Math 67: Modern Linear Algebra (UC Davis, Winter 2020) — Summary of lectures

Dan Romik

[Version of March 14, 2020— this document will be periodically updated with new material]

Lecture 1 (1/6/20)

- High-level description of course goals: 1. linear algebra theory; 2. linear algebra computational skills; 3. introduction to abstract math.

- **Today's topic: introduction to linear algebra.** Conceptually, linear algebra is about sets of quantities (a.k.a. vectors) that are associated with each other by a "linear" relationship, and how to manipulate them, classify the nature of such relationships, and solve equations to determine one set of quantities given another.

  Practically speaking, a lot of the calculations will reduce to solving a system of linear equations. An equation is called "linear" if all the terms on both sides of the equation have the form "a number times one of the unknowns" or "a number". E.g., if x and y are the unknowns, you cannot have the term $10xy$ or $-5x^2$ (let alone more complicated things like $\sqrt{x}$, e^z, $\sin(x + y)$ etc.).

- **What is linear algebra good for?** Almost everything in math, science, engineering. A few examples:

  - 3D graphics
  - Filters (Instagram, Photoshop etc). Also for music and sound processing and to filter other sources of data (such as in astronomy, medicine, nuclear physics, . . .).
  - Antialiasing of text on a phone or computer screen to make it look nice.
  - Analyzing card shuffling, genetic drift and other random processes with many states (Markov chains)
  - Neural networks, the mathematics of Machine Learning, AI
  - Multivariate calculus, optimizing functions of many variables
  - Search engine ranking algorithms
  - Stability analysis in control theory (think robots, rockets, airplanes)
  - Understanding fun things in math and physics such as the moving sofa problem and the Dzhanibekov effect (aka tennis racket theorem)
  - . . . and many more applications.

- **Example 1.** A baker needs one egg and 3 ounces of flour to make a muffin, and 2 eggs and 2 ounces of flour to make a croissant. Given 20 eggs and 28 ounces of flour, how many muffins and croissants can the baker make, assuming all eggs and flour are used?

Solution. Let x denote the number of muffins, and y the number of croissants. The question translates to the equations

$$\begin{cases} x + 2y = 20 \\ 3x + 2y = 28 \end{cases}$$

We can solve this using the "substitution" method:

$$x + 2y = 20 \implies x = 20 - 2y \implies 3(20-2y) + 2y = 28$$
$$\implies 4y = 60 - 28 = 32 \implies y = 8 \implies x = 4$$

Or we can solve by "eliminating variables", i.e., subtracting a multiple of one equation from the other to get an equation with a single variable. For example, subtracting the first equation from the second gives

$$2x = 8 \implies x = 4,$$

which then easily leads to $y = 8$ after substituting the value for x in either of the equations.

- **Abstract algebraic formulation.** Represent the system of equations given above symbolically in the form

$$A \cdot \mathbf{v} = \mathbf{u},$$

where $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\mathbf{u} = \begin{pmatrix} 20 \\ 28 \end{pmatrix}$ and "$\cdot$" refers to "multiplication of a matrix by a vector", a weird form of multiplication defined by

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

To solve this, we simply need to "divide both sides (from the left) by A". What does it mean to "divide" by a matrix? We will learn that there is a special matrix, denoted A^{-1} and called the "inverse" matrix of A, that has the property that

$$A^{-1} \cdot (A \cdot v) = (A^{-1} \cdot A) \cdot v = \text{"1"} \cdot v = v,$$

(here, we need to extend the concept of multiplying a matrix by a vector to multiplying a matrix by a matrix; we also need to show "associativity", i.e. that the order in which we multiply doesn't change the result). So, the original equation translates to

$$\mathbf{v} = A^{-1} \cdot \mathbf{u}.$$

If we knew A^{-1}, we would be able to compute $\mathbf{v}$ by carrying out the multiplication. In this case,

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

(we will learn in the future how to compute such inverse matrices), so we get that

$$\mathbf{u} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & -\frac{1}{4} \end{pmatrix} \cdot \begin{pmatrix} 20 \\ 28 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \cdot 20 + \frac{1}{2} \cdot 28 \\ \frac{3}{4} \cdot 20 - \frac{1}{4} \cdot 28 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix},$$

as before.

- **Geometric formulation.** Each of the two equations represents the equation for a line in the plane. Finding the solution corresponds to finding the point in the plane which is at the intersection of the two lines. In the example above, we had a unique solution, but this geometric way of looking at things tells us that it's also possible to have no solutions (in the case when the two lines are parallel but non-intersecting), or to have infinitely many solutions (if the two lines are identical). In general, "solving" a linear system of equations means "giving a simple description of the set of solutions" (which may be empty, or contain a single point, or be infinite).

- **Linear functions and linear transformations.** Yet another way of thinking about the equation, that we will explore in more detail later on in the course, is that the "operation" that takes a two-dimensional vector $\mathbf{v}$ and returns the new vector $A \cdot \mathbf{v}$ is a special kind of operation which we call a *linear transformation* – it acts on the two-dimensional plane in an interesting way, e.g., by stretching it, rotating it, reflecting it (but not *bending* it – that is why it is called "linear"). Solving the equation corresponds to finding the unknown point $\mathbf{v}$ in the plane that is transformed to the given (known) point $\mathbf{u}$ under the transformation. If we find good ways of visualizing linear transformations or understanding what are the different ways in which they can act, that will help us in solving linear systems.

- **Example 2. The Google PageRank algorithm.** In a simplified model of the internet, there are only 2 websites, (say) Facebook and Twitter. Facebook has 5 links pointing to itself and 2 links to Twitter. Twitter has 2 links to Facebook and 8 to itself. Which site is more important (and should therefore be placed higher up in a search result presented by Google)?

  Solution. The Google PageRank web page ranking algorithm (named after its inventor, Larry Page!) assigns importance to web pages according to the self-referential rule:

  > A web page's importance is proportional to the weighted average of the number of links pointing to it from all web pages, with each page being weighted according to its own importance.

  (This rule can also be applied to other things, for example to the ranking of social status of people: i.e., you are popular in your social group in proportion to how many friends you have, but only to the extent that your friends are themselves popular...).

  The fact that the rule is self-referential, defining "importance" in terms of that involve the same term we are trying to define, is why we are led to having to solve a system of equations (as opposed to just having a plain formula that directly cranks out the importance in terms of the data that's given to us).

  In the case of our model internet, let x denote the relative importance of Facebook and y denote the relative importance of Twitter, then the rule above leads to the equations:

$$\begin{cases} x + y = 1 \\ 5x + 2y = zx \\ 2x + 8y = zy \end{cases}$$

  where z is an additional unknown which represents an arbitrary positive proportion constant. The first equation is what I mean by relative importance — the numbers x, y add up to a "total importance" of 1.

It's important to note that if z is taken into account, then this is a **non-linear** system of equations. However, due to their special structure such equations are still considered a part of linear algebra. The reason is that if we are magically given the value of z by someone with advanced knowledge (in this case $z = 9$), the system becomes a linear system and can be solved with the usual methods:

$$\left\{\begin{array}{rcrcr} x & + & y & = & 1 \\ 5x & + & 2y & = & 9x \\ 2x & + & 8y & = & 9y \end{array}\right. \implies \left\{\begin{array}{rcrcr} x & + & y & = & 1 \\ -4x & + & 2y & = & 0 \\ 2x & - & y & = & 0 \end{array}\right. \implies \left\{\begin{array}{rcrcr} x & + & y & = & 1 \\ 3x & & & = & 1 \end{array}\right.$$

so we finally get $x = 1/3, y = 2/3$.

We will learn later in the course how to deal with finding the value of z in such a situation. This will lead to certain *polynomial* equations (a higher-order generalization of the linear and quadratic equations we are familiar with).

Lectures 2 (1/8/20) and 3 (1/10/20)

- **Coefficients.** In linear algebra we study systems of linear equations. The numbers appearing in these equations are known as *coefficients.*

- **Number systems and number fields.** In elementary applications the coefficients are ordinary real numbers (in fact, usually they are *rational numbers*). However, it turns out that the algebraic rules for manipulating numbers (by adding, subtracting, multiplying them, etc.) that we rely on for linear algebra are not unique to real numbers. There are other number systems that satisfy the "good" properties that are necessary in linear algebra. It makes sense to abstract away those properties for a "good number system". In higher mathematics such a system of numbers is called a *field.* In this course we will focus mainly on the real numbers and complex numbers, but it's interesting to keep in mind that other number fields also arise in real-life applications (notably in computer science, where an important field is the "field with 2 elements" consisting of the numbers 0 and 1), and that essentially all the linear algebra theory we'll develop carries over to that more general setting as well, with little or no modification.

- **Complex numbers.** One of the nicest fields there are (we still haven't defined what a field is, but don't worry about that for now) is that of the *complex numbers.* They are numbers of the form
$$z = a + bi,$$
where i is a hypothetical construct, a symbol representing one of the two square roots of -1. The other square root of -1 is $-i$, since if $i^2 = 1$ then
$$(-i)^2 = (-1) \cdot i \cdot (-1) \cdot i = (-1)^2 i^2 = 1 \cdot (-1) = (-1).$$

- Let's be a bit more formal:

  Definition. The set of complex numbers is the set, denoted $\mathbb{C}$, of pairs (a, b) of real numbers, with the convention that instead of writing (a, b) in the usual vector notation from calculus, we write it in the form $a + bi$. Formally, we can express this definition as
$$\mathbb{C} = \{a + bi \ : \ a, b \in \mathbb{R}\}.$$
  If $z = a + bi$ is a complex number, $a, b \in \mathbb{R}$, we call $a \in \mathbb{R}$ the *real part of* z, and we call b the *imaginary part of* z. We denote $a = \operatorname{Re} z$ and $b = \operatorname{Im} z$. We can think of the complex number $a + bi$ geometrically as a vector with two coordinates a (the x-coordinate) and b (the y-coordinate).

- **Algebraic operations on complex numbers.** What makes $\mathbb{C}$ a field? As a set of numbers, it is not much different than the two-dimensional plane $\mathbb{R}^2$, except that the vector (a, b) is written as $a + bi$. The key to looking at $\mathbb{C}$ as a field is to consider also its *algebraic* properties. It turns out that there is a way to define algebraic operations of addition, subtraction, multiplication and division on complex numbers, and these operations satisfy the "good" properties that we require in a field.

1. **Addition of complex numbers.** If $z = a + bi$ and $w = c + di$ are two complex numbers, we define their sum by

$$z + w = (a + c) + (b + d)i$$

2. **The negative of a complex number.** If $z = a+bi$ is a complex number, its negative is

$$-z = -a - bi$$

3. **Subtraction of complex numbers.** The difference of two complex numbers $z = a+bi$ and $w = c + di$ is defined as

$$z - w = z + (-w)$$

4. **Multiplication of complex numbers.** The product of two complex numbers $z+a+bi$ and $w = c + di$ is defined by

$$z \cdot w = (a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

5. **The reciprocal of a complex number.** The reciprocal of a complex number $z = a + bi$ is defined, assuming $z \neq 0$, by

$$z^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i$$

6. **Division of complex numbers.** The quotient of two complex numbers $z = a + bi$ and $w = c + di$ is defined if $w \neq 0$ by

$$\frac{z}{w} = z \cdot w^{-1}$$

- **Examples.** (fill in the answers)

$$(3 + 5i) + (10 - 2i) =$$
$$(3 + 5i) - (10 - 2i) =$$
$$(3 + 5i) \cdot 4 =$$
$$4 \cdot (3 + 5i) =$$
$$(3 + 5i) \cdot (1 + i) =$$
$$(1 + i)^2 = (1 + i) \cdot (1 + i) =$$
$$(1 + i)^4 =$$
$$(1 + i)^{-1} =$$
$$(2 - i)^{-1} =$$
$$\frac{2 - i}{1 + i} =$$

- **Properties of addition. Theorem:** For any complex numbers z, z_1, z_2, z_3, the following properties are satisfied:

1. ("Commutativity") $z_1 + z_2 = z_2 + z_1$

2. ("Associativity") $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$
   So, in fact, the parentheses are not needed and we will usually omit them in the future, writing this expression simply as $z_1 + z_2 + z_3$.
3. ("Neutral element") $z + 0 = 0 + z = z$
4. ("Additive inverse") $z + (-z) = 0$

Proof: ...

- **Properties of multiplication. Theorem:** For any complex numbers z, z_1, z_2, z_3, the following properties are satisfied:

  1. ("Commutativity") $z_1 \cdot z_2 = z_2 \cdot z_1$
  2. ("Associativity") $(z_1 \cdot z_2) \cdot z_3 = z_1 \cdot (z_2 \cdot z_3)$
  3. ("Neutral element") $z \cdot 1 = 1 \cdot z = z$
  4. ("Multiplicative inverse") If $z \neq 0$ then z^{-1} is defined and $z \cdot z^{-1} = 1$
  5. ("Distributivity") $z_1 \cdot (z_2 + z_3) = z_1 \cdot z_2 + z_1 \cdot z_3$

  Proof: ...

- **The conjugate of a complex number.** If $z = a + bi$ is a complex number, we define its conjugate to be the number $a - bi$, denoted by $\overline{z}$:

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

  Geometrically, taking the conjugate corresponds to reflecting the vector associated with z across the x-axis.

  Why is the conjugate element interesting? Let's see what happens when we multiply z and $\overline{z}$:

$$z \cdot \overline{z} = (a + bi)(a - bi) = (aa - b(-b)) + (a(-b) + ba)i = (a^2 + b^2) + 0 \cdot i = a^2 + b^2$$

  The product is always an ordinary real number, equal to $a^2 + b^2$. In particular, we see that if we take the conjugate and divide it by this real number $a^2 + b^2$ (which is allowed if $a^2 + b^2 \neq 0$, i.e., if $z \neq 0$), we get a number w with the property that $w \cdot z = 1$. This is simply the reciprocal of z. So, we have rederived the formula for the reciprocal element:

$$z^{-1} = \frac{1}{a^2 + b^2} \cdot \overline{z} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2} i$$

- **Properties of conjugation. Theorem:** For any complex numbers $z, z_1, z_2 \in \mathbb{C}$, the following properties hold:

  1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$
  2. $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$
  3. $\overline{z^{-1}} = (\overline{z})^{-1}$ if $z \neq 0$
  4. $z = \overline{z}$ if and only if $\text{Im}(z) = 0$ (i.e., if z is a real number)

5. $\overline{\overline{z}} = z$
6. The real and imaginary part of z can be written as

$$\text{Re}(z) = \tfrac{1}{2}(z + \overline{z}), \qquad \text{Im}(z) = \tfrac{1}{2i}(z - \overline{z})$$

- **The modulus of a complex number.** The quantity $a^2 + b^2$ has a simple geometric meaning as well. By the Pythagorean Theorem, the distance from 0 to $z = a + bi$ is equal to $\sqrt{a^2 + b^2}$. We call this the **modulus** of the number z, and denote it by $|z|$:

$$|z| = \sqrt{a^2 + b^2}$$

(if z is a real number, it is simply the usual absolute value of z). This quantity is also known as the **norm**, **magnitude**, or **length** of z (especially in the general context of vectors, not complex numbers). To summarize the above discussion, we have the identities:

$$z \cdot \overline{z} = |z|^2,$$
$$z^{-1} = \frac{\overline{z}}{|z|^2}$$

- **Properties of the modulus. Theorem:** for any complex numbers $z, z_1, z_2 \in \mathbb{C}$, we have the properties:

1. $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$
2. $|z_1/z_2| = |z_1|/|z_2$ if $z_2 \neq 0$
3. $|\overline{z}| = |z|$
4. $|\text{Re}(z)| \leq |z|$ and $|\text{Im}(z)| \leq |z|$
5. ("triangle inequality") $|z_1 + z_2| \leq |z_1| + |z_2|$
6. (triangle inequality reformulated) $|z_1 - z_2| \geq \big||z_1| - |z_2|\big|$

Proof: ...

