Welcome!

MTH 222

Linear Algebra

Your professor

(a little outdated)

A little bit about me:

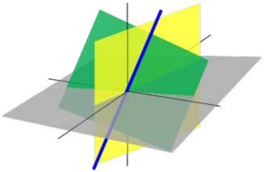
- Went to undergrad at Indiana (Go Hoosiers!)
- Taught high school math in Baltimore, MD
- Earned my PhD from the University of Wisconsin-Milwaukee
- Did postdocs at the University of California, San Diego and Wake Forest University
- In my fourth year at Miami (halfway to tenure!)
- My research area is noncommutative algebra – more on this later in the course

Your professor

When I'm not doing math, or other academia-related stuff, I'm probably either

- spending time with my wife and 5-year-old daughter,
- playing with my pug Rocky,
- listening to music (Tool, Pearl Jam, Radiohead, too many to name),
- playing video games (this summer it was Zelda: BOTW),
- or exercising (I go early if you want to catch me at the rec).

What is linear algebra?



If you've taken any math before, you've probably thought a good bit about 1-dimensional space and 2-dimensional space. If you're one of the many students who has taken multivariable calculus, then you've probably thought a good bit about 3-dimensional space.

In linear algebra, we will study n-dimensional space, but we will only worry about "flat" things. This theory is still extremely useful, because even if things aren't flat, if we look at them closely enough, they can look flat. (This is the essential idea in calculus.)

The main focus of the class will be getting a coherent and somewhat abstract theory of systems of linear equations. The main objects we will talk about are vectors, matrices, systems of equations, linear transformations and vector spaces.

Why is Linear Algebra useful?

Linear algebra is a very important and powerful subject. Here are some examples.

- In physics and engineering, the behavior of electrical circuits is governed by differential equations, and finding solutions of these equations are essentially a problem in linear algebra.
- Linear algebra plays an important role in understanding the behavior of the US economy.
- Linear algebra can be used to rank sports teams, or predict the outcomes of sporting events. Linear algebra plays a big role in sabermetrics, which is used to analyze baseball.
- The theory of eigenvalues and eigenvectors is the main theoretical tool that makes Google as awesome as it is.

What else will I get out of this course?

In any math class, you will probably learn to be more detail-oriented. You will begin to learn to write proofs and think carefully and deeply about mathematics, likely to a greater extent than you have in previous classes. These will help you develop the gray matter in your head, and doing that will make it easier for you to do anything in the future.

The Syllabus

You should read the entire syllabus and email me with any questions. Here are some of the highlights.

Class Meetings:

Section A: MWF 10:05am-11:00am

Section B: MWF 11:40am - 12:35pm.

Our class will meet synchronously on Zoom. Links are posted on Canvas.

Office Hours: Tuesdays 10:00am-11:00am on Zoom.

The Syllabus

Class Meetings (contd):

- Mondays and Wednesdays are scheduled as lecture days. Recordings of these lectures will be posted on Canvas.
- Fridays are reserved for problem sessions. The problem session worksheets are posted by the beginning of the week. Try to work these problems before class! Recordings of these sessions will be posted on Canvas as well.
- Attendance is not required but **highly encouraged**. Being able to ask questions and participate in the discussion is very useful in developing your skills. Students participating in the class asynchronously are assuming a great deal of responsibility for self-learning.

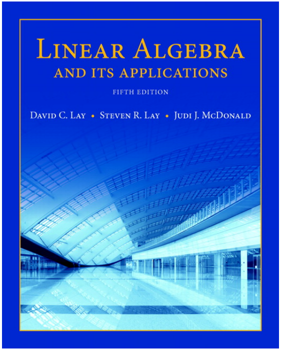
The Syllabus

Grading:

- Concept check quizzes (15%). These are weekly quizzes administered on Canvas. They are open notes but are to be completed individually.
- Two midterm exams (25% each) and a final exam (35%). These are administered online using Proctorio. See the Canvas site for more information. It is your responsibility to be ready for the first exam.

Grading scale is on the syllabus.

The Syllabus



Textbook and homework:

- Text is *Linear Algebra and its Applications (5ed)* by David Lay, Steven Lay, and Judi McDonald. The 4ed is fine. Older editions probably ok too but I haven't reviewed those.
- Homework is posted weekly on Canvas. These are not graded but you need to do them to succeed on quizzes and exams. Solutions will be posted.
- Extra problems are assigned from the text. No solutions will be posted but can discuss in office hours or during problem sessions.

The Syllabus

Other stuff:

- Check Canvas regularly!
- Stay on top of work and don't get behind. If you're having trouble, ask for help.
- I take academic honesty very seriously. See the syllabus for more specifics.
- More on in-person instruction if we ever get to that.
- Have a great semester!

Chapter 1: Linear Equations
§1.1 Systems of linear equations
§1.2 Row reduction and echelon forms

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- State the formal definition of a system of linear equations and discuss the concepts of *existence* and *uniqueness* of solutions.
- Formalize the strategies used in solving systems in terms of elementary row operations on matrices.
- Define the (reduced) row echelon form a matrix and discuss the algorithm for putting a matrix in echelon form.
- Consider different ways to express the solution of a system of equations.

We'll start by looking at some simple systems of linear equations.

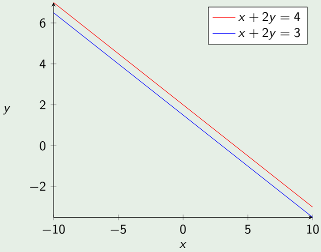
Systems of equations - First examples

Example 1

Consider the following system of equations in two unknowns x and y:

$$\begin{array}{rrcl} & x+2y & = & 4 \\ - & x+2y & = & 3 \\ \hline & 0 & = & 1 \end{array}$$

This equation has no solutions. Geometrically, this system corresponds to two parallel lines.



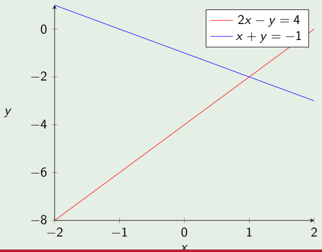
Systems of equations - First examples

Example 2

Consider the following system of equations in two unknowns x and y:

$$\begin{array}{rrcr} & 2x - y & = & 4 \\ + & x + y & = & -1 \\ \hline & 3x & = & 3 \\ & x & = & 1 \end{array}$$

This equation has exactly one solution, $(1, -2)$. Geometrically, this system corresponds to two lines that intersect in exactly one point.



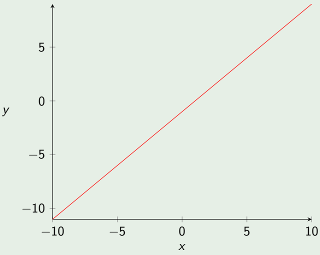
Systems of equations - First examples

Example 3

Consider the following system of equations in two unknowns x and y:

$$\begin{array}{rrcr} & 2(x-y) & = & 1 \\ + & -2x+2y & = & -2 \\ \hline & 0 & = & 0 \end{array}$$

This equation has infinitely many solutions. These two equations represent the same line and so any point on that line is a solution to both.



Systems of linear equations

Definition 4

The equation $a_1x_1 + \cdots + a_nx_n = b$ is called *linear* with variables x_i, coefficients a_i (real or complex), and constant b.

A *solution* to a linear equation is a set $(s_1, \ldots, s_n)$ such that substituting the s_i for x_i in the left-hand side produces a true statement.

Geometrically, a linear equation corresponds to a line in n-dimensional (real or complex) space.

Systems of linear equations

Definition 5

A *system of linear equations* is a set of linear equations in the same variables and a *solution* to the system is a common solution to all the equations in the system.

- A system is *consistent* if it has at least one solution.
- A system is *inconsistent* if it has no solution.
- Two systems with the same solution set are said to be *equivalent*.

When confronted with a system, we are most often interested in the following two questions:

- (Existence) Is the system consistent?
- (Uniqueness) If the system consistent, is there a *unique* solution?

Translating to matrices

Solving a system of equation can be done much more quickly and efficiently using matrix techniques.

Definition 6

An $m \times n$ *matrix* M is a rectangular array with m rows and n columns. We denote by M_{ij} the (i, j)-entry of M, that is, the entry in the ith row and jth column of the matrix M.

The following are examples of matrices:

$$\begin{bmatrix} 1 & 0 & -3 \\ -2 & 5 & 4 \end{bmatrix} \qquad \begin{bmatrix} 1 & -7 \\ 6 & 3 \\ 5 & 11 \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

2×3 matrix $\quad$ 3×2 matrix $\quad$ 2×2 matrix

Translating to matrices

Definition 7

Consider a system of m equations in n variables.

- The $m \times n$ matrix C formed by setting C_{ij} to be the coefficient of x_j in the ith equation is called the *coefficient matrix* of this system.
- The *augmented matrix* A is an $m \times (n+1)$ matrix formed just as C but whose last column contains the constants of each equation.

