is a unique largest eigenvalue λ with a positive eigenvector. This is the value we care about: in the long run, the largest eigenvalue dominates and its eigenvector gives the stable age distribution. If $\lambda = 1$, we expect a stable population. If $\lambda > 1$, we expect exponential growth, and if $\lambda < 1$ we expect exponential decay.

22.4 Stochastic Matrices and Markov Chains

Markov chains are probabalistic linear dynamical systems. Like all dynamical systems, we have a state vector $v \in \mathbb{R}^n$. Here, v represents a probablistic state and the iterated process asks: what are the probabilities in the next state? Often, Markov chains are given graph theoretic interpretation. If $v \in \mathbb{R}^n$, then we can consider a graph on n vertices. Each coefficient of the matrix A_{ij} is the probability of going from vertex j to vertex i. In terms of the graph, these transition number can be thought of as a label on a directed edge.

In this interpretation, the ith column gives the total probabilities of all departures from the ith vertex (including A_{ii}, the probability of staying put). Since something must happen in each time step, we insist that the column sums, $r_j = \sum_i a_{ij}$, all must equal one.

Definition 22.4.1. A non-negative $n \times n$ matrix where all the columns sum to one is called a *(left) stochastic matrix*. (There is a similar definition of a righ stochastic matrix where all the rows sum to one; it is, in fact, the more common definition. However, it doesn't mesh that well with our conventions of columns vectors and matrix actions, so we will work with left stochastic matrices and surpress the word 'left' in what follows.)

Stochastic matrices reflect a probabalistic interpretation. They are are non-negative, so the weak form of Perron-Frobenius applies, but they may not be irreducible. (In this graph, this depends on whether or not each vertex is reachible from all the other vertices).

As with previous iterated processes, we are curious about the long term behaviour. However, for stochastic matrices are are less directly concerned with eigenvalues and eigenvectors. (We can prove, easily from the column sums, that the dominant eigenvalue is $\lambda = 1$ and the dominant eigenvector is the vector where all coefficients sum to 1.) Instead, we are more interested in the matrix A^n itself. The higher powers of the matrix calculate the iterated process. Therefore, we can interpret the coefficient $(A^k)_{ij}$ as the probability of starting in vextex j and ending up in vertex i after k timesteps.