Lecture 4 (1/13/20)

- **Polar representation of complex numbers.** The usual representation $z = x + yi$ of complex numbers tells us where the vector z lies in the plane in terms of the usual Cartesian coordinates. An alternative representation is in terms of the *polar coordinates*, where we give the length r of the vector and its angle θ (measured in the anti-clockwise direction) relative to the x-axis. The numbers (r, θ) are called the polar coordinates of z. They satisfy $r \geq 0$ and $0 \leq \theta < 2\pi$, and θ is only defined if $z \neq 0$.

  From elementary trigonometry, it is easy to see that r, θ are related to x, y by

  $$x = r\cos\theta$$
  $$y = r\sin\theta$$
  $$r = |x + yi| = \sqrt{x^2 + y^2}$$
  $$\theta = \text{“the angle function” of } (x, y)$$

  (there is no standard notation for this; sometimes “arg z” is used, and the angle may be referred to as the “argument” of z).

  To summarize, the complex representation of z is usually written in the form

  $$z = r(\cos\theta + i\sin\theta)$$

- **Multiplication in polar coordinates.** Let's see what happens when we try to multiply two complex numbers z, w that are given in polar coordinates:

  $$\begin{aligned} z &= r_1(\cos\theta_1 + i\sin\theta_1), \\ w &= r_2(\cos\theta_2 + i\sin\theta_2), \\ z \cdot w &= r_1 r_2 \Big[(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2)\Big] \\ &= r_1 r_2 \Big[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + (\cos\theta_1\sin\theta_2 + \cos\theta_2\sin\theta_1)i\Big] \end{aligned}$$

  Because of the trigonometric formulas for the cosine and sine of a sum of two angles, we see that this can be written as

  $$z \cdot w = (r_1 r_2)(\cos\alpha + i\sin\alpha)$$

  where $\alpha = \theta_1 + \theta_2$. We have proved:

  Theorem. The product of two complex numbers is the complex number whose modulus is the product of the moduluses of the two numbers, and whose angle is the sum of the two angles.

  (Note that the sum of the two angles may be bigger than 2π, so we need to subtract 2π to get back to the usual polar representation.)

- **Powers of complex numbers.** From this multiplication rule, it is easy to understand the geometric effect of raising a complex number to the nth power (where n is a positive integer, i.e. the square, cube, fourth power of a number etc.):

  $$z = r(\cos\theta + i\sin\theta) \implies z^n = r^n(\cos n\theta + i\sin n\theta)$$

The modulus gets raised to the nth power, and the angle gets multiplied by n. Note that in Cartesian coordinates the meaning of exponentiation is much less obvious or intuitive, demonstrating the usefulness of polar coordinates.

One interesting observation that we will need for the proof of the Fundamental Theorem of Algebra is: as z goes in a circle of radius r around 0 (the angle increases continuously from 0 to 2π), its image under the nth power function z^n goes n *times* around the circle of radius r^n.

- **Roots of complex numbers.** The inverse operation to the nth power is extracting an nth root. If $z = r(\cos\theta + i\sin\theta)$, its nth root is given by

  $$\sqrt[n]{z} = \sqrt[n]{r}(\cos(\theta/n) + i\sin(\theta/n)),$$

  i.e., we *divide* the angle by n instead of multiplying. However, in this case there is more than one solution, since we can also add to the angle any number which, when multiplied by n, gives an integer multiple of 2π.

  Theorem. If $z = r(\cos\theta + i\sin\theta) \neq 0$ and n is a positive integer, the nth roots of z are the n distinct numbers

  $$r^{1/n}\left(\cos\left(\frac{\theta + 2\pi k}{n}\right) + i\sin\left(\frac{\theta + 2\pi k}{n}\right)\right), \qquad k = 0, 1, 2, \ldots, n-1$$

- **Examples.**
  - The square roots of -1 are $\cos(\pi/2) + i\sin(\pi/2) = i$ and $\cos(3\pi/2) + i\sin(3\pi/2) = -i$.
  - The fourth roots of 1 are $1, -1, i, -i$.
  - The cube roots of 2 are

    $$\begin{aligned} \sqrt[3]{2}(\cos(0) + i\sin(0)) &= \sqrt[3]{2}, \\ \sqrt[3]{2}(\cos(2\pi/3) + i\sin(2\pi/3)) &= \sqrt[3]{2}\cdot\frac{-1+\sqrt{3}i}{2}, \\ \sqrt[3]{2}(\cos(4\pi/3) + i\sin(4\pi/3)) &= \sqrt[3]{2}\cdot\frac{-1-\sqrt{3}i}{2}. \end{aligned}$$

  - The square roots of $1 + i = \sqrt{2}(\cos\pi/4 + i\sin\pi/4)$ are the numbers

    $$\begin{aligned} w &= \sqrt[4]{2}(\cos\pi/8 + i\sin\pi/8) \\ -w &= \sqrt[4]{2}(\cos 9\pi/8 + i\sin 9\pi/8) \end{aligned}$$

- **Polynomial equations.** Extracting an nth root is a special case of solving a polynomial equation:

  $$z = \sqrt[n]{w} \iff z^n = w \iff z^n - w = 0 \iff p(z) = 0$$

  where $p(z) = z^n - w$.

- **Constant polynomials.** A constant function $p(z) = c$ is called a constant polynomial or polynomial of degree 0. The equation $p(z) = 0$ has no solution if $c \neq 0$, or if $c = 0$ then any value of z is a solution.

- **Polynomials of degree 1.** A function of the form $p(z) = az + b$ is called a polynomial of degree 1. Assuming $a \neq 0$ (since if $a = 0$ this is just a constant polynomial in disguise), the equation $p(z) = 0$ has the unique solution $z = -b/a$.

- **Quadratic polynomials.** A function of the form $p(z) = az^2 + bz + c$ is called a *quadratic polynomial*, or a polynomial of degree 2. Assuming $a \neq 0$, the ancient Babylonians figured out (around 2000 BC) how to solve the equation $p(z) = 0$, arriving at the famous formula

$$z_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

  Of course, if $b^2 - 4ac$ is a negative real number, the equation has no solution in real numbers, but it does have solutions in complex numbers.

- **Cubic polynomials.** A function of the form $p(z) = az^3 + bz^2 + cz + d$ is called a *cubic polynomial*, or a polynomial of degree 2. The general solution of the cubic was found around the year 1530 by the Italian mathematicians del Ferro, Tartaglia and Cardano. This was the original motivation for the definition of complex numbers, since the solution involved extracting cube roots, which sometimes involved complex numbers even when the final answer for the roots was comprised of just real numbers.

- **Quartic polynomials.** A polynomial of degree 4 is called a quartic. The equation $p(z) = 0$ for a quartic was solved by the Italian mathematician Ferrari around 1540.

- **General polynomial.** If n is a positive integer and $a_0, a_1, \ldots, a_n$ are complex numbers, the function

$$p(z) = az^n + bz^{n-1} + cz^{n-2} + \ldots + fz^2 + gz + h$$

  is called a polynomial of degree n. A number z that solves the equation $p(z) = 0$ is called a *root* of the polynomial.

- **The Fundamental Theorem of Algebra.** A fundamental fact about polynomial equations is the following famous result, first proved by the famous German mathematician Karl Friedrich Gauss in 1799.

  Theorem (The Fundamental Theorem of Algebra). Every polynomial equation $p(z) = 0$ has a solution over the complex numbers.

  This result has many different proofs (one full proof can be read in the textbook). We will describe Gauss's original proof, which was incomplete since it relied on a "topological" argument that was not understood rigorously at the time (but has been explained since, though understanding it requires more advanced knowledge of topology that we will not discuss).

  Gauss's proof. We may assume that $p(0) \neq 0$, since otherwise we have a root $z = 0$ and there is nothing to prove. Also assume that the polynomial is monic, i.e., the coefficient of z^n is 1 (otherwise, if it is not 1, simply divide by the coefficient a_n to obtain a new coefficient with a leading coefficient of 1 with the same roots).

  Denote $w = p(0)$. Now observe that:

1. as z goes in a circle of very *small* radius r around 0, since $p(z)$ is a continuous function, the value $p(z)$ will traverse some closed curve that stays very close to w. In particular, such a closed curve cannot go around 0 (since w is some fixed distance from 0).
2. On the other hand, as z goes around a circle of very *large* radius R around 0, the polynomial $p(z)$ should behave more and more like the power function $q(z) = z^n$, since

$$\begin{aligned} p(z) &= z^n + az^{n-1} + bz^{n-2} + cz^{n-3} + \ldots + gz + h \\ &= z^n \left(1 + \frac{a}{z} + \frac{b}{z^2} + \ldots + \frac{g}{z^{n-1}} + \frac{h}{z^n}\right), \end{aligned}$$

   and all of the terms except the first are very small in magnitude (since $|z| = R$ is very large).

   In particular, the image of $p(z)$ as z goes around a circle will be a curve that travels n times around 0 in a counter-clockwise direction (it is like a circle with various "wiggles" and fluctuations from the "go in a circle n times around 0" image of a circle under the power function z^n).
3. The conclusion is that as we vary the radius of the circle from being a very small number "r" to being a very large number "R", the image $p(z)$ goes from being a closed curve that doesn't go around the origin to being a closed curve that travels precisely n times around the origin. Thus, at some point during this process, the curve must cross 0. If you don't believe this, try taking a rubber band and making it go around a nail in the wall, without tearing it up, and without the projection of the rubber band into the plane of the wall crossing the nail at any point. It can't be done!

For more details on this and other proofs of the Fundamental Theorem of Algebra, see

`http://en.wikipedia.org/wiki/Fundamental_theorem_of_algebra`

- **Polynomial roots and factoring. Lemma.** a is a root of the polynomial $p(z)$ if and only if $p(z)$ can be written in the form

$$p(z) = (z - a)q(z)$$

  where $q(z)$ is a polynomial of degree 1 lower than p.

- **The Fundamental Theorem of Algebra, reformulated.** Thanks to the lemma, we can reformulate the FTA as follows: If $p(z)$ is a complex polynomial of degree n, then it can be written as

$$p(z) = c(z - a_1)(z - a_2)\ldots(z - a_n)$$

  for some (not necessarily distinct) complex numbers $a_1, a_2, \ldots, a_n$.

  Proof. Take some root a_1, and write $p(z) = (z - a_1)q(z)$ where $q(z)$ is of degree $n - 1$. Now repeat the same process for $q(z)$, getting a second root a_2, etc., until we get to the form of a product of n factors $z - a_j$ times a constant polynomial.

- **Division of polynomials with remainder.** If $f(z)$ and $g(z)$ are two polynomials, we can do a "long division" of $f(z)$ by $g(z)$ and get a "quotient" and a "remainder", i.e., two polynomial $q(z)$ and $r(z)$ such that

$$f(z) = g(z)q(z) + r(z),$$

and $r(z)$ is of lower degree than $g(z)$. The polynomials $q(z)$ and $r(z)$ are determined uniquely.

- **Example.** Will be given in the discussion section.
- **Proof of the lemma.** Proof of the "only if" claim: If $p(z) = (z - a)q(z)$ then of course $p(a) = (a - a)q(a) = 0$, so a is a root of $p(z)$. Proof of the "if" claim: if $p(a) = 0$, we can divide $p(z)$ by the polynomial $z - a$ (which is of degree 1), to get

  $$p(z) = (z - a)q(z) + r(z),$$

  where $r(z)$ is a polynomial of degree 0, i.e., a constant polynomial. But since $p(a) = 0$, we get

  $$0 = p(a) = (a - a)q(a) + r(a) = r(a),$$

  so in fact $r(z)$ is the 0 polynomial, and $p(z) = (z - a)q(z)$, as claimed.

Lecture 5 (1/15/20)

- **Vector spaces.** The abstract setting for systems of linear equations is a structure called a *vector space* (also called a *linear space*). This is a collection of objects (referred to as "vectors") that can be added to each other, and can be multiplied by a (real or complex) scalar. The prototypical examples of vector spaces are

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \;:\; x_1, \ldots, x_n \in \mathbb{R}\},$$
$$\mathbb{C}^n = \{(x_1, x_2, \ldots, x_n) \;:\; x_1, \ldots, x_n \in \mathbb{C}\}.$$

- **Definition of a vector space.** Let $\mathbb{F}$ represent either the set of real numbers $\mathbb{R}$ or the set of complex numbers $\mathbb{C}$. A vector space (called a *real* vector space if $\mathbb{F} = \mathbb{R}$ or a *complex* vector space if $\mathbb{F} = \mathbb{C}$) is a set V on which are defined two operations "addition" and "scalar multiplication"; addition is defined on pairs of vectors, and scalar multiplication is defined between a scalar (an element of $\mathbb{F}$) and a vector. The operations must satisfy the following properties:

  1. $u + v = v + u$ for any $u, v \in V$.
  2. $(u + v) + w = u + (v + w)$ for any $u, v, w \in V$.
  3. $a \cdot (b \cdot v) = (a \cdot b) \cdot v$ for any $a, b \in \mathbb{F}$, $v \in V$.
  4. There exists an element $0 \in V$ such that $v + 0 = 0 + v = v$ for all $v \in V$.
  5. For every $v \in V$, there exists an element $u \in V$ such that $v + u = 0$.
  6. $1 \cdot v = v$ for any $v \in V$
  7. $a \cdot (u + v) = a \cdot u + a \cdot v$ and $(a + b) \cdot u = a \cdot u + b \cdot u$ for any $a, b \in \mathbb{F}$, $u, v \in V$.