Example 8

A system of equations	coefficient matrix	augmented matrix
$\begin{aligned} x_1 + x_2 - x_3 &= 4 \\ 2x_1 - x_2 + 3x_3 &= -13 \\ -x_1 + 2x_2 - x_3 &= 8 \end{aligned}$	$\begin{bmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ -1 & 2 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & -1 & 3 & -13 \\ -1 & 2 & -1 & 8 \end{bmatrix}$

Elementary Row Operations

The method of solving systems by elimination can be translated to the language of matrices. We call the following the *elementary row operations*:

- (Replacement) Replace one row by the sum of itself and a multiple of another row.
- (Interchange) Interchange two rows.
- (Scaling) Multiply all entries in a row by a nonzero constant.

Definition 9

Two matrices are said to be *row equivalent* if one can be obtained from the other by a series of elementary row operations.

We will prove the next theorem eventually.

Theorem 10

Two linear systems are equivalent if and only if their augmented matrices are row equivalent.

Elementary Row Operations - Example

Consider the augmented matrix from Example 8.

Example 11

$$\begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & -1 & 3 & -13 \\ -1 & 2 & -1 & 8 \end{bmatrix} \xrightarrow[\text{R2+(-2)R1}]{\text{Repl}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 5 & -21 \\ -1 & 2 & -1 & 8 \end{bmatrix} \xrightarrow[\text{R3+R1}]{\text{Repl}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 5 & -21 \\ 0 & 3 & -2 & 12 \end{bmatrix}$$

$$\xrightarrow[\text{R3+R2}]{\text{Repl}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 5 & -21 \\ 0 & 0 & 3 & -9 \end{bmatrix} \xrightarrow[\frac{1}{3}\cdot\ \text{R3}]{\text{Scale}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 5 & -21 \\ 0 & 0 & 1 & -3 \end{bmatrix} \xrightarrow[\text{R2+(-5)R3}]{\text{Repl}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 0 & -6 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

$$\xrightarrow[-\frac{1}{3}\cdot\ \text{R2}]{\text{Scale}} \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3 \end{bmatrix} \xrightarrow[\text{R1+R3}]{\text{Repl}} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3 \end{bmatrix} \xrightarrow[\text{R1+(-1)R2}]{\text{Repl}} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

Translating back to a (new) system, we have $x_1 = -1$, $x_2 = 2$, and $x_3 = -3$. Since the two augmented matrices are row equivalent, then by Theorem 10, the systems are equivalence. Hence, the systems have the same solution set.

Said another way, the unique solution to our system is $(-1, 2, -3)$.

Echelon form

The *leading entry* of a row in a matrix is the first nonzero entry when read left to right.

Definition 12

A rectangular matrix is in *row echelon form* (REF) if it has the following three properties.

(1) All nonzero rows are above any rows of all zeros.

(2) Each leading entry of a row is in a column to the right of the leading entry of the row above it.

(3) All entries in a column below a leading entry are zeros.

This form is *reduced* (RREF) if in addition

(4) The leading entry in each nonzero row is 1.

(5) Each leading 1 is the only nonzero entry in its column.

Echelon form

The following matrices are in row echelon form (REF), but only the second one is in reduced row echelon form (RREF).

Example 13

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

(REF) (RREF)

Theorem 14

Every matrix is equivalent to one and only one matrix in RREF. Thus, two matrices with the same RREF are row equivalent.

Gaussian elimination

Definition 15

A *pivot position* in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A *pivot column* is a column of A that contains a pivot position.

Row reduction algorithm (Gaussian Elimination)

(1) Begin with the leftmost nonzero column (this is a pivot column).

(2) Interchange rows as necessary so the top entry is nonzero.

(3) Use row operations to create zeros in all positions below the pivot.

(4) Ignoring the row containing the pivot, repeat 1-3 to the remaining submatrix. Repeat until there are no more rows to modify.

(5) Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. Make each pivot 1 by scaling.

Gaussian elimination

Example 16

Put the following matrix in RREF:

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{bmatrix} \xrightarrow[\text{R2+(-1)R1}]{\text{Repl}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{bmatrix} \xrightarrow[\text{R3+(-2)R1}]{\text{Repl}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -6 \\ 4 & 5 & 4 & 2 \end{bmatrix} \xrightarrow[\text{R4+(-4)R1}]{\text{Repl}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -6 \\ 0 & -3 & -12 & -18 \end{bmatrix}$$

$$\xrightarrow[\text{R2} \leftrightarrow \text{R4}]{\text{IC}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & -3 & -12 & -18 \\ 0 & 0 & -3 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{(REF)} \xrightarrow[-\frac{1}{3}\cdot\text{R2}]{\text{Scale}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & -3 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[-\frac{1}{3}\cdot\text{R3}]{\text{Scale}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[\text{R2+(-4)R3}]{\text{Repl}} \begin{bmatrix} 1 & 2 & 4 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{R1+(-2)R2}]{\text{Repl}} \begin{bmatrix} 1 & 0 & 4 & 9 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{R1+(-4)R3}]{\text{Repl}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{(RREF)}$$

This example corresponds to a system with a unique solution, $(x_1, x_2, x_3) = (1, -2, 2)$.

Solution sets

A solution need not be unique (or exist at all).

Example 17

The following matrix row reduces as

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 2 & 4 & 6 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The reduced form of the augmented matrix has a pivot in each column. This corresponds to an inconsistent system. The reason is that, translating back to a system, we get the equations $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, and $0 = 1$. This last equation is impossible so there is no solution to the system.

Solution sets

Example 18

The following matrix is in RREF.

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This corresponds to the system

$$\begin{aligned} x_1 - 5x_3 &= 1 \\ x_2 + x_3 &= 4 \\ 0 &= 0 \end{aligned}$$

Any choice of x_3 gives a solution. For example, setting $x_3 = 0$ gives the solution $(1, 4, 0)$. Setting $x_3 = 1$ gives $(6, 3, 1)$. Hence, there are infinitely many solutions to this system.

Solution sets

Definition 19

A variable corresponding to a pivot column in the augmented matrix of a system is a *basic variable*. Variables corresponding to non-pivot columns are called *free variables*.

Example 20

Consider the previous example.

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Here, the basic variables are x_1 and x_2 because they correspond to pivot columns. Since x_3 does not correspond to a pivot column, it is a free variable.

Solution sets

Theorem 21

A (linear) system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column. If a linear system is consistent, then the solution set contains either (i) a unique solution (no free variables) or (ii) infinitely many solutions.

The first condition in the theorem is equivalent to no row of the form

$$\begin{bmatrix} 0 & \cdots & 0 & b \end{bmatrix}, \qquad b \neq 0,$$

in the RREF form of the augmented matrix.

Solution sets - parametric form

One way to write the solution set is with *parametric form*. Here, free variables are listed as such and basic variables are solved for in terms of the free variables.

Example 22

a system of equations

$$\begin{aligned} x_1 - 5x_3 &= 1 \\ x_2 + x_3 &= 4 \end{aligned}$$

solution set in parametric form

$$\begin{cases} x_1 = 5x_3 + 1 \\ x_2 = -x_3 + 4 \\ x_3 \text{ is free} \end{cases}$$

Example 23

The following matrix is in RREF. Translate this into a system and identify the basic and free variables. Write the solution to this system in parametric form.

$$\begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

Next time

In the next lecture we will:

- Define vectors and discuss their algebraic properties.
- Define linear combinations and span.
- Define the product of a matrix and a vector.
- Discuss the connection between the matrix equation $A\mathbf{x} = \mathbf{b}$, vector equations, and systems of linear equations.

Chapter 1: Linear Equations
§1.3 Vector equations
§1.4 The matrix equation $A\mathbf{x} = \mathbf{b}$

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define vectors and discuss their algebraic properties.
- Define linear combinations and span.
- Define the product of a matrix and a vector.
- Discuss the connection between the matrix equation $A\mathbf{x} = \mathbf{b}$, vector equations, and systems of linear equations.

Vectors

Definition 1

A matrix with one column (an $n \times 1$ matrix) is said to be a *column vector*, which for now we will just call a vector and denote it by $\mathbf{v}$ or $\vec{v}$.

- The *dimension* of a vector is the number of rows (n).
- Two vectors are *equal* if they have the same dimension and all corresponding entries are equal.
- A vector whose entries are all zero is called a *zero vector* and denoted $\mathbf{0}$.

There is a corresponding notion of a row vector but columns are more appropriate for our use now. The following are all examples of vectors.

Example 2

$$\mathbf{u} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} e \\ \pi \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, w_i \in \mathbb{R}.$$

Vectors

We denote the set of n-dimensional vectors with entries in $\mathbb{R}$ by $\mathbb{R}^n$. There are two standard operations on $\mathbb{R}^n$.

For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we have:

$$c\mathbf{v} = c \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix} \qquad \mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

(Scalar Multiplication) (Addition)

Example 3

Let $\mathbf{u} = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]$ and $\mathbf{v} = \left[\begin{smallmatrix} -3 \\ 4 \end{smallmatrix}\right]$. Then

$$\mathbf{u} + 2\mathbf{v} = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] + 2\left(\left[\begin{smallmatrix} -3 \\ 4 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] + \begin{bmatrix} 2(-3) \\ 2(4) \end{bmatrix} = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} -6 \\ 8 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2-6 \\ -1+8 \end{smallmatrix}\right] = \left[\begin{smallmatrix} -4 \\ 7 \end{smallmatrix}\right].$$

Algebraic properties of vectors:

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all scalars $c, d \in \mathbb{R}$, we have the following:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = \mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

These are easy to prove using the corresponding properties for real numbers. For example, for (i) we use commutivity of addition (in $\mathbb{R}$) to find

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

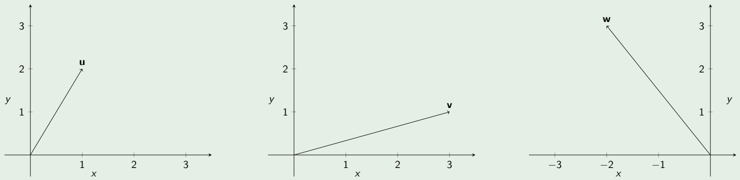
The others follow similarly.

An aside on $\mathbb{R}^2$

Geometrically, a vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2$ is an arrow from $(0,0)$ to (a,b).

Example 4

The vectors $\mathbf{u} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$, and $\mathbf{w} = \left[\begin{smallmatrix} -2 \\ 3 \end{smallmatrix}\right]$ are visualized below.



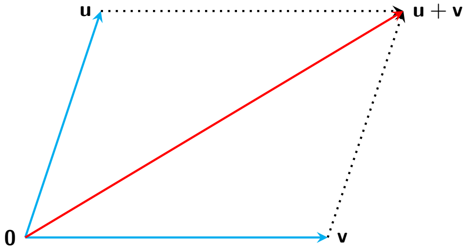
Any scalar multiple by $c \in \mathbb{R}$, $c \neq 0, 1$, of $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ points in the same direction but is longer if $c > 1$ and shorter if $0 < c < 1$. A scalar multiple by -1 reverses the direction of the vector.

An aside on $\mathbb{R}^2$

Vector addition can be visualized via the parallelogram rule. It is important to remember this geometric fact: a parallelogram is uniquely determined by three points in the plane.

Parallelogram Rule for Addition

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ are represented by points in the plan, then $\mathbf{u}+\mathbf{v}$ corresponds to the fourth vertex of a parallelogram whose other vertices are $\mathbf{0}, \mathbf{u}$, and $\mathbf{v}$.

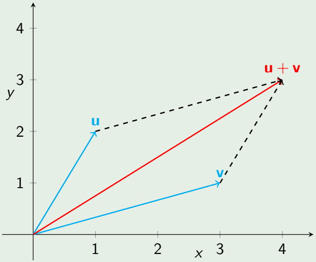


An aside on $\mathbb{R}^2$

Example 5

Let $\mathbf{u} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and $\mathbf{v} = \left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$.

Geometrically, we find that $\mathbf{u} + \mathbf{v} = \left[\begin{smallmatrix}4\\3\end{smallmatrix}\right]$ as below:



We can confirm this algebraically:

$$\mathbf{u} + \mathbf{v} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] + \left[\begin{smallmatrix}3\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}1+3\\2+1\end{smallmatrix}\right] = \left[\begin{smallmatrix}4\\3\end{smallmatrix}\right].$$

Span, span, span, span

Definition 6

A *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ with *weights* $c_1, \ldots, c_p \in \mathbb{R}$ is defined as the vector $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$. The set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is called the *span* of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ (or the subset of $\mathbb{R}^n$ *spanned by* $\mathbf{v}_1, \ldots, \mathbf{v}_p$) and is denoted $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Geometrically, we think of the span of one nonzero vector $\mathbf{v}$ as line through the origin since any vector in $\mathrm{Span}\{\mathbf{v}\}$ is of the form $c\mathbf{v}$ for some scalar (weight) c.

Similarly, the span of two nonzero vectors $\mathbf{v}_1, \mathbf{v}_2$ which are not scalar multiples forms a plane through the three points $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2$.

Example 7

Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}$. Is $\mathbf{w} \in \operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$?

We are asking whether $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. That is, are there weights x_1, x_2 such that $x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{w}$. This is equivalent to

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 3x_1+2x_2 \\ -x_1+x_2 \\ x_1-3x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}.$$

This gives a system of three equations and two unknowns. We form the corresponding augmented matrix and row reduce

$$\begin{bmatrix} 3 & 2 & 1 \\ -1 & 1 & 3 \\ 1 & -3 & -7 \end{bmatrix} \xrightarrow[\text{R1} \leftrightarrow \text{R3}]{\text{IC}} \begin{bmatrix} 1 & -3 & -7 \\ -1 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow[\text{R2+R1}]{\text{Repl}} \begin{bmatrix} 1 & -3 & -7 \\ 0 & -2 & -4 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow[\text{R3+(-3)R1}]{\text{Repl}} \begin{bmatrix} 1 & -3 & -7 \\ 0 & -2 & -4 \\ 0 & 11 & 22 \end{bmatrix}$$

$$\xrightarrow[-\frac{1}{2}\cdot \text{R2}]{\text{Scale}} \begin{bmatrix} 1 & -3 & -7 \\ 0 & 1 & 2 \\ 0 & 11 & 22 \end{bmatrix} \xrightarrow[\text{R3+(-11)R2}]{\text{Repl}} \begin{bmatrix} 1 & -3 & -7 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{R1+(3)R2}]{\text{Repl}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

This gives the (unique) solution $x_1 = -1$ and $x_2 = 2$, so $\mathbf{w} \in \operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$.

Span, span, span, span

Definition 8

Let $\mathbf{e}_i \in \mathbb{R}^n$ denote the vector of all zeros except a 1 in the ith spot. We call the $\mathbf{e}_i$ the *standard basis vectors* of $\mathbb{R}^n$.

Theorem 9

For any $n \in \mathbb{N}$, $\mathbb{R}^n = \text{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Proof.

To show that two sets are equal, we show that each set is contained in the other. Any linear combination of the $\mathbf{e}_i$ lives in $\mathbb{R}^n$, so $\text{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset \mathbb{R}^n$. On the other hand, if $\mathbf{a} \in \mathbb{R}^n$, then $\mathbf{a} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n \in \text{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Thus, $\mathbb{R}^n \subset \text{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. $\square$

Vector equations

A vector equation $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{b}$ has the same solution as the linear system whose augmented matrix is $\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n & \mathbf{b}\end{bmatrix}$. In particular, $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ if and only if there exists a solution to the corresponding linear system.

Example 10

Let $\mathbf{v}_1 = \begin{bmatrix}5\\-5\\25\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\3\\-6\end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix}2\\-2\\10\end{bmatrix}$.
Find all solutions to the system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$.

We form the augmented matrix of the system and row reduce,

$$\begin{bmatrix}5 & 0 & 2 & 0\\-5 & 3 & -2 & 0\\25 & -6 & 10 & 0\end{bmatrix} \xrightarrow[\text{R2+R1}]{\text{Repl}} \begin{bmatrix}5 & 0 & 2 & 0\\0 & 3 & 0 & 0\\25 & -6 & 10 & 0\end{bmatrix} \xrightarrow[\text{R3+(-5)R1}]{\text{Repl}} \begin{bmatrix}5 & 0 & 2 & 0\\0 & 3 & 0 & 0\\0 & -6 & 0 & 0\end{bmatrix} \xrightarrow[\text{R3+(2)R2}]{\text{Repl}} \begin{bmatrix}5 & 0 & 2 & 0\\0 & 3 & 0 & 0\\0 & 0 & 0 & 0\end{bmatrix}$$

$$\xrightarrow[\frac{1}{5}\cdot\text{ R1}]{\text{Scale}} \begin{bmatrix}1 & 0 & 2/5 & 0\\0 & 3 & 0 & 0\\0 & 0 & 0 & 0\end{bmatrix} \xrightarrow[\frac{1}{3}\cdot\text{ R2}]{\text{Scale}} \begin{bmatrix}1 & 0 & 2/5 & 0\\0 & 1 & 0 & 0\\0 & 0 & 0 & 0\end{bmatrix}$$

Hence, x_3 is a free variable. The solution is $\begin{cases} x_1 = -\frac{2}{5}x_3 \\ x_2 = 0 \\ x_3 \text{ is free.}\end{cases}$

Matrix equation

Definition 11

If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$, then the *product of A and* $\mathbf{x}$, denoted $A\mathbf{x}$, is the linear combination

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Observe that the number of columns of A must match the dimension of $\mathbf{x}$, and that the result will be a vector of dimension m.