- **Examples.**

  1. $\mathbb{F}$ is itself a vector space where scalars are also considered as vectors.
  2. $\mathbb{F}^n = \{(x_1, \ldots, x_n) \;:\; x_j \in \mathbb{F} \text{ for } j = 1, \ldots, n\}$ — the "standard", or "canonical", n-dimensional real/complex vector space.
  3. $\mathbb{R}^\infty = \{(x_1, x_2, \ldots) \;:\; x_j \in \mathbb{R} \text{ for } j = 1, 2, \ldots\}$ — an "infinite-dimensional" version of $\mathbb{R}^n$ (note that we haven't defined what "dimension" means just yet, but we will).
  4. The space of polynomials of degree n:

$$P_n = \{f : \mathbb{R} \to \mathbb{R} \;:\; f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0\}$$

  5. The space of all polynomials (of any degree).
  6. The "trivial" space $V = \{0\}$.
  7. The space of solutions to a system of linear equations... (more on this later).

Lecture 6 (1/17/20)

- **Properties of vector spaces.**
  - **Proposition 1.** The additive neutral element is unique. **Proof.** ...
  - **Proposition 2.** The additive inverse element is unique. **Proof.** ...
  - **Proposition 3.** $0v = 0$ for all $v \in V$. (Note: the 0 on the left is the scalar 0; the 0 on the right is the *vector* 0. Mathematicians often like to confuse you by using the same symbol to mean two different things!) **Proof.** ...
  - **Proposition 4.** $a0 = 0$ for any $a \in \mathbb{F}$. **Proof.** ...
  - **Proposition 5.** $(-1)v = -v$ for any $v \in V$ (here, $-v$ is the notation for the negative inverse element of v, guaranteed to exist by property 5 in the definition of a vector space). **Proof.** ...

- **Subspaces.** A (vector/linear) subspace is a sub*set* of a vector space which also happens to be by itself a vector space, with the same operations of addition and scalar multiplication.

- **Subspace criterion.** To check that a subset $U \subset V$ of a vector space V is a subspace, one does not need to check all 7 properties in the list, which is tedious. Instead we have an easier criterion:

  Lemma. $U \subset V$ is a subspace if and only if the following conditions hold:

  1. $0 \in U$.
  2. ("closure under addition") If $u, v \in U$ then $u + v \in U$.
  3. ("closure under scalar multiplication") If $a \in \mathbb{F}$, $u \in U$ then $au \in U$.

  Proof. ...

- **Examples.**

  1. $U_{\text{smallest}} = \{0\} \subset V$ and $U_{\text{largest}} = V \subset V$ are both subspaces, respectively the smallest and largest possible subspaces of V.
  2. $\{(x, 0) \ : \ x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.
  3. The space P_n of real polynomials of degree $\leq n$ is a subspace of the vector space of all real polynomials. The space of all polynomials p of degree $\leq n$ such that $p(5) = 0$ is a subspace of *that* subspace.
  4. The set of solutions of a homogeneous system of linear equations in k unknowns is a subspace of $\mathbb{R}^k$.
  5. The union of the x- and y-axes is *not* a subspace of $\mathbb{R}^2$.
  6. If U_1, U_2 are both subspaces of a larger vector space V, then their intersection $U_1 \cap U_2$ is also a subspace.

- **Creating new spaces from old.** There are several ways to get new vector spaces from existing ones:

  1. Subspaces (described above)

2. The intersection of subspaces
3. The *sum* of subspaces
4. The *linear span* of a collection of vectors

We describe these constructions next.

- **Intersection of subspaces.** The intersection $U_1 \cap U_2$ of subspaces of a vector space V is also a vector space, in fact quite a natural one since it is *the largest subspace of V contained in both U_1 and U_2.*

Lecture 7 (1/22/20)

- **Sum of subspaces.** Going in the opposite direction, given subspaces U_1, U_2, we might ask what is *the smallest subspace of V that contains both U_1 and U_2*? The answer is not the *union* of the subspaces, since that is not a subspace, but a new subspace called the *sum* of U_1 and U_2

  Definition. The sum of U_1 and U_2 is the set

  $$U_1 + U_2 = \{u_1 + u_2 \ : \ u_1 \in U_1,\ u_2 \in U_2\}$$

  It is easy to check that $U_1 + U_2$ is itself a subspace of V, and contains U_1 and U_2.

- **Example.** If $U_1 = \{(x, 0, 0) \ : \ x \in \mathbb{R}\}$, $U_2 = \{(0, y, 0) \ : \ y \in \mathbb{R}\}$ then

  $$U_1 + U_2 = \{(x, y, 0) \ : \ x, y \in \mathbb{R}\}.$$

  If we change U_2 to $U_2' = \{(y, y, 0) \ : \ y \in \mathbb{R}\}$ (a different subspace), the sum $U_1 + U_2'$ remains the same as $U_1 + U_2$ before.

- **Direct sum.** The sum $U_1 + U_2$ is called a *direct sum*, and denoted $U_1 \oplus U_2$, if we require the further property that each $v \in U_1 + U_2$ can be written *in a unique way* as a sum of the form $v = u_1 + u_2$, where $u_1 \in U_1$, $u_2 \in U_2$.

  Lemma. If $U_1, U_2 \subset V$ are subspaces, and $W = U_1 + U_2$, then the following conditions are equivalent:

  (1) $W = U_1 \oplus U_2$
  (2) If $0 = u_1 + u_2$ for some $u_1 \in U_1, u_2 \in U_2$, then $u_1 = u_2 = 0$.
  (3) $U_1 \cap U_2 = \{0\}$.

  Proof. It would be enough to prove that: (1) $\implies$ (2); (2) $\implies$ (1); and (3) $\implies$ (1). ...

- **Example.** Define

  $$U_1 = \{(x, y, 0) \in \mathbb{R}^3 \ x, y \in \mathbb{R}\},$$
  $$U_2 = \{(0, 0, z) \in \mathbb{R}^3 \ z \in \mathbb{R}\},$$
  $$U_3 = \{(0, w, z) \in \mathbb{R}^3 \ w, z \in \mathbb{R}\}.$$

  Then $\mathbb{R}^3 = U_1 \oplus U_2$, but $\mathbb{R}^3 = U_1 + U_3$ is *not* a direct sum.

- **Example.** Define

  $$P_3 = \{p(x) = ax^3 + bx^2 + cx + d \ : \ a, b, c, d \in \mathbb{R}\},$$

  the vector space of polynomials of degree ≤ 3 with real coefficients. Define

  $$U_1 = \{p(x) \in P_3 \ : \ p(0) = 0 \ \},$$
  $$U_2 = \{p(x) = ax^3 + bx^2 + cx + d \in P_3 \ : \ a = 0\}$$

  Then $P_3 = U_1 + U_2$ (a small exercise: check this), but it is not a direct sum.

Lecture 8 (1/24/20)

- **Linear combinations.** Let V be a vector space over $\mathbb{F}$. If $v_1, v_2, \dots, v_m$ are vectors in V, an expression of the form

$$a_1 v_1 + a_2 v_2 + \ldots + a_m v_m,$$

where $a_1, \dots, a_m$ are scalars from $\mathbb{F}$, is called a *linear combination*.

- **Linear span.** If $v_1, \dots, v_m$ are vectors in a vector space V, define the *linear span* of $v_1, \dots, v_m$ as

$$\mathrm{span}(v_1, \dots, v_m) = \{a_1 v_1 + \ldots + a_m v_m \ : \ a_1, \dots, a_m \in \mathbb{F}\},$$

i.e., the span is the set of all linear combinations of $v_1, \dots, v_m$.

Lemma. The span of $v_1, \dots, v_m$ is a subspace. Furthermore, it is the smallest possible subspace that contains $v_1, \dots, v_m$ (more precisely, what this means is that if $U \subset V$ is any subspace that contains $v_1, \dots, v_m$, then $\mathrm{span}(v_1, \dots, v_m) \subset U$).

Proof: This is similar to the property of a sum of subspaces mentioned above. In fact, an equivalent way to define $\mathrm{span}(v_1, \dots, v_m)$ would be to write it as

$$\mathrm{span}(v_1, \dots, v_m) = V_1 + V_2 + \ldots + V_m,$$

where $V_i = \{t v_i \ : \ t \in \mathbb{F}\}$ is the subspace consisting of v_i and all its scalar multiple (geometrically, V_i is a line through 0 and v_i).

Finite and infinite dimension. If V is a vector space, we say that it is *finite-dimensional* if there are vectors $v_1, \dots, v_m \in V$ such that $V = \mathrm{span}(v_1, \dots, v_m)$. Otherwise, we say V is infinite-dimensional, i.e., if it is not spanned by a finite set of vectors.

If $V = \mathrm{span}(v_1, \dots, v_m)$, does it make sense to say that V has dimension m? No, since there may be many spanning sets, not all of them containing the same number of elements. But we will soon figure out the right way to define the actual dimension.

Examples. $\mathbb{F}^n$ is finite-dimensional since it is spanned by

$$e_1 = (1, 0, \dots, 0), e_2 = (0, 1, \dots, 0), \dots, v_n = (0, 0, \dots, 1).$$

Similarly, the space P_n of polynomials of degree $\leq n$ is finite-dimensional; it is spanned by the monomials $1, z, \dots, z^n$. The space of all polynomials of arbitrary degree is infinite-dimensional (proof: ...).

- **Linear independence. Definition.** The vectors $v_1, \dots, v_m \in V$ are called *linearly independent* if whenever we have

$$a_1 v_1 + \ldots a_m v_m = 0$$

it follows that $a_1 = a_2 = \ldots = a_m = 0$. I.e., the only linear combination of $v_1, \dots, v_m$ that equals the zero vector is the obvious one with all coefficients equal to 0. If the vectors are not linearly independent, they are called *linearly dependent*.

Examples.

1. The vectors $e_1, \ldots, e_n$ mentioned before as a spanning set for $\mathbb{R}^n$ are linearly independent. Proof: ...
2. The vectors $v_1 = (1, 1, 1), v_2 = (0, 1, -1), v_3 = (1, 2, 0)$ in $\mathbb{R}^3$ are linearly dependent. To see this, we look for coefficients a_1, a_2, a_3 such that $a_1 v_1 + a_2 v_2 + a_3 v_3 = (0, 0, 0)$. This leads to a system of linear equations. It is not difficult to find that it has the solution $(a_1, a_2, a_3) = (1, 1, -1)$ (as well as any scalar multiple of this solution).

Lemma. $v_1, \ldots, v_m$ are linearly independent if and only if every vector $v \in \text{span}(v_1, \ldots, v_m)$ can be written in a unique way as a linear combination of $v_1, \ldots, v_m$.

Proof. ...

- **Linear dependence means the spanning set can be made smaller. Lemma.** If vectors $v_1, \ldots, v_m$ in a vector space V are linearly dependent, then there is an index $1 \leq j \leq m$ such that

  1. $v_j \in \text{span}(v_1, \ldots, v_{j-1})$ (here, if $j = 1$ this statement should be interpreted as saying that $v_1 = 0$).
  2. $\text{span}(v_1, \ldots, \widehat{v}_j, \ldots, v_m) = \text{span}(v_1, \ldots, v_m)$, where $\widehat{v}_j$ means that v_j is omitted from the list.

  Proof. ...

Lecture 9 (1/27/20)

- **A spanning set which is linearly independent is minimal.** The lemma above could actually be formulated as an "if and only if" statement. The easy converse ("only if") part says that if the vectors are linearly independent, then removing any of them from the list makes the span a strictly smaller subspace. The following theorem makes a much stronger claim:

  Theorem. If vectors $v_1, \ldots, v_m$ in a vector space V are linearly independent, then for any vectors $w_1, \ldots, w_n$, if they span V (i.e., if $V = \text{span}(w_1, \ldots, w_n)$) then $n \geq m$.

  Proof. (Important) See pages 43–44 in the textbook.

- **Bases.** We now get to the important concept of a *basis*. **Definition.** A sequence of vectors $v_1, \ldots, v_m$ in a finite-dimensional vector space V is called a *basis* if the vectors are linearly independent and $V = \text{span}(v_1, \ldots, v_m)$.

- **Dimension.** By the theorem above, all bases have the same size (any spanning set has at least as many elements as any linearly independent set; a basis has both properties so we'll get an inequality in both directions), so we can use it to define the dimension. **Definition.** The *dimension* of a finite-dimensional vector space V is the size of any basis (and therefore all bases) of V.

- **Is this a good definition?** Note that we haven't yet proved that any finite-dimensional space has even one basis, which leaves the theoretical possibility of a finite-dimensional vector space whose dimension is undefined. This problem is remedied by the following theorem:

  Basis reduction theorem. If $V = \text{span}(v_1, \ldots, v_m)$ then either $v_1, \ldots, v_m$ is a basis of V or some vectors can be removed from the list to obtain a basis for V.

  Proof. Successively remove any v_i which is in the span of the ones preceding it in the list. Each such removal does not change the span. Eventually we end up with a spanning set in which no vector is in the span of the ones preceding it, and by the lemma above that set must be linearly independent, hence it is a basis.

- **Corollary.** Every finite-dimensional vector space has a basis, since it has a spanning set (by the definition of finite-dimensionality) which, by the theorem above, can be thinned to give a basis.

- **The dimension is well-defined.** If V is finite-dimensional, it has a basis, therefore its dimension is defined.

- **Examples**

  1. $e_1, \ldots, e_n$ form a basis for $\mathbb{R}^n$.
  2. The polynomials $1, z, z^2, \ldots, z^n$ are a basis for the space P_n of polynomials of degree at most n.
  3. Let $\mathcal{S} = \{(1, -1, 0), (2, -2, 0), (-1, 0, 1), (0, -1, 1), (0, 1, 0)\}$. One can verify that an arbitrary vector $v = (x, y, z) \in \mathbb{R}^3$ can be written as a linear combination

$$v = (x + z)(1, -1, 0) + 0(2, -2, 0) + z(-1, 0, 1) + 0(0, -1, 1) + (x + y + z)(0, 1, 0)$$

of the elements of $\mathcal{S}$. Therefore $\mathcal{S}$ is a spanning set for $\mathbb{R}^3$. However, $\dim \mathbb{R}^3 = 3$, so we can replace $\mathcal{S}$ with the smaller set $\mathcal{B} = \{(1,-1,0), (-1,0,1), (0,1,0)\}$, which is linearly independent and therefore a basis.

- **Basis extension theorem.** If $v_1, \ldots, v_m$ are linearly independent vectors in a finite-dimensional vector space V, either they are a basis, or we can add additional vectors $v_{m+1}, \ldots, v_n$ to them to form a basis.

  Proof. ...

Lecture 10 (1/29/20)

- **Summary.** To summarize the discussion on bases and dimension, here are a few additional facts that follow as easy consequences of the results we showed:

  1. The dimension is the smallest size of a spanning set.

     Proof. If there is a spanning set of size n, then as we saw from the basis reduction theorem, the dimension is $\leq n$. But if there is no spanning set of smaller size, in particular the smallest basis has size $\geq n$ and therefore the dimension is $\geq n$.

  2. The dimension is the largest size of a linearly independent set.

     Proof. If there is a linearly independent set of size n, then as we saw from the basis extension theorem, the dimension is $\geq n$. But if there is no linearly independent set of bigger size, in particular the largest basis has size $\leq n$ and therefore the dimension is $\leq n$.

  3. Any spanning set whose size is the dimension is a basis.

     Proof. If a spanning set is not a basis, it can be reduced to a basis, so its size must be strictly bigger than the dimension.

  4. Any linearly independent set whose size is the dimension is a basis.

     Proof. If a linearly independent set is not a basis, it can be extended to a basis, so its size must be strictly smaller than the dimension.

  5. If $U \subset V$ is a subspace then $\dim U \leq \dim V$.

     Proof. Take a basis for U. In particular it is linearly independent (in U, and therefore also in V) and hence is either a basis for V or can be extended to a basis of V by adding more vectors to it.

- **Theorem.** If $U, W \subset V$ are subspaces of a finite-dimensional vector space V, then

  $$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

  In particular, if $U + W = U \oplus W$ then $\dim(U + W) = \dim(U) + \dim(W)$.

  Proof. (Important) See page 48 in the textbook.

Lecture 11 (1/31/20)

- **Linear transformations.** If V, W are vector spaces, a function $T : V \to W$ is called a *linear transformation*, or *linear map*, if it satisfies the properties

  1. For any $u, v \in V$, $T(u + v) = T(u) + T(v)$.
  2. For any $a \in \mathbb{F}$, $v \in V$, we have $T(av) = a\,T(v)$.

  We can write a single condition that encompasses both conditions at the same time: it is easy to see that T is linear if and only if it satisfies

  $$T(au + bv) = a\,T(u) + b\,T(v)$$

  for any $a, b \in \mathbb{F}$ and $u, v \in V$.

  The space V is called the *domain* of the linear transformation. The space W is called the *co-domain*.

  Note that a linear transformation always maps the zero vector in V to the zero vector in W, i.e., it satisfies $T(0) = 0$, since $T(0) = T(0 \cdot 0) = 0 \cdot T(0) = 0$.

  The set of linear transformations from V to W is denoted by $\mathcal{L}(V, W)$ (this is a vector space!). If $V = W$ we denote $\mathcal{L}(V) = \mathcal{L}(V, V)$ and refer to linear transformations $T : V \to V$ as *linear operators*.

- **Examples.**

  1. **The zero map** $0 : V \to W$ sends every vector $v \in V$ to $0 \in W$.
  2. **The identity map** $I : V \to V$ (also denoted 1) is defined by $I(v) = v$.
  3. A function $f : \mathbb{R} \to \mathbb{R}$ is linear if and only if $f(x) = ax$ for some $a \in \mathbb{R}$. So $f(x) = e^x$ for example is not linear. A first-degree polynomial $g(x) = ax + b$ is sometimes referred to as a linear function or linear polynomial, but this terminology is inconsistent with our current one so we will not use it.
  4. On the space P_∞ of polynomials (or more generally on the vector space of differentiable functions) we can define the differentiation operator $T(f) = f'(x)$.
  5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (x - 2y, 3x + y)$ is linear. This is an example of matrix multiplication, since we can write

     $$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$$

     We shall see that linear transformations can in general be encoded (in some sense that will be explained) by such matrix multiplication operations, and that is what ties them to systems of linear equations, which can also be written in terms of matrix multiplication.

- **Describing a linear transformation.** A linear transformation seems to contain a lot of information — all possible values $T(v)$ for all possible vectors in V — but in fact, as the following lemma shows, it is enough to specify their values on a much smaller set.

Theorem. Let $v_1, \ldots, v_m$ be a basis for V. A linear transformation $T : V \to W$ is uniquely determined by the vectors

$$w_1 = T(v_1), w_2 = T(V_2), \ldots, w_m = T(v_m).$$

That is, for any list of vectors $w_1, \ldots, w_m \in W$ there exists exactly one linear transformation $T : V \to W$ for which $T(v_j) = w_j$ for $j = 1, \ldots, m$.

Proof. ...

Lecture 12 (2/3/20)

- **Composition/product of linear maps.** If V, U, W are vector spaces and $S \in \mathcal{L}(V, U), T \in \mathcal{L}(U, W)$, we can define a function $R : V \to W$ by

$$R(v) = T(S(v))$$

  R is called the composition of the functions T and S, and denoted $R = T \circ S$. It is easy to verify that it is also a linear transformation. In linear algebra, sometimes R will be referred to as the *product* of the linear transformations, and denoted $R = TS$. Talking about composition in this way makes sense, since it shares the following properties with other "products":

  1. **Associativity:** $(T_1T_2)T_3 = T_1(T_2T_3)$
  2. **Neutral element:** $TI = IT = T$ where I is the identity map on the appropriate space (if $T \in \mathcal{L}(V, W)$ then the I in TI is $I : V \to V$ and the I in IT is $I : W \to W$).
  3. **Distributivity:** $(T_1 + T_2)S = T_1S + T_2S$ and $T(S_1 + S_2) = TS_1 + TS_2$.

  The main place where multiplication of linear maps differs from normal multiplication is:

  4. **NO Commutativity:** It is **not** true that $TS = ST$ for all linear transformations S, T (even when both ST and TS are defined; sometimes only one of them makes sense). Here is an example where $TS \neq ST$: on $\mathbb{R}^2$ take $T(x, y) = (y, x)$ and $S(x, y) = (x, -y)$.

- **Null space and range.** Given a linear transformation $T : V \to W$, we can define two interesting linear subspaces:

  1. The *null space* of T, denoted null(T) (also sometimes called the *kernel* of T and denoted ker(T)), is the set vectors in V that T maps to $0 \in W$:

$$\text{null}(T) = \{v \in V \ : \ T(v) = 0\}.$$

  2. The *range* of T, denoted range(T) (also sometimes called the *image* of T and denoted im(T)), is the set of vectors in W to which T maps *some* vector in V:

$$\text{range}(T) = \{T(v) \ : \ v \in V\}.$$

- **Claim:** (easy) null(T) is a linear subspace of V, and range(T) is a linear subspace of W.

- **Injective transformations. Definition.** $T \in \mathcal{L}(V, W)$ is called *injective* (or *one-to-one*) if for any $u, v \in V$, if $T(v) = T(u)$ then $u = v$. Equivalently, T is injective if it maps any distinct vectors $u \neq v \in V$ to distinct vectors $T(u) \neq T(v) \in W$.

- **Proposition.** $T \in \mathcal{L}(V, W)$ is injective if and only if null$(T) = \{0\}$.

  Proof. ...

- **Examples**

  1. The differentiation operator on polynomials is not injective.

2. The identity map is injective.
3. The linear map that sends a polynomial $p(z)$ to $z^2p(z)$ is injective.
4. The linear map $T(x, y) = (x - 2y, 3x + y)$ is injective.

- **Surjective transformations. Definition.** A linear transformation $T \in \mathcal{L}(V, W)$ is called *surjective* (or (*onto*) if for any $w \in W$ there exists a $v \in V$ such that $T(v) = w$. Equivalently, T is injective if range$(T) = W$, i.e., the range of T is equal to its co-domain.

- **Examples**

1. The identity map is surjective.
2. The differentiation operator on polynomials is surjective.
3. The map $T(x, y) = (x - 2y, 3x + y)$ is surjective. Given a vector $w = (a, b) \in \mathbb{R}^2$, we can solve the equation $T(x, y) = (a, b)$ and obtain $(x, y) = \frac{1}{7}(a + 2b, -3a + b)$. Note that the fact that there is a solution means T is surjective; the fact that there is a *unique* solution means it's injective.
4. The map on polynomials that sends $p(z)$ to $z^2p(z)$ is not surjective, since no polynomials of degree 0 or 1 are in its image.

Lectures 13 (2/5/20) and 14 (2/7/20)

- **The dimension formula. Theorem.** If V, W are vector spaces and $T : V \to W$ is a linear transformation, then

$$\dim(V) = \dim(\mathrm{null}(T)) + \dim(\mathrm{range}(T)).$$

Proof. (Important) See pages 56–57 in the textbook.

- **Corollary.** Let $T \in \mathcal{L}(V, W)$.
  1. If $\dim(V) > \dim(W)$ then T is not injective.
  2. If $\dim(V) < \dim(W)$ then T is not surjective.

- **Coordinate vectors.** Let $B = \{v_1, \dots, v_n\}$ be a basis of a vector space V. Any other vector v can be written in a unique way as a linear combination

$$v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n.$$

Another way of thinking about this is that the vector of coefficients $(a_1, \dots, a_n)$ (which is a vector in $\mathbb{F}^n$) encodes the vector v (which is an element of an abstract vector space V).

Definition. The vector of coefficients in the linear combination (traditionally written as a column vector) is called the *coordinate vector* of v in the basis B, and denoted

$$[v]_B = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Note that the function mapping a vector v to its coordinate vector $[v]_B$ is linear, i.e., we have the relations $[u+v]_B = [u]_B + [v]_B$ and $[av]_B = a[v]_B$ for $u, v \in V$ and $a \in \mathbb{F}$.

- **Representing a linear transformation as a matrix.** If $T : V \to W$ is a linear transformation, we can encode it as a matrix. Fix a basis $B = \{v_1, \dots, v_n\}$ of V and a basis $C = \{w_1, \dots, w_m\}$ of W. For each $j = 1, \dots, n$, the vector $T(v_j)$ is in W so we can consider its coordinate vector $u_j = [T(v_j)]_C$ in the basis C (which is a vector in $\mathbb{F}^m$). Now package all the vectors $u_1, \dots, u_n$ in a matrix with m rows and n columns. This is called the *matrix representing T in the bases B and C*, and denoted $M(T)$:

$$M(T) = \begin{pmatrix} | & | & & | \\ [T(v_1)]_C & [T(v_2)]_C & \cdots & [T(v_n)]_C \\ | & | & & | \end{pmatrix}.$$

We can also write the matrix in terms of its entries

$$M(T) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

or abbreviate it as $M(T) = (a_{ij})_{1\le i\le m, 1\le j\le n}$. Its defining property is that for each $1 \le j \le n$ we have the equation

$$T(v_j) = a_{1j}w_1 + a_{2j}w_2 + \ldots + a_{nj}w_n.$$

expressing $T(v_j)$ as a linear combination of the elements of the basis C.

- **Example.** If $T(x, y) = (ax+by, cx+dy)$ (which can also be written as matrix multiplication of $\begin{pmatrix} x \\ y \end{pmatrix}$ by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$), then the matrix $M(T)$ representing T in the standard basis $B = C = \{e_1, e_2\}$ of $\mathbb{R}^2$ is again the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

  More generally, for each $m \times n$ matrix $M = (m_{i,j})_{1\le i\le m, 1\le j\le n}$ we have the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by matrix multiplication:

$$T(v) = M \cdot v.$$

  If we take the bases B and C to be the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, then the representing matrix $M(T)$ is simply the original matrix M we started with.

- **Example.** Let $T(x, y) = (x + 2y, -x + 3y)$. In the standard basis $B = C = \{(1, 0), (0, 1)\}$ the matrix representing T is

$$M(T)^B_C = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}.$$

  What happens if we change the basis? Take $B = \{(1, 1), (0, 1)\}$ and $C = \{(1, 0), (0, 1)\}$. In this case we may compute:

$$T(1, 1) = (3, 2) = 3(1, 0) + 2(0, 1), \quad T(0, 1) = (2, 3) = 2(1, 0) + 3(0, 1),$$

  so

$$M(T)^B_C = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

  If we exchange the roles of C and B we get (check!)

$$M(T)^C_B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

- **Example.** Let $B = C = \{1, z, z^2, z^3\}$ be bases for the space P_3 of polynomials of degree ≤ 3, and let $T(p) = p'$ be the derivative map. In this case it is not difficult to check that the representing matrix is

$$M(T) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- **Matrix multiplication.** If $M = (a_{ij})_{1\le i\le m, 1\le j\le n}$ and $N = (b_{ij})_{1\le i\le n, 1\le j\le p}$, the matrix product MN is computed by multiplying the matrix M by each of the columns of N, and packaging the results in a new $m \times p$ matrix. Formally, we have $MN = (c_{ij})_{1\le i\le m, 1\le j\le p}$, where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Another way of thinking about it is that we compute all possible "dot products" of rows of M with columns of N, where a dot product of two (row or column) vectors $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ is equal to the sum of the products of the coordinates, i.e.

$$(x_1, \ldots, x_n) \cdot (y_1, \cdot y_n) = x_1 y_1 + \ldots + x_n y_n.$$

The matrix product MN is the rectangular table of numbers containing all these dot products.

- **Examples:** ...

- **Matrix multiplication and composition of linear transformations.** A fundamental fact related matrix multiplication with composition of linear maps:

  Theorem. U, V, W be finite-dimensional vector spaces. Let $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$ be linear transformations. Fix bases $B = \{v_1, \ldots, v_m\}$, $C = \{w_1, \ldots, w_n\}$, $D = \{u_1, \ldots, u_p\}$ of V, W, U respectively. Then we have

  $$M(S \circ T) = M(S)M(T),$$

  where $M(T) = M(T)_C^B$ is the matrix representing T in the bases B and C; $M(S) = M(S)_D^C$ is the matrix representing S in the bases C and D; and $M(S \circ T)$ is the matrix representing the composition $S \circ T$ in the bases B and D.