Example 12

Let $A = \begin{bmatrix} -1 & 2 & 0 \\ 3 & 1 & 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$. Then

$$A\mathbf{x} = (2)\begin{bmatrix} -1 \\ 3 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 1 \end{bmatrix} + (5)\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5 \end{bmatrix} = \begin{bmatrix} -4 \\ 10 \end{bmatrix}.$$

Matrix equations vs systems of equations

Theorem 13

If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b} \in \mathbb{R}^m$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

which in turn has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}.$$

Another way to read the previous theorem is this: The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

An example

Example 14

Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and write $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent *for all* choices of $\mathbf{b}$?

(Partial solution) Row reduce $[A|\mathbf{b}]$ to REF form

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}.$$

The corresponding system has a solution if and only if the last column *is not* a pivot. Hence, the matrix equation has a solution if and only if $b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) = 0$.

Theorem 15

Let A be an $m \times n$ matrix. The following are equivalent.

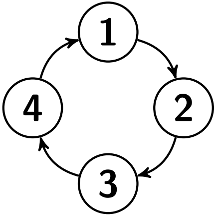
(1) *For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.*

(2) *Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.*

(3) *The columns of A span $\mathbb{R}^m$.*

(4) *A has a pivot in every row.*

We prove all of these statements are equivalent by showing the following equivalences:



Equivalence of notions

Theorem 15

Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Proof.

(1) $\Rightarrow$ (2). Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ denote the columns of A. By definition of the matrix product,

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

for some $\mathbf{x} \in \mathbb{R}^n$. Hence, $\mathbf{b}$ is a linear combination of the $\mathbf{a}_i$ with weights x_i.

Equivalence of notions

Theorem 15

Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Proof.

(2) $\Rightarrow$ (3). By definition, $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$. If $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A, then $\mathbf{b} \in \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Hence, $\mathbb{R}^m \subset \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. It follows that $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathbb{R}^m$.

Equivalence of notions

Theorem 15

Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Proof.

(3) $\Rightarrow$ (4). Suppose A does not have a pivot in every row. Then, in particular, A does not have a pivot in the last row. But then $\mathbf{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$ is not in the span of RREF(A). But this implies that $\text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \neq \mathbb{R}^m$.

Equivalence of notions

Theorem 15

Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Proof.

(4) $\Rightarrow$ (1). Since A has a pivot in every row, then the statement is clearly true for RREF(A). Since $[A|b]$ and RREF($[A|b]$) have the same solution space, then the claim holds. $\square$

Next time

In the next lecture we will:

- Discuss homogeneous versus non-homogeneous system solutions.
- Write solutions in parametric vector form.
- Discuss the geometric interpretation of solutions sets.
- Consider some applications of linear systems.

Chapter 1: Linear Equations
§1.5 Solution sets of linear systems
§1.6 Applications of linear systems

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Discuss homogeneous versus non-homogeneous system solutions.
- Write solutions in parametric vector form.
- Discuss the geometric interpretation of solutions sets.
- Consider some applications of linear systems.

Homogeneous systems

Definition 1

A linear system/vector equation is said to be *homogeneous* if it can be written as $A\mathbf{x} = \mathbf{0}$ where A is an $m \times n$ matrix and $x \in \mathbb{R}^n$. Such a system always has one solution, $\mathbf{0}$, called the *trivial solution*.

Theorem 2

A homogeneous system has a nontrivial solution if and only if the system has at least one free variable.

Homogeneous systems

Example 3

Solve the following homogeneous system

$$\begin{aligned} 3x_2 + x_3 &= 0 \\ -2x_1 + 5x_2 + 5x_3 &= 0 \\ x_1 + 2x_2 - x_3 &= 0. \end{aligned}$$

We form the augmented matrix of the system and row reduce:

$$\begin{bmatrix} 0 & 3 & 1 & 0 \\ -2 & 5 & 5 & 0 \\ 1 & 2 & -1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -5/3 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution in parametric form is $\begin{cases} x_1 = (5/3)x_3 \\ x_2 = (-1/3)x_3 \\ x_3 \text{ is free.} \end{cases}$

Parametric vector form

Here is another way to represent our solution sets.

Definition 4

The *parametric vector form* of a solution set is of the form

$$\mathbf{x} = \mathbf{u}_1 x_{i_1} + \cdots + \mathbf{u}_k x_{i_k} + \mathbf{p}$$

where the x_{i_j} are free variables and $\mathbf{p}$ is a particular solution not contained in $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

Note that $\mathbf{p}$ can be *any* solution not contained in $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. This means that solutions sets may appear different, but in fact are the same.

Parametric vector form

Example 5

Recall our last example, the solution in parametric form was $\begin{cases} x_1 = (5/3)x_3 \\ x_2 = (-1/3)x_3 \\ x_3 \text{ is free.} \end{cases}$

In parametric vector form, the solution would be

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (5/3)x_3 \\ (-1/3)x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5/3 \\ -1/3 \\ 1 \end{bmatrix} x_3.$$

Hence, the solution set is $\text{Span}\left(\begin{bmatrix} 5/3 \\ -1/3 \\ 1 \end{bmatrix} \right)$, which represents a line through the origin in $\mathbb{R}^3$. Note that

$$\text{Span}\left(\begin{bmatrix} 5/3 \\ -1/3 \\ 1 \end{bmatrix} \right) = \text{Span}\left(\begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix} \right).$$

Homogeneous systems

Example 6

Solve the following homogeneous "system" with one equation:

$$3x_1 + 2x_2 - 5x_3 = 0.$$

A general solution is $x_1 = -\frac{2}{3}x_2 + \frac{5}{3}x_3$ with x_2 and x_3 free. The solution in parametric vector form is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3}x_2 + \frac{5}{3}x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3.$$

Let $\mathbf{u} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix}$. Then the solution set is $\mathrm{Span}\{\mathbf{u}, \mathbf{v}\} \subset \mathbb{R}^3$, which represents a plane through the origin in $\mathbb{R}^3$.

Nonhomogeneous systems

Example 7

Solve the following "system" with one equation: $3x_1 + 2x_2 - 5x_3 = -1$.

This is almost the same system as Example **??**. Since the systems have the same coefficient matrix, then their homogenous solutions will be the same. We need a *particular solution*. One such solution is $(-1/3, 0, 0)$.

As before, we write the solution in parametric vector form as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3}x_2 + \frac{5}{3}x_3 - \frac{1}{3} \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}.$$

This is a translation of the homogeneous solution by the vector $\mathbf{p} = \begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}$.

An example

Example 8

The system $3x_1 + 2x_2 - 5x_3 = -1$ has solution set in parametric vector form:

$$\mathbf{x} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}.$$

Another choice of particular solution is $(0, -1/2, 0)$ and so we could write the solution set as

$$\mathbf{x}' = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} 0 \\ -1/2 \\ 0 \end{bmatrix}.$$

Why are these sets the same? Notice that taking $x_2 = 1/2$ and $x_3 = 0$ in $\mathbf{x}'$ gives $\begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}$. Similarly, taking $x_2 = -1/2$ and $x_3 = 0$ in $\mathbf{x}$ gives $\begin{bmatrix} 0 \\ -1/2 \\ 0 \end{bmatrix}$.

We say that parametric form is *independent* of the choice of particular solution.

Solution sets

Theorem 9

Suppose $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$ and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Proof.

Suppose $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$ as given above. Then

$$A\mathbf{w} = A(\mathbf{p} + \mathbf{v}_h) = A\mathbf{p} + A\mathbf{v}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

That is, $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$. Now let $\mathbf{y}$ be some solution. We claim $\mathbf{y}$ has the form given in the theorem. Then

$$A(\mathbf{y} - \mathbf{p}) = A\mathbf{y} - A\mathbf{p} = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

so $(\mathbf{y} - \mathbf{p})$ is a solution of $A\mathbf{x} = \mathbf{0}$. Hence, $\mathbf{y} = \mathbf{p} + (\mathbf{y} - \mathbf{p})$ has the desired form. $\square$

Translations

Geometrically, we think of the solution set of $A\mathbf{x} = \mathbf{b}$ as a *translation* of the solution set of $A\mathbf{x} = \mathbf{0}$ in the direction of the particular solution.

Example 10

Solve the following system

$$\begin{aligned} 3x_2 + x_3 &= 1 \\ -2x_1 + 5x_2 + 5x_3 &= 1 \\ x_1 + 2x_2 - x_3 &= 1. \end{aligned}$$

The homogeneous part of this system is the same as in Example **??**. Solving, we find that the solution in parametric vector form is

$$\mathbf{x} = \begin{bmatrix} -5/3 \\ 1/3 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} 1/3 \\ 1/3 \\ 0 \end{bmatrix}.$$

The solution set is a line in $\mathbb{R}^3$ that has been translated in the direction of $\mathbf{p} = \begin{bmatrix} 1/3 \\ 1/3 \\ 0 \end{bmatrix}$.

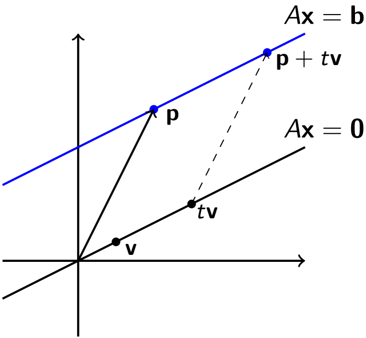
Translations

Definition 11

The equation

$$\mathbf{x} = \mathbf{p} + t\mathbf{v}, \qquad t \in \mathbb{R}$$

is the *parametric equation of the line through* $\mathbf{p}$ *parallel to* $\mathbf{v}$.



Applications: Economics

Example 12

Suppose an economy has four sectors: Agriculture (A), Energy (E), Manufacturing (M), and Transportation (T).

- A sells 10% of its output to E, 25% to M, and retains the rest.
- E sells 30% of its output to A, 35% to M, 25% to T, and retains the rest.
- M sells 30% of its output to A, 15% to E, 40% to T, and retains the rest.
- T sells 20% of its output to A, 10% to E, 30% to M, and retains the rest.

Leontif proved that there exist *equilibrium prices* that can be assigned to the total outputs of the various sectors so that the income of each sector balances its expenses. First, we construct an exchange table for this economy:

Distribution of output from:	A	E	M	T	Purchased by
	.65	.3	.3	.2	A
	.1	.1	.15	.1	E
	.25	.35	.15	.3	M
	0	.25	.4	.4	T

Applications: Economics

Example 12

Distribution of output from:	A	E	M	T	Purchased by
	.65	.3	.3	.2	A
	.1	.1	.15	.1	E
	.25	.35	.15	.3	M
	0	.25	.4	.4	T

Denote the prices of the total annual outputs of Agriculture, Energy, Manufacturing, and Transportation by p_A, p_E, p_M, and p_T, respectively. Agriculture pays for 65% of its own output, 30% of Energy's output, 30% of Manufacturing's output, and 20% of Transportation's output. To make Agriculture's income equal its expenses, we need

$$p_A = .65p_A + .3p_E + .3p_M + .2p_T$$

or

$$.35p_A - .3p_E - .3p_M - .2p_T = 0.$$

Applications: Economics

Example 12

This leads to the following system of equations:

$$\begin{aligned} .35p_A - .3p_E - .3p_M - .2p_T &= 0 \\ -.1p_A + .9p_E - .15p_M - .1p_T &= 0 \\ -.25p_A - .35p_E + .85p_M - .3p_T &= 0 \\ -.25p_E - .4p_M + .6p_T &= 0. \end{aligned}$$

We form the corresponding augmented matrix and row reduce:

$$\begin{bmatrix} .35 & -.3 & -.3 & -.2 & 0 \\ -.1 & .9 & -.15 & -.1 & 0 \\ -.25 & -.35 & .85 & -.3 & 0 \\ 0 & -.25 & -.4 & .6 & 0. \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2.03 & 0 \\ 0 & 1 & 0 & -.53 & 0 \\ 0 & 0 & 1 & -1.17 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution in parametric vector form is

$$p = \begin{bmatrix} p_A \\ p_E \\ p_M \\ p_T \end{bmatrix} = \begin{bmatrix} 2.03 \\ .53 \\ 1.17 \\ 1 \end{bmatrix} p_T.$$

Setting $p_T = 100$ gives $p_A = 203$, $p_E = 53$, and $p_M = 117$.

Applications: Chemical equations

Chemical equations describe quantities of substances consumed and produced by chemical reactions. (Friendly reminder: atoms are neither destroyed nor created, just changed.)

Example 13

When propane gas burns, propane C_3H_8 combines with oxygen O_2 to form carbon dioxide CO_2 and water H_2O according to an equation of the form

$$(x_1)\mathrm{C_3H_8} + (x_2)\mathrm{O_2} \longrightarrow (x_3)\mathrm{CO_2} + (x_4)\mathrm{H_2O}.$$

To *balance the equation* means to find x_i such that the total number of atoms on the left and right are equal. We translate the chemicals into vectors $\begin{bmatrix} C \\ H \\ O \end{bmatrix}$,

$$\mathrm{C_3H_8} = \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix}, \qquad \mathrm{O_2} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathrm{CO_2} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathrm{H_2O} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Balancing now becomes the linear system,

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Applications: Chemical equations

Example 13

Equivalently,

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} - x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \mathbf{0}.$$

We form the augmented matrix and row reduce,

$$\begin{bmatrix} 3 & 0 & -1 & 0 & 0 \\ 8 & 0 & 0 & -2 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -1/4 & 0 \\ 0 & 1 & 0 & -5/4 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{bmatrix}.$$

Thus, the solution is $x_1 = (1/4)x_4$, $x_2 = (5/4)x_4$, $x_3 = (3/4)x_4$ with x_4 free. Since only integer solutions make sense in this context, a solution is to take $x_4 = 4$, which gives

$$\mathrm{C_3H_8} + 5\mathrm{O_2} \longrightarrow 3\mathrm{CO_2} + 4\mathrm{H_2O}.$$

Next time

In the next lecture we will:

- Define linear independence and efficient ways to write the span of a set of vectors.
- Practice checking for linear independence and discuss some special cases.
- Prove a theorem giving equivalent criteria for linear independence.
- Discuss linear independence in $\mathbb{R}^n$.

Chapter 1: Linear Equations
§1.7 Linear independence

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define linear independence and efficient ways to write the span of a set of vectors.
- Practice checking for linear independence and discuss some special cases.
- Prove a theorem giving equivalent criteria for linear independence.
- Discuss linear independence in $\mathbb{R}^n$.

Linear independence

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$. Writing $\text{Span}\{\mathbf{u}, \mathbf{v}\}$ is somewhat deceptive because we might expect that this represents a plane in $\mathbb{R}^3$. In fact, $\mathbf{v} = 2\mathbf{u}$, so $\text{Span}\{\mathbf{u}, \mathbf{v}\} = \text{Span}\{\mathbf{u}\}$, so this represents a line in $\mathbb{R}^3$. Today we discuss the most "efficient" way to express the span of the vectors.

Definition 1

An index set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be *linearly independent* if the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_p\mathbf{v}_p = \mathbf{0} \tag{1}$$

has only the trivial solution. Otherwise the set is said to be *linearly dependent*. That is, there exist weights $c_1, \ldots, c_p$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}.$$

Linear independence - examples

Example 2

Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix}$.

To determine whether the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent, we set up the augmented matrix corresponding to the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$$

and reduce:

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 3 & 5 & 2 & 0 \\ -2 & -1 & 7 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Hence, the only solution is the trivial one and so the set is linearly independent.

Linear independence - examples

Example 3

Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

To determine whether the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent, we set up the augmented matrix corresponding to the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$$

and reduce:

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

There is a nontrivial solution (in particular, x_3 is free). Hence, the set is linearly dependent.

Since x_3 is free, then $\mathbf{v}_3$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Hence, $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

Linear independence - examples

Here are some special cases.

Example 4

The set $\{\mathbf{0}\}$ is linearly dependent because $c\mathbf{0} = \mathbf{0}$ for all $c \in \mathbb{R}$. Moreover, any set containing the zero vector is linearly dependent. To see this, suppose $\mathbf{v}_1 = \mathbf{0}$, then $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$. A similar argument holds if any $\mathbf{v}_i$ is the zero vector.

Example 5

Let $\mathbf{v}$ be a nonzero vector. Then $\{\mathbf{v}\}$ is linearly independent because $c\mathbf{v} = \mathbf{0}$ implies $c = 0$.

Example 6

A set of two vectors is linearly dependent if and only if one vector is a multiple of the other. (Exercise)

A test for linear dependence

Theorem 7

An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, $p \geq 2$, is linearly dependent if and only if at least one of the vectors is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq 0$, then some $\mathbf{v}_j$, $j > 1$, is a linear combination of the preceding vectors.

On one hand, the theorem gives us a way to recognize when a set is linearly dependent. At times, this may be more useful than row reduction. However, the real power of this theorem is in its applications to other results down the road.

A test for linear dependence

Theorem 7

An indexed set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$, $p \geq 2$, is linearly dependent if and only if at least one of the vectors is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq 0$, then some $\mathbf{v}_j$, $j > 1$, is a linear combination of the preceding vectors.

Proof.

Suppose S is linearly dependent. Then there exists weights $c_1, \dots, c_p$ such that $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$. If $c_1 \neq 0$, then $\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_p}{c_1}\mathbf{v}_p$, so $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_p$. (Note that it must be true that at least one of $c_2, \dots, c_p$ is nonzero.)