  Proof. Denote $M(T) = (a_{ij})_{1\le i\le n, 1\le j\le m}$, $M(S) = (b_{ij})_{1\le i\le p, 1\le j\le n}$.

  $$\begin{aligned}(S \circ T)v_j &= S(Tv_j) = S(b_{1j}w_1 + b_{2j}w_2 + \ldots + b_{nj}w_n) = b_{1j}S(w_1) + \ldots + b_{nj}S(w_n) \\ &= b_{1j}(a_{11}u_1 + a_{21}u_2 + \ldots + a_{p1}u_p) + b_{2j}(a_{12}u_1 + a_{22}u_2 + \ldots + a_{p2}u_p) \\ &\quad + \ldots + b_{nj}(a_{1n}u_1 + a_{2n}u_2 + \ldots + a_{pn}u_p) \\ &= c_{1j}u_1 + c_{2j}u_2 + \ldots + c_{pj}u_p,\end{aligned}$$

  where

  $$\begin{aligned}c_{1j} &= a_{11}b_{1j} + a_{12}b_{2j} + \ldots + a_{1n}b_{nj}, \\ c_{2j} &= a_{21}b_{1j} + a_{22}b_{2j} + \ldots + a_{2n}b_{nj}, \\ &\vdots \\ c_{ij} &= a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}, \\ &\vdots \\ c_{pj} &= a_{p1}b_{1j} + a_{p2}b_{2j} + \ldots + a_{pn}b_{nj}.\end{aligned}$$

  By the definition of the representing matrix, the numbers c_{ij} are exactly the entries of the matrix $M(S \circ T)$.

- **Corollary.** Matrix multiplication has the same properties that composition of linear transformations has:

  1. **Associativity:** $(M_1M_2)M_3 = M_1(M_2M_3)$
  2. **Neutral element:** $MI = IM = M$ where I is the identity matrix of the appropriate size.
  3. **Distributivity:** $(M_1 + M_2)N = M_1N + M_2N$ and $M(N_1 + N_2) = MN_1 + MN_2$.
  4. **NO Commutativity:** It is **not** true that $MN = NM$ for all matrices M, N (even when both MN and NM are defined).

- **Inverse functions.** A function $f : A \to B$ (where A and B are sets) is called invertible if for each $b \in B$ there is a unique $a \in A$ for which $f(a) = b$. In this case we denote $a = f^{-1}(b)$ and call $f^{-1} : B \to A$ the inverse function of f, and say that f is invertible. It has the property that $f^{-1} \circ f =$ the identity function on A, $f \circ f^{-1} =$ the identity function on B.

  It is easy to see that f is invertible if and only if it is both injective and surjective: injectivity means for any $b \in B$ there is *at most one* $a \in A$ such that $f(a) = b$, and surjectivity means for any $b \in B$ there is *at least one* $a \in A$ such that $f(a) = b$.

Lecture 15 (2/10/20)

- **Invertible transformations.** A linear transformation $T : V \to W$ is called *invertible* if it is invertible as a function, and also its inverse $T^{-1} : W \to V$ is itself a linear transformation. It turns out this last condition is redundant: T^{-1} is automatically linear; in other words we have the following result:

  Proposition. The following conditions are equivalent:

  1. T is invertible as a function, i.e., is injective and surjective.
  2. T is invertible as a linear transformation.
  3. $T : V \to W$ is invertible if and only if there exists a linear transformation $S : W \to V$ such that
  $$T \circ S = I_W \quad \text{and} \quad S \circ T = I_V$$
  (here I_W denotes the identity map on W and I_V denotes the identity map on V). Note that in this case, S is determined uniquely, since if $S' : V \to W$ also satisfies $T \circ S' = I_W, S' \circ T = I_V$, then
  $$S = S \circ I_W = S \circ (T \circ S') = (S \circ T) \circ S' = I_V \circ S' = S'$$

  Proof. (1) $\Longrightarrow$ (2): If $w_1, w_2 \in W$, $a, b \in \mathbb{F}$, denote $v_1 = T^{-1}(w_1), v_2 = T^{-1}(w_2)$. Then
  $$T(av_1 + bv_2) = aT(v_1) + bT(v_2) = aw_1 + bw_2,$$
  but that means that $av_1 + bv_2$ is the inverse image of $aw_1 + bw_2$, i.e.,
  $$T^{-1}(aw_1 + bw_2) = aT^{-1}(w_1) + bT^{-1}(v_2),$$
  which is exactly what we need to know that T^{-1} is a linear transformation.

  (2) $\Longrightarrow$ (3): T^{-1} is exactly the S we need.

  (3) $\Longrightarrow$ (1): If such an S exists then T is injective and surjective and S is its inverse (as a function): if $T(v) = T(v')$ for some $v, v' \in V$ then $v = I_V(v) = S \circ T(v) = S \circ T(v') = I_V(v') = v'$ (proving injectivity), and for any $w \in W$, $T(S(w)) = T \circ S(w) = I_W(w) = w$ (proving surjectivity).

- **Example.** The transformation $T(x, y) = (x - 2y, 3x + y)$ is invertible. Compute its inverse. **Solution.** ...

- Isomorphic vector spaces. **Definition.** If V, W are vector spaces, they are called *isomorphic* if there exists an invertible linear map $T \in \mathcal{L}(V, W)$.

  For finite-dimensional vector spaces, the question of whether two given spaces are isomorphic has an easy answer.

  Theorem. If V, W are finite-dimensional vector spaces, then they are isomorphic if and only if $\dim(V) = \dim(W)$.

  Proof. ...

- **Invertible linear operators. Theorem.** If V is a finite-dimensional vector space and $T : V \to V$ is a linear operator, then the following are equivalent:

1. T is invertible.
2. T is injective.
3. T is surjective.

Proof. ...

- **Invertible matrices.** A square matrix $M = (a_{ij})_{1 \leq i,j \leq n}$ is invertible if there is a matrix $B = (b_{ij})_{1 \leq i,j \leq n}$ such that $AB = BA = I$ (where I denotes the identity matrix of order n). This is equivalent to the linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ defined by $T(v) = A \cdot v$ being invertible. We can rewrite the theorem above in terms of matrices as follows:

  Theorem. For a square matrix A, the following conditions are equivalent:

  1. A is invertible.
  2. the system of linear equations $Av = 0$ has the unique solution $v = 0$ (the zero vector).
  3. The columns of the matrix A span $\mathbb{F}^n$.
  4. The reduced row-echelon form of A is the identity matrix.

  Proof. ...

Lecture 16 (2/14/20)

- **Algorithm to find the inverse of a matrix.** In the discussion section you will learn a simple method to find the inverse of a matrix by using Gaussian elimination.

- **Inverse of a 2×2 matrix.** We saw that the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. In that case it is not difficult to verify that its inverse matrix is

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

For higher-order square matrices, it would be nice if we had a simple criterion to determine whether the matrix is invertible. It turns out that there is a function of the matrix entries, generalizing the function $ad - bc$, that gives such a criterion. This function is called the *determinant*.

Permutations. To define the determinant of an $n \times n$ matrix we need the concept of a permutation.

Definition. A *permutation* (of order n) is a function $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ which is a bijection (a.k.a invertible, i.e., it is injective and surjective).

One can think of a permutation as a way of arranging n objects in a line: object number 1 goes in position number $\sigma(1)$ (from the left), object number 2 goes in position number $\sigma(2)$, etc. Because permutations are such special functions, we use a special notation to denote them in what's known as two-line notation: the permutation σ will be written as

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}$$

Note that the top line in two-line notation is redundant — it will always consist of the numbers $1, \ldots, n$ arranged in increasing order — but it is useful for the purposes of getting a "feel" for what the permutation does and for manipulating it in some ways, as we will see.

- **Example.** There are 2 different permutations of order 2: ..., and here are the 6 different permutations of order 3: ...

- Denote by S_n the set of permutations of order n (this is a standard notation used pretty much in all of mathematics).

Theorem. The number $|S_n|$ of permutations of order n is given by the factorial function

$$n! = 1 \cdot 2 \cdot 3 \ldots (n-1)n.$$

Proof. ...

Note. The factorial function grows very fast. Here are its first few values:

n	1	2	3	4	5	6	7	8	9	10
$n!$	1	2	6	24	120	720	5040	40320	362880	3628800

Unfortunately, the determinant, which is the magical function that will tell us if a square matrix is invertible, will be defined as a sum of $n!$ terms, one for each permutation. So, for large matrices, computing it seems hopeless (but it isn't, as it turns out).

Lecture 17 (2/19/20)

- **The identity permutation.** The most important permutation in S_n is the *identity permutation* $\sigma = \text{id}$ defined by $\sigma(j) = j$ for $j = 1, \ldots, n$.

- **Composition of permutations.** Permutations, like general functions, can be composed with each other. If $\sigma, \pi \in S_n$ we define their composition $\sigma \circ \pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ by

$$(\sigma \circ \pi)(j) = \sigma(\pi(j)), \qquad j = 1, \ldots, n$$

  Examples. ...

- **Properties of composition of permutations. Theorem.** The composition of permutations is a "nice" operation; it has the following properties:

  1. The composition of two permutations is itself a permutation.
  2. Associativity: $\pi \circ (\sigma \circ \tau) = (\pi \circ \sigma) \circ \tau$ for any $\pi, \sigma, \tau \in S_n$.
  3. Neutral element: $\pi \circ \text{id} = \text{id} \circ \pi = \pi$ for any $\pi \in S_n$.
  4. Inverse element: for any $\pi \in S_n$ there is a unique inverse permutation π^{-1} such that $\pi \circ \pi^{-1} = \pi^{-1} \circ \pi = \text{id}$.
  5. NO commutativity: the composition is in general not commutative, i.e. $\pi \circ \sigma$ is not necessarily equal to $\sigma \circ \pi$ (and in most cases it isn't).

- **Inversions.** There is a way to measure how "badly ordered" a permutation is, using a concept called *inversions*. If i, j are two numbers between 1 and n and $\sigma \in S_n$, then we say that the pair (i, j) is an inversion pair of σ if $i < j$ but $\sigma(i) > \sigma(j)$. The total number of inversion pairs is called the *inversion number* of the permutation.

  Examples. The identity permutation has inversion number 0 (the smallest possible). The *reverse permutation* $\begin{pmatrix} 1 & 2 & \ldots & n-1 & n \\ n & n-1 & \ldots & 2 & 1 \end{pmatrix}$ has inversion number

$$1 + 2 + 3 + \ldots + n = \frac{n(n-1)}{2},$$

  which is the largest possible inversion number for a permutation of order n.

  In general, the inversion number can be visualized as the number of "line crossings" in a diagram showing the numbers $1, \ldots, n$ in the top row, the same numbers (in the same order) in the bottom row, with lines connecting each number j on the top row to the number $\sigma(j)$ on the bottom row.

- **The sign of a permutation.** With each permutation we associate a sign $\text{sign}(\sigma) = \pm 1$ given by

$$\text{sign}(\sigma) = (-1)^{(\text{the inversion number of } \sigma)} = \begin{cases} +1 & \text{even number of inversions,} \\ -1 & \text{odd number of inversions.} \end{cases}$$

Theorem. Assume that a permutation σ' is obtained from $\sigma \in S_n$ by swapping the values of i and j. In other words, $\sigma' = t_{i,j} \circ \sigma$, where $t_{i,j}$ is the permutation (called a *transposition*) which leaves all numbers unchanged except for swapping the values i and j, i.e.,

$$t_{i,j} = \begin{pmatrix} 1 & 2 & 3 & \dots & i & \dots & j & \dots & n-1 & n \\ 1 & 2 & 3 & \dots & j & \dots & i & \dots & n-1 & n \end{pmatrix}.$$

Then we have $\operatorname{sign}(\sigma') = -\operatorname{sign}(\sigma)$.

Sketch of proof. The proof is a bit elaborate so we won't write it fully, but here's the idea. First, prove this in the case when the transposition $t_{i,j}$ is an *adjacent transposition*, meaning that $j = i \pm 1$. In this case, it is not difficult to see that the inversion number of σ' differs from that of σ by either $+1$ or -1, and that means that the sign gets inverted. Next, for a general transposition $t_{i,j}$ with $1 < i < j \leq n$, show that $t_{i,j}$ can be expressed as a composition of *an odd number* of adjacent transpositions. Since each adjacent transposition has the effect of inverting the sign of the permutation, composing an odd number of them will also have that effect.

- **Corollary.** $\operatorname{sign}(\sigma \circ \pi) = \operatorname{sign}(\sigma)\operatorname{sign}(\pi)$.

  Proof. By the result above, this is true when σ is a transposition. By induction, it is also true when σ is a composition of many transpositions. It remains to observe that any permutation can be represented as a composition of transpositions — this is left to the reader as an exercise.

- **Definition of the determinant.** Finally, we can define the determinant of a square matrix.

  Definition. If $M = (a_{ij})_{i,j=1}^n$ is an $n \times n$ square matrix, its determinant is defined as

$$\det(M) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

- **Examples.** The determinants of a general 2×2 matrix and of a general 3×3 matrix are:
  ...

Lectures 18 (2/21/20) and 19 (2/24/20)

- **Properties of the determinant.**

  1. $\det(I_n) = 1$. More generally, if $A = (a_{ij})_{i,j=1}^n$ is a *diagonal matrix*, i.e., all of its off-diagonal entries are 0, then $\det(A) = a_{11}a_{22}\ldots a_{nn}$ is the product of the diagonal entries.
  2. More generally, even if A is *upper triangular* or *lower triangular*, i.e., all of its entries below (respectively, above) the main diagonal are 0, then the formula $\det(A) = a_{11}a_{22}\ldots a_{nn}$ still holds.
  3. $\det(A) = \det(A^\top)$, where $A^\top$ denotes the *transpose matrix* of A, whose entry in position (i,j) (ith row, jth column) is a_{ji} (the entry in the jth row, ith column of the original matrix).
  4. If A has two identical rows/columns, then $\det(A) = 0$.
  5. If A has a row/column of 0s, then $\det(A) = 0$.
  6. $\det(A)$ is linear in each of the rows of A. I.e., if A, A' are identical except in the jth row then $\det(A + A') = \det(A) + \det(A')$; if A' is obtained from A by multiplying the jth row by a scalar c then $\det(A') = c\det(A)$.
  7. If E is an elementary matrix (a matrix such that multiplying A from the left by E has the effect of performing an elementary row operation on A; see section 12.3.1 in the textbook) then $\det(EA) = \det(E)\det(A)$. It is easy to compute $\det(E)$ for each of the elementary matrix types.
  8. $\det(A) = \det(\mathrm{RREF}(A))/(\det(E_1)\ldots\det(E_k))$, where $\mathrm{RREF}(A) = E_k \ldots E_1 A$ represents the sequence of elementary operations used to bring A to reduced row-echelon form. Note that all of the numbers $\det(E_j)$ are non-zero. Therefore we get:
  9. $\det(A) \neq 0$ if and only if A is invertible.
  10. $\det(AB) = \det(A)\det(B)$ for every $n \times n$ square matrices A, B — to see this, factor A into a product $E_k E_{k-1} \ldots E_2 E_1 \cdot \mathrm{RREF}(A)$ of elementary matrices followed by a reduced row-echelon form matrix. If A is invertible then the RREF is the identity matrix I and the claim follows from the previous property; otherwise AB cannot be invertible and therefore $\det(AB) = 0 = \det(A) = \det(A)\det(B)$.