Conversely, if $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_p$, then there exists weights $c_2, \dots, c_p$ such that $\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$. Equivalently, $-\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$, so S is linearly dependent.

Both arguments hold with any vector in place of $\mathbf{v}_1$. $\square$

A test for linear dependence - consequences

Example 8

The set $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\1\end{bmatrix}\right\}$ is linearly dependent.

Observe that

$$\begin{bmatrix}-4\\1\end{bmatrix} = (-2)\begin{bmatrix}1\\0\end{bmatrix} + (-1)\begin{bmatrix}2\\-1\end{bmatrix}.$$

One can find the weights above using row reduction.

Example 9

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent if and only if one of the vectors is in the span of the other two.

Note: the previous example does not require that $\mathbf{w} \in \text{Span}\{\mathbf{u}, \mathbf{v}\}$. Consider, for example, the set

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

Linear independence in $\mathbb{R}^n$

Theorem 10

Any set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subset \mathbb{R}^n$ *is linearly dependent if* $p > n$.

Proof.

The matrix $\begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_p \end{bmatrix}$ has dimensions $n \times p$. If $p > n$, then there are more columns than rows, so there will not be a pivot in every column. Hence, there are free variables and so the solution to

$$x_1\mathbf{v}_1 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

is not unique. (That is, there are nonzero solutions to this equation.) □

Next time

In the next lecture we will:

- Discuss properties of functions on vector spaces defined by matrices.
- Define linear transformations on $\mathbb{R}^n$.
- Discuss properties of linear transformations (1-1 and onto).
- Consider the relationship between maps defined by matrices and linear transformations.

Chapter 1: Linear Equations
§1.8 Introduction to linear transformations
§1.9 The matrix of a linear transformation

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Discuss properties of functions on vector spaces defined by matrices.
- Define linear transformations on $\mathbb{R}^n$.
- Discuss properties of linear transformations (1-1 and onto).
- Consider the relationship between maps defined by matrices and linear transformations.

Maps defined by matrices

Given an $m \times n$ matrix A, the rule $T(\mathbf{x}) = A\mathbf{x}$ defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Example 1

Define $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]$, $\mathbf{b} = \left[\begin{smallmatrix} 3 \\ 2 \\ -5 \end{smallmatrix}\right]$, $\mathbf{c} = \left[\begin{smallmatrix} 3 \\ 2 \\ 5 \end{smallmatrix}\right]$.

Since A is 3×2, then T is a function from $\mathbb{R}^2$ to $\mathbb{R}^3$. We'll consider several questions related to T.

(1) Find $T(\mathbf{u})$.

We compute

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] = (2)\left[\begin{smallmatrix} 1 \\ 3 \\ -1 \end{smallmatrix}\right] + (-1)\left[\begin{smallmatrix} -3 \\ 5 \\ 7 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 5 \\ 1 \\ -9 \end{smallmatrix}\right].$$

Maps defined by matrices

Given an $m \times n$ matrix A, the rule $T(\mathbf{x}) = A\mathbf{x}$ defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Example 1

Define $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$.

(2) Find some $\mathbf{x} \in \mathbb{R}^2$ whose image under T is $\mathbf{b}$.

We need to solve the matrix equation $A\mathbf{x} = \mathbf{b}$, so we form the augmented matrix $[A \mid \mathbf{b}]$ and row reduce.

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 3/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{bmatrix}.$$

There are no free variables, so the unique solution is $\mathbf{x} = \begin{bmatrix} 3/2 \\ -1/2 \end{bmatrix}$.

Maps defined by matrices

Given an $m \times n$ matrix A, the rule $T(\mathbf{x}) = A\mathbf{x}$ defines a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Example 1

Define $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$.

(3) Is $\mathbf{c}$ in the range of T?

This is basically the same question as before. We need to solve the matrix equation $A\mathbf{x} = \mathbf{c}$, so we form the augmented matrix $[A \mid \mathbf{c}]$ and row reduce.

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since there a pivot in the last column then there is no solution. Hence, $\mathbf{c}$ is not in the range of T.

Linear transformations

Definition 2

A *transformation* T from $\mathbb{R}^n$ to $\mathbb{R}^m$ (written $T : \mathbb{R}^n \to \mathbb{R}^m$) is a rule that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$, called the *image of* $\mathbf{x}$. We call $\mathbb{R}^n$ the domain and $\mathbb{R}^m$ the *codomain*. The set of all images under T is called the *range*.

We will be primarily interested in transformations that respect the algebraic properties of vectors.

Definition 3

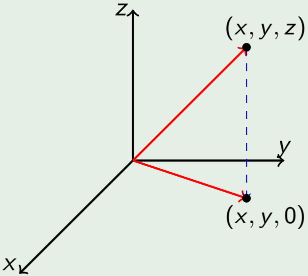
A transformation is *linear* if

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.
- $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $c \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^n$.

Example 4

Let $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Then T is a transformation $\mathbb{R}^3 \to \mathbb{R}^3$ but the range of T is "equivalent" in some way to $\mathbb{R}^2$. We say T is a *projection* of $\mathbb{R}^3$ onto $\mathbb{R}^2$.

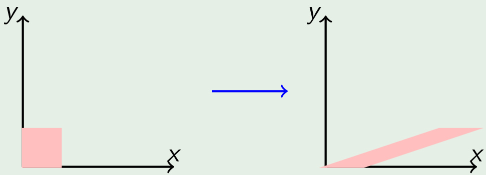


Linear transformation - examples

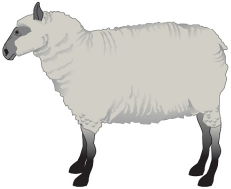
Example 5

Let $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$.

Then T is called a *shear transformation* because it leaves fixed the second component of $\mathbf{x}$.



Linear transformation - examples



sheep

sheared sheep

Linear transformation - examples

All of our examples of linear transformations have been defined by matrices. You might ask whether we can define a rule that is not determined by a matrix.

Example 6

Consider the map $T : \mathbb{R} \to \mathbb{R}^2$ given by $T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$.

This map T is not linear (because $(a+b)^2 \neq a^2 + b^2$ in general). For example,

$$\left[\begin{smallmatrix} 3 \\ 9 \end{smallmatrix}\right] = T(3) = T(1+2) = T(1) + T(2) = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right],$$

which is absurd.

The zero vector

As mentioned before, linear transformations preserve the algebraic properties of vectors. The zero vector $\mathbf{0}$ is a special vector so we might expect that linear transformations preserve this vector.

Theorem 7

If T is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

Proof.

Let $\mathbf{y} = T(\mathbf{0})$. By linearity,

$$\mathbf{y} = T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}) = 2\mathbf{y}.$$

Hence, $\mathbf{0} = 2\mathbf{y} - \mathbf{y} = \mathbf{y} = T(\mathbf{0})$. □

Now we reveal the intimate connection between linear transformations and matrices. In some sense, they are the same thing, as the next theorem explains.

Linear transformations and matrices

Theorem 8

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists a unique $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

Proof.

($\Leftarrow$) Suppose $T(\mathbf{x}) = A\mathbf{x}$. We must show that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(c\mathbf{u}) = cA\mathbf{u}$. Denote the columns of A by $\mathbf{a}_1, \dots, \mathbf{a}_n$.

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= \mathbf{a}_1(u_1 + v_1) + \cdots + \mathbf{a}_n(u_n + v_n) \\ &= (\mathbf{a}_1 u_1 + \mathbf{a}_1 v_1) + \cdots + (\mathbf{a}_n u_n + \mathbf{a}_n v_n) \\ &= (\mathbf{a}_1 u_1 + \cdots + \mathbf{a}_n u_n) + (\mathbf{a}_1 v_1 + \cdots + \mathbf{a}_n v_n) \\ &= A\mathbf{u} + A\mathbf{v}. \end{aligned}$$

$$\begin{aligned} A(c\mathbf{u}) &= \mathbf{a}_1(cu_1) + \cdots + \mathbf{a}_n(cu_n) \\ &= c\mathbf{a}_1 u_1 + \cdots + c\mathbf{a}_n u_n \\ &= c(\mathbf{a}_1 u_1 + \cdots + \mathbf{a}_n u_n) \\ &= cA\mathbf{u}. \end{aligned}$$

Hence, T is a linear transformation.

Linear transformations and matrices

Theorem 8

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists a unique $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

Proof.

($\Rightarrow$) Suppose T is linear. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be the standard basis vectors of $\mathbb{R}^n$. For $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$. By linearity,

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) \\ &= x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) \\ &= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \mathbf{x}. \end{aligned}$$

Set $A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$.

Linear transformations and matrices

Theorem 8

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists a unique $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

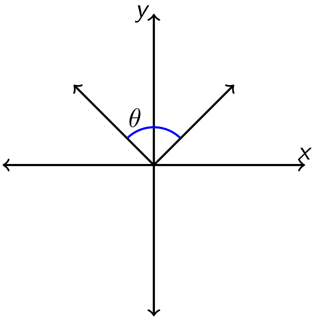
Proof.