- **Minors.** Let $A = (a_{ij})_{i,j=1}^n$ be a square matrix. For any $1 \le i, j \le n$, the determinant of the matrix obtained from A by deleting the ith row and jth column is called the (i,j)-minor of A, and we will sometimes denote it by M_{ij} or $M_{ij}(A)$.

- **Cofactors.** The cofactor is almost the same as the minor, but we add a sign of $+1$ or -1 to each minor by imposing the following "checkerboard" pattern of signs on top of the matrix:

$$\begin{pmatrix} + & - & + & - & \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \\ & \vdots & \vdots & & \ddots \end{pmatrix}$$

To make this precise, the (i, j)-cofactor of the matrix A is $C_{ij} = (-1)^{i+j} M_{ij}$ where M_{ij} is the (i, j)-minor.

- **Computing determinants by row expansion.** One reason why cofactors are useful is that they allow us to easily compute the determinant of a matrix using the method known as *row expansion*. In the simplest case of a "first row expansion", the formula looks like this:

  $$\det(A) = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n}$$

  Sketch of proof. Divide the $n!$ permutations in the sum defining the determinant into classes according to the value $j = \sigma(1)$. In each class we have $(n-1)!$ permutation-like functions (which are not quite permutations, since they map the numbers $\{2, \ldots, n\}$ to the numbers $\{1, 2, \ldots, \widehat{j}, \ldots, n\}$ — the "hat" notation $\widehat{j}$ means j is missing from the list). These can be identified with the permutations with the permutations involved in computing each of the cofactors C_{1j}.

- **Expansion by other rows.** For a general row i we can write the ith row expansion of the determinant as

  $$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}.$$

  To prove this, we can reduce it to the case of a first row expansion by swapping the ith row with the first row (this has the effect of only changing the sign of the determinant), then doing a first row expansion and noticing that all the $(1, j)$-cofactors in the new matrix (after swapping the rows) correspond to (i, j)-cofactors of the original matrix with their signs inverted to account for the fact that the values of 1 and j are transposed for each permutation appearing in the sum.

- **Column expansion.** Since we know that $\det(A) = \det(A^\top)$, by taking the transpose of the matrix and doing a row expansion, we see that we could have simply done an analogous *column expansion* to compute the determinant of the original matrix A, i.e., for each $1 \leq j \leq n$ we have the formula for a jth column expansion, given by

  $$\det(A) = \sum_{i=1}^{n} a_{ij}C_{ij}$$

  (note that here the summation is on the row index i, not the column index j).

Lecture 20 (2/26/20)

- **The adjoint matrix.** We can package all the cofactors C_{ij} into a matrix to get what is known as the *adjoint matrix* (or *adjugate matrix*) of A, denoted $\text{adj}(A)$; except that, for reasons that will be explained in the next theorem, to get the adjoint matrix one first writes down all the cofactors in a matrix and then *takes the transpose* **(do not forget this step!)**, i.e,

$$\text{adj}(A) = (C_{ji}(A))_{i,j=1}^{n}.$$

  The usefulness of this matrix is explained by the next result:

  Theorem. We have the equation of matrices:

$$A \cdot \text{adj}(A) = \text{adj}(A) \cdot A = \det(A) \cdot I$$

  (In words: when multiplying A by its adjoint, in either order, we get the identity matrix of order n multiplied by the scalar $\det(A)$.) In particular, if A is invertible then the adjoint matrix is related to the inverse matrix A^{-1} by

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

  Thus, the adjoint matrix gives a formula of sorts for the inverse matrix. For small matrices this is actually a reasonably efficient way of computing the inverse matrix, as long as one remembers the definition of the adjoint and is skilled in computing determinants.

- **Proof idea.** When we multiply A by the adjoint matrix $\text{adj}(A)$, the diagonal entries are seen to be precisely representations of $\det(A)$ as row expansions of the different rows of A. The off-diagonal entries are also determinants of various matrices which are obtained from A by replacing row i by row j for various different values of (non-equal) i, j. All such matrices have two identical rows and therefore their determinants are all equal to 0.

Lecture 20 (2/26/20) — eigenvectors, eigenvalues, and diagonalization of linear operators

- **Invariant subspaces.** Let $T : V \to V$ be a linear operator on a vector space V. A linear subspace $U \subseteq V$ is called an *invariant subspace* for T if for any $u \in U$, also $T(u) \in U$.
- **Examples.** The spaces null(T) and range(T) are both invariant subspaces.
- **Eigenvectors and eigenvalues.** An eigenvector corresponds to the simplest type of invariant subspace which is a 1-dimensional subspace.

  Definition. A vector $v \in V$ is called an *eigenvector* of a linear operator $T \in \mathcal{L}(V)$ if $v \neq 0$ and $T(v) = \lambda v$ for some scalar $\lambda \in \mathbb{F}$. The number λ is called the *eigenvalue* associated with the eigenvector v.
- **Properties of eigenvectors.**
  1. null(T) is the set of eigenvectors corresponding to the eigenvalue 0 (plus the zero vector, which is not considered an eigenvector).
  2. More generally, null($T-\lambda I$) is the set of eigenvectors corresponding to the eigenvalue λ. This is sometimes referred to as the *eigenspace* of T associated with λ.
  3. **Theorem.** The following conditions are equivalent:
     (a) λ is an eigenvalue.
     (b) null($T - \lambda I) \neq \{0\}$.
     (c) $T - \lambda I$ is not injective.
     (d) $T - \lambda I$ is not surjective.
     (e) $T - \lambda I$ is not invertible.
     (f) $\det(M - \lambda I) = 0$ where $M = M(T)^B_B$ is a matrix representing T in some basis B of V.

     Concerning the last condition on this list, note that if we think of the matrix M as fixed, then $p(\lambda) = \det(M - \lambda I)$ as a function of λ is a polynomial, called the *characteristic polynomial of* M, whose roots are the eigenvalues. This is why polynomial equations are so important in linear algebra.
- **Examples.**
  - $T = 0$ (the zero map): all nonzero vectors are eigenvectors associated with the eigenvalue 0.
  - $T = I$ (the identity map): all nonzero vectors are eigenvectors associated with the eigenvalue 1.
  - $P : \mathbb{R}^3 \to \mathbb{R}^3, P(x, y, z) = (x, y, 0)$ has eigenvalues 0 and 1. The eigenvectors for the eigenvalue 0 are of the form $(0, 0, z)$ for $z \in \mathbb{R}$. The eigenvectors for the eigenvalue 1 have the form $(x, y, 0)$ for $x, y \in \mathbb{R}$.
- **Eigenvectors associated with distinct eigenvalues are linearly independent.**

  Theorem. If $T \in \mathcal{L}(V)$ has eigenvectors $v_1, \ldots, v_n$ where v_i is associated with eigenvalue λ_i, i.e., $T(v_i) = \lambda_i v_i$, and $\lambda_1, \ldots, \lambda_n$ are distinct, then $v_1, \ldots, v_n$ are linearly independent.

  Proof. ...

Lecture 21 (2/28/20)

- **Corollary.** An operator $T \in \mathcal{L}(V)$ where $\dim(V) = n$ has at most n eigenvalues.

- **Diagonalization.** Assume that we can find a basis $B = \{v_1, \ldots, v_n\}$ consisting of eigenvectors of T, where each v_i is associated with an eigenvalue λ_i. If we represent the transformation T in the basis B we get a diagonal matrix:

$$M(T)_B^B = \begin{pmatrix} \lambda_1 & 0 & 0 & & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

  In that case we can say that we *diagonalized* T.

- **Example:** Let us try to diagonalize the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ -2 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5x + 3y \\ -2x \end{pmatrix}$$

  First, we find the eigenvalues. A number λ is an eigenvalue if and only if $\det(A - \lambda I) = 0$, so we compute

$$A - \lambda I = \begin{pmatrix} 5 - \lambda & 3 \\ -2 & -\lambda \end{pmatrix} = (5 - \lambda)(-\lambda) - 3(-2) = \lambda^2 - 5\lambda + 6$$

  This is 0 when $\lambda = 2, 3$, so these are the possible eigenvalues. Next, for each of the eigenvalues we find an associated eigenvector: For $\lambda_1 = 2$, we need to solve the equation $Av = 2v$ or $(A - 2I)v = 0$:

$$(A - 2I)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ -2 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  It is easy to find that $x = 1, y = -1$ is a solution (in fact the vector $(1, -1)$ spans the space of solutions, i.e., any other solution is a scalar multiple of this solution). This gives us our first eigenvector $v_1 = (1, -1)$.

  For the second eigenvalue $\lambda_2 = 3$, we solve

$$(A - 3I)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ -2 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  and we get a solution $x = 3, y = -2$, which gives us the second eigenvector $v_2 = (3, -2)$.

  To summarize, we found a basis $B = \{(1, -1), (3, -2)\}$ of eigenvectors of T, so when represented in the basis B, the representing matrix for T will be the diagonal matrix.

$$M(T)_B^B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

This is especially helpful, for example, if we wanted to compute a high power

$$T^n = T \circ T \circ \ldots \circ T$$

of the transformation: in this case, when representing everything in the basis B the problem would be reduced to computing a high power of the matrix $M(T)^B_B$, which is easy since this is a diagonal matrix. (In the next lecture we'll see how this can be used to find a formula for the famous Fibonacci numbers).

- **The characteristic polynomial.** As the example above illustrates, an important role in the discussion about eigenvectors and diagonalization is played by the function $\det(M - \lambda I)$, so we take a closer look at this function.

  Definition. Let $M = (m_{ij})_{i,j=1}^n$ be a square matrix of order n. The *characteristic polynomial of* M is the function

$$p(x) = p_M(x) = \det(xI - M) = \det \begin{pmatrix} x - m_{11} & -m_{12} & -m_{13} & \ldots & -m_{1n} \\ -m_{21} & x - m_{22} & -m_{23} & \ldots & -m_{2n} \\ -m_{31} & -m_{32} & x - m_{33} & \ldots & -m_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -m_{n1} & -m_{n2} & -m_{n3} & \ldots & x - m_{nn} \end{pmatrix}$$

- **Properties of the characteristic polynomial.**

  1. The characteristic polynomial is a polynomial of degree n in x. There are several ways to see this, by thinking of the determinant as a sum over permutations or by doing a first-row expansion.
  2. The leading coefficient of $p(x)$ (i.e., the coefficient of x^n) is 1, since that comes from expanding the product $(x - m_{11})(x - m_{22}) \ldots (x - m_{nn})$ associate with the identity permutation in the sum over the permutations that defines the determinant.
  3. The coefficient of x^{n-1} is $-(m_{11} + m_{22} + \ldots + m_{nn})$. The number $\sum_j m_{jj}$ is called the *trace* of the matrix M and is denoted $\text{tr}(M)$, so a shorter way to write this coefficient is $-\text{tr}(M)$.
  4. The constant coefficient is $p(0) = \det(-M) = (-1)^n \det(M)$.
  5. The roots of the polynomial $p(x)$ (i.e., solutions of the equation $p(x) = 0$) are exactly the eigenvalues of the matrix M.
  6. For an upper or lower triangular matrix $A = (a_{ij})_{i,j=1}^n$, the characteristic polynomial is $(x - a_{11})(x - a_{22}) \ldots (x - a_{nn})$, since the matrix $xI - A$ is also triangular and its determinant is the product of the diagonal entries.

  Using the characteristic polynomial we can prove the following fundamental result:

- **Theorem.** Any matrix M over the complex numbers has at least one eigenvalue.

  Proof. Over the complex numbers any polynomial has a root, by the Fundamental Theorem of Algebra.

Note that this fact is not true when working over the real numbers. For example, the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has no real eigenvalues. This difference is yet another reason why the complex numbers are considered so useful and important in mathematics.

- **Powers of matrices and the Fibonacci numbers.** Let us illustrate the power of matrix diagonalization by using these ideas to find a formula for the Fibonacci numbers[1]. These numbers, named after an Italian mathematician of the 13th century, are defined by the recursive equations

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}, \quad n \geq 2.$$

Here are the first few Fibonacci numbers:

n	0	1	2	3	4	5	6	7	8	9	10
F_n	0	1	1	2	3	5	8	13	21	34	55

We can represent the recursion in terms of matrix multiplication, as follows;

$$\begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix}$$

In other words, if we define vectors $v_n = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}$ for $n = 1, 2, 3, \ldots$ and a matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, then we have the equation

$$v_n = A v_{n-1}$$

which by iteration leads to

$$v_n = A^{n-1} v_1 = A^{n-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So, if we could easily compute powers of the matrix A we would be able to find a formula for v_n (and hence for F_n). The idea is to find the eigenvalues and eigenvectors of A. Following the usual method, we compute the characteristic polynomial and look for its roots:

$$p_A(x) = \det(xI - A) = \det \begin{pmatrix} x-1 & -1 \\ -1 & x \end{pmatrix} = (x-1)x - 1 = x^2 - x - 1 = 0.$$

The solutions of this equation are $\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.61803$ (a famous mathematical constant known as the golden ratio) and $\lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.61803$. For each of them it is not difficult to find an eigenvalue

$$e_1 = \begin{pmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix} \quad (\text{for } \lambda_1),$$

$$e_2 = \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix} \quad (\text{for } \lambda_2).$$

[1]The Fibonacci numbers are well-known for their amusing mathematical properties and for appearing in nature in connection with biological phenomena such as growth patterns of pine cones and sunflowers — see https://en.wikipedia.org/wiki/Fibonacci_number.

Finally, to compute $v_n = A^{n-1}v_1$, we expand v_1 in the basis $B = \{e_1, e_2\}$ of eigenvectors:

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{5}}e_1 - \frac{1}{\sqrt{5}}e_2$$

This allows us to write

$$\begin{aligned} v_n &= A^{n-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = A^{n-1}\left(\frac{1}{\sqrt{5}}e_1 - \frac{1}{\sqrt{5}}e_2\right) = \frac{1}{\sqrt{5}}(A^{n-1}e_1 - A^{n-1}e_2) \\ &= \frac{1}{\sqrt{5}}(\lambda_1^{n-1}e_1 - \lambda_2^{n-1}e_2) = \frac{1}{\sqrt{5}}\begin{pmatrix} \lambda_1^n - \lambda_2^n \\ \lambda_1^{n-1} - \lambda_2^{n-1} \end{pmatrix} \end{aligned}$$

So, we have derived the famous (and rather surprising) formula

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

Lecture 22 (3/2/20)

- **Final remarks on matrix diagonalization.** To conclude the discussion on diagonalization, I'll illustrate another way of thinking about diagonalization of matrices. If a square matrix A can be diagonalized, that means that we can find a basis $v_1, \ldots, v_n$ of eigenvectors, with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. If we create a matrix S whose columns are the eigenvectors, then the equations $Av_j = \lambda_j v_j$ translate into the matrix equation

$$AS = SD$$

where D is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

This can be written equivalently as

$$A = SDS^{-1} \qquad \Longleftrightarrow \qquad D = S^{-1}AS$$

(recall that the columns of S form a basis, which is one of the equivalent conditions for a matrix to be invertible; so we can be sure that S^{-1} exists). Thus, "diagonalizing a matrix" can refer to the process of finding a basis of eigenvectors and their associated eigenvalues, or, equivalently, as the process of finding a pair of matrices S and D, where S is invertible and D is a diagonal matrix, such that the matrix equations $A = SDS^{-1}$, $D = S^{-1}AS$ hold. The implication also goes in the opposite direction: if we found such matrices S, D, then working backwards through this reasoning we see that the columns of the matrix, regarded as column vectors $v_1, \ldots, v_n$, satisfy the eigenvector equation

$$Av_j = \lambda_j v_j,$$

where $\lambda_1, \ldots, \lambda_n$ are the diagonal entries of D.