($\Rightarrow$) We still need to show that A is unique. Suppose $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B with columns $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. Then $T(\mathbf{e}_1) = B\mathbf{e}_1 = \mathbf{b}_1$. But $T(\mathbf{e}_1) = A\mathbf{e}_1 = \mathbf{a}_1$, so $\mathbf{a}_1 = \mathbf{b}_1$. Repeating for each column of B gives $B = A$. □

We call A the *standard matrix* of T.

Example 9

Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point θ° counterclockwise. Then $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] \mapsto \left[\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \mapsto \left[\begin{smallmatrix}-\sin\theta\\ \cos\theta\end{smallmatrix}\right]$. Hence, the standard matrix of T is $A = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$.



Properties of transformations

Definition 10

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be

- *onto* (surjective) if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at least one $\mathbf{x} \in \mathbb{R}^n$
- *one-to-one* (injective) if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at most one $\mathbf{x} \in \mathbb{R}^n$.
- *an isomorphism* (bijective) if it is both onto and one-to-one.

Another way to phrase one-to-one (1-1) is $T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$.

Example 11

- The map $T : \mathbb{R}^2 \to \mathbb{R}$ given by $T(\mathbf{x}) = x_1$ is onto but not 1-1.
- The map $T : \mathbb{R} \to \mathbb{R}^2$ given by $T(x) = \begin{bmatrix} x \\ 2x \end{bmatrix}$ is 1-1 but not onto.
- The map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ is 1-1 and onto.

Criteria for injectivity

Theorem 12

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is 1-1 if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Proof.

($\Rightarrow$) Assume T is 1-1. Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. Because T is 1-1, $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$ implies $\mathbf{x} = \mathbf{0}$. Hence $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

($\Leftarrow$) Assume $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. Suppose $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. Thus, $\mathbf{x} = \mathbf{y}$ and T is 1-1. $\square$

The set $\{\mathbf{x} : T(\mathbf{x}) = \mathbf{0}\}$ is called the *kernel* of T. Another way to state the previous theorem is to say that T is 1-1 if and only if its kernel contains only $\mathbf{0}$.

The next theorem frames onto and 1-1 in terms of properties of the standard matrix.

The standard matrix

Theorem 13

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A.

(1) T is onto if and only if the columns of A span $\mathbb{R}^m$.

(2) T is 1-1 if and only if the columns of A are linearly independent.

Proof.

(1) The columns of A span $\mathbb{R}^m$ if and only if for every $\mathbf{b} \in \mathbb{R}^m$ there exists $\mathbf{x} \in \mathbb{R}^n$ such that $x_1\mathbf{a}_1 + \cdots + x_1\mathbf{a}_n = \mathbf{b}$. This is equivalent to $A\mathbf{x} = \mathbf{b}$, which is equivalent to $T(\mathbf{x}) = \mathbf{b}$ and this holds if and only if T is onto.

(2) T is 1-1 if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution and this holds if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. That is, $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$ has only the trivial solution. This is equivalent to the columns of A being linearly independent. □

Isomorphisms

Theorem 13

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A.

(1) T is onto if and only if the columns of A span $\mathbb{R}^m$.

(2) T is 1-1 if and only if the columns of A are linearly independent.

Corollary 14

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. If T is an isomorphism, then A is a square matrix.

Proof.

Suppose T is an isomorphism, that is, T is bijective (1-1 and onto). By Theorem 13 (1), A has a pivot in every row. By Theorem 13 (2), A has a pivot in every column. This implies that $m = n$, so A is a square matrix. $\square$

Next time

In the next lecture we will:

- Define the operations of addition and scalar multiplication of matrices.
- Consider algebraic properties of these matrix operations.
- Define the operation of matrix multiplication and consider its algebraic properties.
- Define the operation of matrix transpose and consider its algebraic properties.

Chapter 2: Matrix algebra
§2.1 Matrix operations

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define the operations of addition and scalar multiplication of matrices.
- Consider algebraic properties of these matrix operations.
- Define the operation of matrix multiplication and consider its algebraic properties.
- Define the operation of matrix transpose and consider its algebraic properties.

Matrix operations

Recall that matrices correspond to linear transformations in a natural way. Today we discuss algebraic operations on matrices and how these correspond to operations on linear transformations.

Definition 1

Let $A = (a_{ij})$ be an $m \times n$ matrix.

- An $m \times n$ matrix is *square* if $m = n$.
- The *diagonal entries* of a matrix A are those a_{ij} with $i = j$.
- A square matrix A is
  - *diagonal* if $a_{ij} = 0$ for all $i \neq j$,
  - *upper triangular* if $a_{ij} = 0$ for all $i > j$, and
  - *lower triangular* if $a_{ij} = 0$ for all $i < j$.
- A matrix A is a *zero matrix* if $a_{ij} = 0$ for all i, j. We often write 0 for the zero matrix when m and n are implied.
- The (square) $n \times n$ *identity matrix* is a diagonal matrix I_n with 1 for every diagonal entry.

Matrix operations

Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices of the *same size* $m \times n$ and $c \in \mathbb{R}$.

- Matrix addition: $A + B$ is the $m \times n$ matrix whose (i, j)-entry is $a_{ij} + b_{ij}$.
- Scalar multiplication: cA is the $m \times n$ matrix whose (i, j)-entry is ca_{ij}.

Example 2

Let $A = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix}$. Then

$$A + 3B = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} + 3 \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} + \begin{bmatrix} 3 & 15 \\ -6 & 9 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 17 \\ -5 & 14 \\ 2 & 3 \end{bmatrix}.$$

Algebraic properties

Algebraic properties of matrices

Let A, B, C be matrices of the same size and $r, s \in \mathbb{R}$ scalars. We have the following:

(i) $A + B = B + A$

(ii) $(A + B) + C = A + (B + C)$

(iii) $A + 0 = A$

(iv) $r(A + B) = rA + rB$

(v) $(r + s)A = rA + sA$

(vi) $r(sA) = (rs)A$

These properties are similar to those of vectors. Later we will study *vector spaces* in general. The space of $m \times n$ matrices is an example of a vector space (as in $\mathbb{R}^n$) because it obeys these linearity properties.

The above properties are easy to prove. For example, for (i), we have $a_{ij} + b_{ij} = b_{ij} + a_{ij}$ for all i, j (because addition is commutative in $\mathbb{R}$).

Algebraic properties

Another way to prove the algebraic properties of matrices is to recognize that the matrices A, B, and C all correspond to linear transformations.

If T and S are linear transformations with standard matrices A and B, respectively, then $T + S$ is a linear transformation with standard matrix $A + B$. Similarly, if c is a scalar then the standard matrix of (cT) is just cA.

Question

What is the standard matrix associated to the composition of two linear transformations?

Matrix multiplication

If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^p \to \mathbb{R}^q$, then $T \circ S$ only makes sense of $q = n$. Let A and B be the standard matrices of T and S, respectively, so A is $m \times n$ and B is $n \times p$. Then

$$(T \circ S)(\mathbf{x}) = T(S(\mathbf{x})) = T(B\mathbf{x}) = A(B\mathbf{x}) = AB(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^p.$$

Hence, the matrix product AB is the standard matrix of $T \circ S$.

$$\begin{tikzcd}
\mathbb{R}^p \arrow[r, bend left, "S", "B"'] \arrow[rr, bend right, "AB", "T \circ S"'] & \mathbb{R}^n \arrow[r, bend left, "T", "A"'] & \mathbb{R}^m
\end{tikzcd}$$

Note that AB should then be $m \times p$.

Matrix multiplication

Definition 3

If A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_p$, then the *matrix product* AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \ldots, A\mathbf{b}_p$. That is,

$$AB = A\begin{bmatrix}\mathbf{b}_1 & \cdots \mathbf{b}_p\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & \cdots A\mathbf{b}_p\end{bmatrix}.$$

Example 4

Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{bmatrix}$. Then $\mathbf{b}_1 = \left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$, $\mathbf{b}_2 = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$, and $\mathbf{b}_3 = \left[\begin{smallmatrix} 4 \\ 6 \end{smallmatrix}\right]$. So $A\mathbf{b}_1 = \left[\begin{smallmatrix} -1 \\ -16 \end{smallmatrix}\right]$, $A\mathbf{b}_2 = \left[\begin{smallmatrix} 5 \\ 3 \end{smallmatrix}\right]$, and $A\mathbf{b}_3 = \left[\begin{smallmatrix} 14 \\ 26 \end{smallmatrix}\right]$. Hence,

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} -1 & 5 & 14 \\ -16 & 3 & 26 \end{bmatrix}.$$

In Example **??**, the product BA is not defined. In general, $AB \neq BA$ even if both are defined. Also, cancellation does not work. That is, it is possible for $AB = 0$ even if $A, B \neq 0$.