As a final comment, the representation of A as SDS^{-1} is also convenient for computing the matrix powers A^n of A:

$$\begin{aligned} A^n &= \overbrace{(SDS^{-1})(SDS^{-1})\cdots(SDS^{-1})}^{n \text{ times}} = SD(S^{-1}S)D(S^{-1}S)D\cdots(S^{-1}S)DS^{-1} \\ &= SD^nS^{-1} = S\begin{pmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_n^n \end{pmatrix}, S^{-1} \end{aligned}$$

where in the last step we exploited the useful property that when two diagonal matrices are multiplied, the result is a diagonal matrix with the entries being the products of the respective diagonal entries of the two matrices.

As an exercise, I suggest trying to use this alternative formalism for matrix diagonalization to derive the formula for Fibonacci numbers we found earlier.

- **Inner product spaces.** We now begin the new topic of inner product spaces.

  An inner product space is a vector space with an additional "geometric" structure that enables us to measure lengths of vectors, angles between vectors, and a notion of orthogonality of vectors.

  Definition. Let V be a vector space over the field $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$). An *inner product* on V is a function that takes two vectors u, v and returns a scalar $\langle u, v\rangle$, that satisfies the following properties:

  1. Linearity in the first argument: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ and $\langle au, v\rangle = a\langle u, v\rangle$ for all $u, v, w \in V$ and $a \in \mathbb{F}$.
  2. Positivity: $\langle u, u\rangle \geq 0$ (and in particular, is a real number) for all $v \in V$.
  3. Positive definiteness: $\langle v, v\rangle = 0$ if and only if $v = 0$.
  4. (Conjugate)-symmetry: if $\mathbb{F} = \mathbb{R}$ then $\langle u, v\rangle = \langle v, u\rangle$. If $\mathbb{F} = \mathbb{C}$ then $\langle u, v\rangle = \overline{\langle v, u\rangle}$ (recall that $\overline{z}$ denotes the complex conjugate of a complex number z).
  5. (Conjugate)-linearity in the second argument (follows from 1 and 4 above): if $\mathbb{F} = \mathbb{R}$ then $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ and $\langle u, av\rangle = a\langle u, v\rangle$ for all $u, v, w \in V$ and $a \in \mathbb{R}$. If $\mathbb{F} = \mathbb{C}$ then $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ and $\langle u, av\rangle = \overline{a}\langle u, v\rangle$ for all $u, v, w \in V$ and $a \in \mathbb{C}$.

  A vector space V equipped with an inner product is called an *inner product space.*

Lecture 23 (3/4/20)

- **Examples of inner product spaces.**

  1. On $\mathbb{R}^n$, the usual dot product defined by

  $$(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = x_1 y_1 + \dots + x_n y_n$$

  is an inner product.

  2. On $\mathbb{C}^n$, the product defined by

  $$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = x_1 \overline{y_1} + \dots + x_n \overline{y_n} = \sum_{j=1}^{n} x_j \overline{y_j}$$

  is an inner product.

  3. On the space P of polynomials (say with complex coefficients), we can define an inner product

  $$\langle f, g \rangle_P = \int_0^1 f(x) \overline{g(x)} \, dx.$$

  4. On each of the above spaces we can define a non-standard inner product. For example, on $\mathbb{R}^2$ we can define

  $$\langle (x_1, y_1), (x_2, y_2) \rangle = x_1 x_2 + 3 y_1 y_2.$$

  It is not difficult to check that this function is an inner product.

- **Orthogonality.** Two vectors u, v in an inner product space V are called *orthogonal* if $\langle u, v \rangle = 0$. This is denoted by $u \perp v$.

- **Norm.** The *norm* (a.k.a. length) of a vector v in an inner product space V is defined as

  $$\|v\| = \sqrt{\langle v, v \rangle}.$$

  For example, for the standard inner product in $\mathbb{R}^n$ we have

  $$\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \dots + x_n^2}.$$

- **Theorem.** The norm satisfies the following properties:

  1. $\|v\| \geq 0$ for any v, and $\|v\| = 0$ if and only if $v = 0$.
  2. $\|\alpha v\| = |\alpha| \|v\|$ for any $v \in V$, $\alpha \in \mathbb{F}$.
  3. The triangle inequality: $\|u + v\| \leq |u\| + \|v\|$ for any $u, v \in V$.

  Proof. The first two claims are immediate from the definition. The third requires first proving the Cauchy-Schwartz inequality, which we'll do shortly.

- **Theorem.** (Pythagorean theorem.) If $u \perp v$ then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

  Proof. For general u, v we have

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle = \langle u, u+v\rangle + \langle v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u,v\rangle + \langle v,u\rangle = \|u\|^2 + \|v\|^2 + \langle u,v\rangle + \overline{\langle u,v\rangle} \\ &= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\langle u,v\rangle.\end{aligned}$$

  If they are orthogonal then $\langle u,v\rangle = 0$ so the Pythagorean theorem holds.

- **Corollary.** If $v_1, \ldots, v_k$ are pairwise orthogonal (i.e., each two of them are orthogonal) and $u = a_1 v_1 + \ldots + a_k v_k$ is a linear combination of $u_1, \ldots, u_k$, then

$$\|u\| = \sqrt{a_1^2 \|v_1\|^2 + \ldots + a_k^2 \|v_k\|^2}.$$

  Note that if also $\|v_1\| = \ldots = \|v_k\| = 1$ then we get the simpler formula

$$\|u\| = \sqrt{a_1^2 + \ldots + a_k^2}.$$

- **Orthogonal decomposition.** Fix a vector $v \neq 0$ in an inner product space V. We claim that any vector u can be decomposed as a sum $u = u_1 + u_2$ where u_1 is parallel to v (i.e., $u_1 = av$ for some scalar a) and $u_2 \perp v$. To find this decomposition, write $u_2 = u - av$. The fact that $u_2 \perp v$ leads to the equation

$$0 = \langle u - av, v\rangle = \langle u,v\rangle - a\langle v,v\rangle = \langle u,v\rangle - a\|v\|^2.$$

  Solving for the unknown a we get

$$a = \frac{\langle u,v\rangle}{\|v\|^2},$$

  so the orthogonal decomposition is

$$u = \frac{\langle u,v\rangle}{\|v\|^2} v + \left(u - \frac{\langle u,v\rangle}{\|v\|^2} v\right).$$

  The vector $u_1 = av$ is called the *orthogonal projection of u in the direction of v*. Note that $u_2 = 0$ if and only if $u = u_1 = av$, i.e., u is parallel to v.

- **Angles between vectors.** Let V be an inner product space *over the real numbers*. The formula

$$\|u-v\|^2 = \|u\|^2 + |v\|^2 - 2\langle u,v\rangle$$

  is reminiscent of the law of cosines from plane geometry, which says that

$$\|u-v\|^2 = \|u\|^2 + |v\|^2 - 2\|u\| \cdot \|v\| \cos\theta_{u,v},$$

  where $\theta_{u,v}$ denotes the angle subtended between the two vectors u and v. The following definition therefore seems like a natural generalization of the concept of angles to inner product spaces:

Definition. The angle $0 \leq \theta_{u,v} \leq \pi$ between nonzero vectors u, v in a general inner product space is defined by the formula

$$\cos\theta_{u,v} = \frac{\langle u, v\rangle}{\|u\| \cdot \|v\|}$$

Note that in order for this to make sense, we need to convince ourselves that the number on the right-hand side is between -1 and 1. A famous inequality (which is also true in complex inner product spaces, where angles are not defined) will come to our rescue:

- **The Cauchy-Schwartz inequality.** For any $u, v \in V$, we have

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|,$$

The two sides are equal if and only if u, v are linearly dependent, i.e., one of them is a scalar multiple of the other.

Proof. Assume $v \neq 0$, since if $v = 0$ both sides are 0 so there's nothing to prove. Let $u = u_1 + u_2$ be the orthogonal decomposition of u with respect to v as described above. Applying the Pythagorean theorem, we have

$$\|u\|^2 = \|u_1\|^2 + \|u_2\|^2 = \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 + \|u_2\|^2 \geq \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 = \frac{\|v\|^2 |\langle u, v\rangle|^2}{\|v\|^4} = \frac{|\langle u, v\rangle|^2}{\|v\|^2}.$$

Multiplying the extreme sides of this inequality by $\|v\|^2$ and taking square roots gives the claim. The claim about when we have inequality follows from the remark above about u_2 being 0 if and only if u is parallel to v.

Note that the Cauchy-Schwartz inequality can be rewritten as

$$-1 \leq \frac{\langle u, v\rangle}{\|u\| \cdot \|v\|} \leq 1$$

so the definition of the angle $\theta_{u,v}$ indeed makes sense. The case when $\theta_{u,v} = \pi/2$ corresponds to orthogonal vectors $u \perp v$, and the case when the vectors are parallel $u = av$ gives an angle $\theta_{u,v} = 0$ if $a > 0$ or $\theta_{u,v} = \pi$ if $a < 0$.

Lecture 24 (3/6/20)

- **The triangle inequality.** $\|u+v\| \le \|u\| + \|v\|$. Equality holds if and only if $u = \alpha v$ or $v = \alpha u$ for some scalar $\alpha \ge 0$.

  Proof. Write the inequality

  $$\begin{aligned}\|u+v\|^2 &= \|u\|^2 + |v\|^2 + 2\|u\| \cdot \|v\| \cos\theta_{u,v} \\ &\le \|u\|^2 + |v\|^2 + 2\|u\| \cdot \|v\| = (\|u\| + \|v\|)^2\end{aligned}$$

  and take square roots. The verification of the condition for equality is left as an exercise to the reader.

- **The parallelogram law. Theorem.** For any $u, v \in V$ we have

  $$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right).$$

  Note that the name comes from interpreting the quantities $\|u+v\|$ and $\|u-v\|$ as the side lengths of a parallelogram with one corner in the origin and two of whose sides are represented by the vectors u and v.

  Proof. ...

- **Orthogonal system and basis.** A set of nonzero vectors $\{e_1, \ldots, e_n\}$ of an inner product space V is called an *orthogonal system* if $e_i \perp e_j$ for any $i \ne j$. It is called an *orthogonal basis* if it is a basis and an orthogonal system.

- **Orthonormal system and basis.** A set of nonzero vectors $\{e_1, \ldots, e_n\}$ of an inner product space V is called an *orthonormal system* if it is an orthogonal system and $\|e_i\| = 1$ for $i = 1, \ldots, n$. It is called an *orthonormal basis* if it is a basis and an orthonormal system.

  The two conditions for orthonormality $e_i \perp e_j$ and $\|e_i\| = 1$ can be unified into one equation

  $$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \ne j. \end{cases}$$

- **The Kronecker delta.** The quantity on the right-hand side of the above equation is often denoted by the abbreviated notation $\delta_{i,j}$ and referred to as the *Kronecker delta* function (of i and j). (You do not need to remember this term.)

- **Properties of orthogonal/orthonormal bases.**

  1. If $\{e_1, \ldots, e_k\}$ is an orthogonal system and $u = a_1 e_1 + \ldots + e_k$ zis a linear combination of $e_1, \ldots, e_k$, then we can recover the coefficients $a_1, \ldots, a_k$ by taking an inner product of u with each of the vectors $e_1, \ldots, e_k$:

     $$\begin{aligned}\langle u, e_j \rangle &= \langle a_1 e_1 + \ldots + e_k, e_j \rangle \\ &= a_1 \langle e_1, e_j \rangle + a_2 \langle e_2, e_j \rangle + \ldots + a_j \langle e_j, e_j \rangle + \ldots + a_k \langle e_k, e_j \rangle = a_j \|e_j\|^2,\end{aligned}$$

     so we get that $a_j = \frac{\langle u, e_j \rangle}{\|e_j\|^2}$. In the case of an orthonormal system we get the even simpler formula $a_j = \langle u, e_j \rangle$. (Note that this is exactly the scalar that appeared in the expression we called earlier the *projection of u in the direction e_j.*) This is one of the reasons orthonormal bases are useful: to find the coordinates of a vector u with respect to an orthonormal basis, we don't need to solve systems of linear equations, but can instead compute the coefficients directly using the inner product.

2. As an immediate corollary of the property mentioned above, we get:

   Proposition. Every orthogonal system is linearly independent.

   Proof. If $0 = a_1 e_1 + \ldots + a_k e_k$ then $a_j = \frac{\langle 0, e_j \rangle}{\|e_j\|^2} = 0$.

3. It follows from the above result that any orthonormal system with $\dim(V)$ vectors is a basis (since it is a linearly independent set of maximal size).

- **Examples.**
  1. The standard basis in $\mathbb{R}^n$ is orthonormal.
  2. The basis $\{(1,1),(1,-1)\}$ in $\mathbb{R}^2$ is orthogonal.
  3. The basis $\left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\right\}$ in $\mathbb{R}^2$ is orthonormal.

- **A surprisingly non-obvious result: Theorem.** Every finite dimensional inner product space has an orthonormal basis.

  To prove the theorem, we need an interesting algorithm known under the (slightly bombastic) name:

- **The Gram-Schmidt orthogonalization procedure.** Given linearly independent vectors $v_1, \ldots, v_n$, we construct a new set of vectors $u_1, \ldots, u_n$ that satisfy

  1. For each $1 \leq k \leq n$, $\text{span}(u_1, \ldots, u_k) = \text{span}(v_1, \ldots, v_k)$.
  2. $u_1, \ldots, u_n$ are orthogonal.

  The idea is to define each u_k by taking v_k and subtracting from it some linear combination of $v_1, \ldots, v_{k-1}$ (or equivalently, of $u_1, \ldots, u_k$) to cause the resulting vector to be orthogonal to the ones already defined. More precisely, we define

$$
\begin{aligned}
u_1 &= v_1, \\
u_2 &= v_2 - \frac{\langle v_2, u_1 \rangle}{\|u_1\|^2} u_1 && (= v_2 \text{ minus its projection in the direction } u_1), \\
u_3 &= v_3 - \frac{\langle v_3, u_1 \rangle}{\|u_1\|^2} u_1 - \frac{\langle v_3, u_2 \rangle}{\|u_2\|^2} u_2 && (= v_3 \text{ minus its projections in the directions } u_1, u_2), \\
&\vdots \\
u_n &= v_n - \sum_{j=1}^{n-1} \frac{\langle v_n, u_j \rangle}{\|u_j\|^2} u_j && (= v_n \text{ minus its projections in the directions } u_1, \ldots, u_{n-1}).
\end{aligned}
$$

  It is easy to verify that $\langle u_k, u_j \rangle = 0$ for all $j < k$. To check the claim about the span being preserved, note that

$$\text{span}(u_1, \ldots, u_k) \subset \text{span}(v_1, \ldots, v_k)$$

since each u_k is defined as a linear combination of v_k and the previous u_j's (which are by induction linear combinations of $v_1, \ldots, v_{k-1}$. Conversely, given $u_1, \ldots, u_k$ we can recover v_k as a linear combination of $u_1, \ldots, u_k$ by moving the linear combination of $u_1, \ldots, u_{k-1}$ to the other side of the equation in the definition of u_k; then proceeding by induction we can similarly represent each v_k as a linear combination of $u_1, \ldots, u_k$, proving that

$$\text{span}(v_1, \ldots, v_k) \subset \text{span}(u_1, \ldots, u_k).$$

- **Normalizing the vectors.** As a final step, we can replace $u_1, \ldots, u_n$ with new vectors $e_1, \ldots, e_n$ which give an orthonormal basis, by defining

$$e_k = \frac{u_k}{\|u_k\|}.$$

- **Example.** Take $v_1 = (1, 1, 0), v_2 = (2, 1, 1)$ in $\mathbb{R}^3$. Applying the Gram-Schmidt procedure we get: ...

- **Proof that any IPS has an orthonormal basis.** Start with an arbitrary basis (not necessarily orthogonal) of the space, and apply the Gram-Schmidt procedure to get an orthonormal system of the same size as the original basis, which therefore must be a basis.

Lecture 25 (3/9/20)

- **Orthogonal complement.** Let V be a finite-dimensional inner product space, and let $U \subset V$ be a subspace of V. The *orthogonal complement* of U is the subspace $U^\perp \subset V$, defined by
$$U^\perp = \{v \in V \,:\, v \perp u \text{ for all } u \in U\}.$$
- **Properties of $U^\perp$:**
  1. $U^\perp$ is a linear subspace of V.
     Proof. ...
  2. $\{0\}^\perp = V$, $V^\perp = \{0\}$.
  3. $U \cap U^\perp = \{0\}$.
     Proof. if $v \in U \cap U^\perp$ then v is orthogonal to itself, so $\|v\|^2 = \langle v, v\rangle = 0$ and therefore $v = 0$.
  4. $V = U + U^\perp$.
     Proof. if $e_1, \ldots, e_k$ is an orthonormal basis for U, we define
$$\begin{aligned} v_1 &= \langle v, e_1\rangle e_1 + \ldots + \langle v, e_k\rangle e_k, \\ v_2 &= v - v_1. \end{aligned}$$
     We have $v = v_1 + v_2$, and it is easy to check that $v_1 \in U, v_2 \in U^\perp$.
  5. Combining the last two properties we get: **Theorem.** $V = U \oplus U^\perp$.
  6. $(U^\perp)^\perp = U$.
     Proof. Any vector $u \in U$ is orthogonal to any vector $w \in U^\perp$. This proves that $U \subset (U^\perp)^\perp$. Conversely, if $v \in (U^\perp)^\perp$, write $v = u + w$ where $u \in U, w \in U^\perp$. We have
$$\|w\|^2 = \langle w, w\rangle = \langle w, w\rangle + 0 = \langle w, w\rangle + \langle u, w\rangle = \langle u + w, w\rangle = \langle v, w\rangle = 0$$
     (the last equality holding since $w \in U^\perp$ and $v \in (U^\perp)^\perp$). Thus $w = 0$ and $v = u \in U$. This shows that $U^\perp \subset (U^\perp)^\perp$ and finishes the proof.
     Alternative proof of the second part. By the direct sum decomposition (property 5) above,
$$\dim(V) = \dim(U) + \dim(U^\perp),$$
     and similarly (applying the same argument to $U^\perp$ instead of U),
$$\dim(V) = \dim(U^\perp) + \dim((U^\perp)^\perp).$$
     Comparing the two equations we conclude that $\dim(U) = \dim((U^\perp)^\perp)$. We also know (from the easy part of the proof above) that $U \subset (U^\perp)^\perp$, so the two spaces have to be equal.
- **Orthogonal projections.** If $v \in V = U \oplus U^\perp$, we can write v in a unique way as a sum $v = u + w$ where $u \in U$ and $w \in U^\perp$. The vector u is called the *orthogonal projection of v onto U* and is denoted $u = P_U(v)$. This generalizes the idea of the orthogonal decomposition of a vector u in the direction of a vector v discussed in a previous lecture.

  Note that $P_U : v \mapsto P_U(v)$ is a function from V to V. It is not hard to check that it is a linear transformation. We call it the *orthogonal projection operator associated with U*.

- **Properties of P_U.**
  1. $\mathrm{range}(P_U) = U$.
  2. $\mathrm{null}(P_U) = U^\perp$.
  3. For any $v \in V$, we have $v = P_U(v) + P_{U^\perp}(v)$. Equivalently, this can be written as the operator identity
  $$P_U + P_{U^\perp} = I$$
  4. **Theorem.** $P_U(v)$ is the vector in U that is closest to v. More precisely, for every $u \in U$ we have
  $$\|v - u\| \geq \|v - P_U(v)\|,$$
  with equality holding if and only if $u = P_U(v)$.
  Proof.
  $$\begin{aligned} \|v - u\|^2 &= \|(v - P_U(v)) + (P_U(v) - u)\|^2 = \|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 \\ &\geq \|v - P_U(v)\|^2 \end{aligned}$$
  (note that $v - P_U(v) \in U^\perp$ and $P_U(v) - u \in U$, which is why the second equality follows from the Pythagorean theorem).

- **Example.** In $\mathbb{R}^3$, let us find the distance between the vector $v = (1, 2, 3)$ and the plane U where $U = W^\perp, W = \mathrm{span}\{(1, 1, 1)\}$, and the orthogonal projection $P_U(v)$. In this case we have
$$\begin{aligned} P_U(v) = v - P_W(v) &= (1, 2, 3) - \frac{\langle (1, 2, 3), (1, 1, 1) \rangle}{\|(1, 1, 1)\|^2}(1, 1, 1) \\ &= (1, 2, 3) - \frac{6}{3}(1, 1, 1) = (1, 2, 3) - (2, 2, 2) = (-1, 0, 1). \end{aligned}$$
Therefore the distance between v and U is
$$\|v - P_U(v)\| = \|(2, 2, 2)\| = \sqrt{12}.$$
Note that it was easier to compute $P_{U^\perp}(v)$ than $P_U(v)$ since U is 2-dimensional and its orthogonal complement is 1-dimensional. In general, to compute $P_U(v)$ one has to first find an orthogonormal (or orthogonal) basis for the space.

Lecture 26 (3/11/20)

[**Note.** The following lecture corresponds to parts of sections 11.1–11.3 in the textbook, but covers only part of the material from those sections and follows a simplified approach that gives a proof of the spectral theorem only for the case of self-adjoint operators.]

- **Adjoint operators.** Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional inner product space V. We associate to T a new operator denoted T^* that will be called the *adjoint operator* of T. The defining equation of T^* is

$$\langle Tv, w\rangle = \langle v, T^* w\rangle$$

Theorem. There is a unique operator satisfying this equation for all $v, w \in V$.

Proof. For simplicity, assume that $V = \mathbf{C}^n$ equipped with the standard inner product, and $T(v) = Av$ where A is an $n \times n$ square matrix of complex numbers. The proof in the general case is similar. Let $e_1, \ldots, e_n$ be the standard orthonormal basis of $\mathbf{C}^n$. Let $S(v) = Bv$ be a linear operator, and assume that the equation

$$\langle Tv, w\rangle = \langle v, Sw\rangle = \overline{\langle Sw, v\rangle}$$

holds for all $v, w \in \mathbf{C}^n$. Applying it with $v = e_j$ and $w = e_i$ gives that the entry in the ith row, jth column of the matrix A must be equal to the complex conjugate of the entry in the jth row, ith column of the matrix B; i.e.,

$$b_{ji} = \overline{a_{ij}}, \qquad i, j = 1, \ldots, n.$$

This proves uniqueness: the matrix B is determined uniquely by the condition. The existence also follows, since we can define the matrix B according to the same rule and get an operator for which the defining equation of the adjoint operator holds for vectors v, w belonging to the standard basis, and hence (by linearity in the first argument and conjugate-linearity in the second argument) it is easy to check that the same identity holds for general vectors.

- The above definition also gives rise to an analogous concept in the language of matrices:

Definition. If $A = (a_{ij})_{i,j=1}^n$ is a square matrix, the *conjugate transpose matrix* (a.k.a. *Hermitian conjugate*) of A is the matrix $A^* = (b_{ij})_{i,j=1}^n$ whose entries are given by $b_{ji} = \overline{a_{ij}}$. Equivalently, we can write

$$A_* = \overline{A^\top},$$

where $A^\top$ is the transpose matrix and the "bar" $\overline{\cdots}$ indicates that we take the complex conjugate of each entry. **In the case of matrices of real numbers, the concept of the conjugate transpose matrix is identical to taking the transpose.**

- **Note.** The conjugate transpose matrix is also sometimes called the adjoint matrix. This creates a conflict with our earlier definition of the entirely different concept of the adjoint matrix (which is also known as the adjugate matrix). Try not to get confused by this sometimes inconsistent terminology...

- **Self-adjoint operators and matrices.** An operator $T \in \mathcal{L}(V)$ is called *self-adjoint* if $T = T^*$. A square matrix A is called *self-adjoint* if $A = A^*$. **For matrices of real numbers, a self-adjoint matrix is the same as a symmetric matrix, i.e., a matrix satisfying $A = A^\top$.**

- **Properties of adjoint operators.** The following properties are all easy to verify:

  1. $(S+T)^* = S^* + T^*$.
  2. $(aT)^* = \overline{a}T^*$ for any $a \in \mathbb{C}$.
  3. $(T^*)^* = T$.
  4. $I^* = I$ (the identity operator is self-adjoint).
  5. $(ST)^* = T^*S^*$.

- **The eigenvalues of a self-adjoint operator are all real.** Let $\lambda \in \mathbb{C}$ be an eigenvalue of a self-adjoint operator T. That means that there is a nonzero eigenvector $v \in V$ such that the equation $T(v) = \lambda v$ holds. In this case, we can write

  $$\lambda\langle v,v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v,v\rangle.$$

  It follows that $\lambda = \overline{\lambda}$, i.e., λ is a real number! We have proved:

  Theorem. All the eigenvalues of a self-adjoint operator are real.

- Self-adjoint operators are extremely interesting and useful. The following two theorems explain why.

  Theorem. If u, v are eigenvectors of a self-adjoint operator T corresponding to two different eigenvalues $\lambda_1 \neq \lambda_2$, then $u \perp v$.

  Proof.
  $$\lambda_1\langle u,v\rangle = \langle \lambda_1 u, v\rangle = \langle Tu, v\rangle = \langle u, Tv\rangle = \langle u, \lambda_2 v\rangle = \lambda_2\langle u,v\rangle.$$

  Since at least one of λ_1, λ_2 is not zero, the above equation implies that $\langle u,v\rangle = 0$.

  Theorem. (The spectral theorem for self-adjoint operators). Any self-adjoint operator is diagonalizable. Furthermore, we can choose the basis of eigenvectors to be an orthonormal basis.

  To prove the theorem, we need a lemma:

  Lemma. Let T be a self-adjoint operator. If $U \subseteq V$ is a linear subspace of V that is an invariant subspace under T (i.e., if $u \in U$ then $T(u) \in U$), then its orthogonal complement $U^\perp$ is also invariant under T.

  Proof. If $w \in U^\perp$ then $w \perp u$ for any $u \in U$. Therefore, for any $u \in U$ we have

  $$\langle T(w), u\rangle = \langle w, Tu\rangle = 0$$

  since $T(u) \in U$ by the fact that U is T-invariant.

  Proof of the spectral theorem. Use induction on $\dim V$. If $\dim V = 1$ the claim is obvious (an operator is just a number). Assume we proved it for spaces of dimension $n-1$ and let V be n-dimensional. By the fundamental theorem of algebra, the characteristic polynomial $P_T(x)$ has a root λ_n, which is an eigenvalue. Let v_n be an associated eigenvector. Consider now the subspace

$$U = \text{span}\{v_n\}^\perp,$$

the subspace of vectors which are orthogonal to v_n. Since v_n is an eigenvector, span(v_n) is an invariant subspace, and therefore (by the lemma), so is its orthogonal complement U. It follows that we can restrict the operator T to the subspace U (this restricted operator is usually denoted $T_{|U}$). The restricted operator $T_{|U}$ is a self-adjoint operator on U, which is an $(n-1)$-dimensional space. By the induction hypothesis, it has an orthonormal basis $v_1, \ldots, v_{n-1}$ of eigenvectors. Adding the vector v_n (which is orthogonal to $v_1, \ldots, v_n$ gives an orthonormal system of n eigenvectors (which must therefore be a basis).

- **Example.** Let $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}$. Since A is a symmetric real matrix, we immediately know it is diagonalizable and has real eigenvalues. Let us diagonalize it. The characteristic polynomial is

$$p_A(x) = \det(xI - A) = \det \begin{pmatrix} x-1 & -1 & 0 \\ -1 & x & 1 \\ 0 & 1 & x-1 \end{pmatrix} = x^3 - 2x^2 - x + 2 = (x-1)(x+1)(x-2),$$

so the eigenvalues are $\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 2$. With a short computation we find corresponding eigenvectors

$$v_1 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix},$$

$$v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

$$v_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Note that these vectors are orthogonal (as predicted by the theorem we proved) but not orthonormal. By normalizing them, we get a basis of orthonormal vectors:

$$u_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix},$$

$$u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

$$u_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$