Matrix multiplication

Here is an alternate way to compute AB.

Row-column rule

Let $A = (a_{ij})$ be $m \times n$ matrix and let $B = (b_{ij})$ be an $n \times p$ matrix. Then

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}.$$

Example 5

Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 2(1) + 1(-3) & 2(2) + 1(1) & 2(4) + 1(6) \\ -1(1) + 5(-3) & -1(2) + 5(1) & -1(4) + 5(6) \end{bmatrix} = \begin{bmatrix} -1 & 5 & 14 \\ -16 & 3 & 26 \end{bmatrix}.$$

More algebraic properties

Algebraic properties of matrices (contd)

Let A, B, C be matrices of appropriate sizes such that each sum/product is defined in each of the follow, and let $r \in \mathbb{R}$ be a scalar. We have the following:

(i) $A(BC) = (AB)C$

(ii) $A(B + C) = AB + AC$

(iii) $(B + C)A = BA + CA$

(iv) $r(AB) = (rA)B = A(rB)$

(v) $I_m A = A = AI_n$

If A is a square $n \times n$ matrix and k a positive integer, then the *matrix power* A^k denotes the product of k copies of A. By definition, $A^0 = I_n$.

The transpose

Definition 6

If A is an $m \times n$ matrix, the *transpose* of A, denoted A^T, is an $n \times m$ matrix whose columns are the rows of A. That is, $(A^T)_{ij} = A_{ji}$.

Example 7

Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{bmatrix}$. Then $A^T = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}$, $B^T = \begin{bmatrix} 1 & -3 \\ 2 & 1 \\ 4 & 6 \end{bmatrix}$.

Moreover,

$$B^T A^T = \begin{bmatrix} 1(2) + (-3)(1) & 1(-1) + (-3)(5) \\ 2(2) + 1(1) & 2(-1) + 1(5) \\ 4(2) + 6(1) & 4(-1) + 6(5) \end{bmatrix} = \begin{bmatrix} -1 & -16 \\ 5 & 3 \\ 14 & 26 \end{bmatrix} = (AB)^T.$$

The transpose

Algebraic properties of the transpose

Let A and B be matrices of appropriate sizes and let $r \in \mathbb{R}$ be a scalar. We have the following:

(i) $(A^T)^T = A$

(ii) $(A+B)^T = A^T + B^T$

(iii) $(AB)^T = B^T A^T$

(iv) $(rA)^T = r(A^T)$

Properties (i), (ii), and (iv) are easy to verify. We verify (iii) below.

Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be an $n \times p$ matrix. By the row column rule,

$$(B^T A^T)_{ij} = b_{1i} a_{j1} + \cdots + b_{ni} a_{jn}.$$

On the other hand, the (i,j)-entry of $(AB)^T$ is the (j,i)-entry of AB, which is

$$(AB)_{ji} = a_{j1} b_{1i} + a_{j2} b_{2i} + \cdots + a_{jn} b_{ni}.$$

Next time

In the next lecture we will:

- Define the inverse of a matrix.
- Use matrix inversion to solve matrix equations.
- Study algebraic properties of the matrix inverse.

Chapter 2: Matrix algebra
§2.2 The inverse of a matrix

MTH 222

Linear Algebra

Goals for today:

- Define the inverse of a matrix.
- Use matrix inversion to solve matrix equations.
- Study algebraic properties of the matrix inverse.

Matrix inverse

Recall that the $n \times n$ identity matrix I_n is the diagonal matrix with 1s along the diagonal. For any $m \times n$ matrix A, we have $I_m A = A I_n = A$.

Definition 1

Let A be an $m \times n$ matrix.

- A is said to be *left invertible* if there exists an $n \times m$ matrix C such that $CA = I_n$.
- A is said to be *right invertible* if there exists an $n \times m$ matrix D such that $AD = I_m$.
- A is said to be *invertible* if it is both left and right invertible.

An invertible matrix must be square ($n \times n$). We will prove this later but it essentially comes down to the idea that invertible matrices correspond to linear transformations that are isomorphisms.

Example 2

Let $A = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & 5 \\ 3 & -2 \end{bmatrix}$.

Then $CA = I_2$ and $AC = I_2$, so C is both the left and right inverse of A.

Theorem 3

Let A be an invertible $n \times n$ matrix. Then the left and right inverse of A are the same and the inverse is unique.

Proof.

Let C be the left inverse of A and D the right inverse. Then

$$C = CI_m = C(AD) = (CA)D = I_nD = D.$$

Hence, $C = D$. Similarly, suppose B and C are inverses of A. Then

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

Hence, $B = C$. □

If A is invertible, we denote *the* inverse of A as A^{-1}.

The 2×2 case

Theorem 4

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *If* $ad - bc \neq 0$, *then* A *is invertible and*

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

If $ad - bc = 0$, *then* A *is singular.*

The value $ad - bc$ in Theorem **??** is called the *determinant* of A. We will learn more about determinants in Chapter 3.

We do not prove Theorem **??** as we will prove a more general theorem later. For now, let's see how to apply this

Example 5

Consider the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

We set up and solve the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right], \quad \mathbf{b} = \left[\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right].$$

By Theorem **??**,

$$A^{-1} = \frac{1}{(3)(6) - (4)(5)} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \frac{-1}{2} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix}.$$

So,

$$\mathbf{x} = I_2\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} = \left[\begin{smallmatrix} 5 \\ -3 \end{smallmatrix}\right].$$

The next theorem generalizes the previous example.

Theorem 6

If A is an $n \times n$ invertible matrix, then for each $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof.

If A is invertible, then clearly,

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}.$$

Hence, $A^{-1}\mathbf{b}$ is a solution. We claim this solution is unique. Suppose $\mathbf{u}$ is another solution, then $A\mathbf{u} = \mathbf{b} = A(A^{-1}\mathbf{b})$. Multiplying both sides by A^{-1} gives,

$$\begin{aligned} A^{-1}(A\mathbf{u}) &= A^{-1}(A(A^{-1}\mathbf{b})) \\ \Rightarrow (A^{-1}A)\mathbf{u} &= (A^{-1}A)(A^{-1}\mathbf{b}) \\ \Rightarrow I_n\mathbf{u} &= I_n(A^{-1}\mathbf{b}) \\ \Rightarrow \mathbf{u} &= A^{-1}\mathbf{b}. \end{aligned}$$

Hence, the solution is unique. $\square$

Theorem 7

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n.

Proof.

Suppose that A is invertible. Then the equation $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^n$. This is equivalent (by an earlier theorem) to the statement that A has a pivot in each row. Since A is square, this is equivalent to A having n pivots. Thus, $RREF(A)$ has n pivots, so $RREF(A) = I_n$.

On the other hand, if A is row equivalent to I_n, then $A\mathbf{x} = \mathbf{e}_1$ has a unique solution. Let B be the $n \times n$ matrix whose first column is this unique solution. In general, let B be the $n \times n$ matrix whose kth column is the unique solution to $A\mathbf{x} = \mathbf{e}_k$. One verifies that $B = A^{-1}$. □

The process of finding the inverse is also very much related to row reduction.

Theorem 8

Suppose A is an invertible $n \times n$ matrix. Any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

We'll put off the proof of Theorem **??** for a bit because it requires more machinery. What it tells us is that to find A^{-1}, then we form the augmented matrix $[A \mid I_n]$ and row reduce A to I_n. The result will be the augmented matrix $[I_n \mid A^{-1}]$.

Example 9

Find the inverse of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$.

We form the augmented matrix $[A|I_3]$ and row reduce A to I_3.

$$[A|I_3] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right]$$

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

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right].$$

Thus, $A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$.

Algebraic properties of the inverse

Theorem 10

Let A and B be invertible $n \times n$ matrices.

(1) A^{-1} is invertible and $(A^{-1})^{-1} = A$.

(2) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

(3) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Proof.

(1) follows from the symmetry in the definition of A^{-1}.

(2) By uniqueness, it suffices to show that $B^{-1}A^{-1}$ satisfies the definition of the inverse for AB. By associativity,

$$(AB)(B^{-1}A^{-1}) = (A(BB^{-1}))A^{-1} = (AI_n)A^{-1} = AA^{-1} = I_n.$$

Similarly, $(B^{-1}A^{-1})(AB) = I_n$.

(3) Since $(A^T)(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$ and similarly $(A^{-1})^T A^T = I_n$, then the result follows. $\square$

We now set up some technology to prove Theorem **??**.

Definition 11

An *elementary matrix* is one obtained by performing a single row operation to an identity matrix.

Example 12

The following are elementary matrices.

- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$, Repl $R_3 + (-4)R_1$
- $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, IC $R_1 \leftrightarrow R_2$
- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$, Scale $5R_3$

Lemma 13

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Proof.

Let E be the elementary matrix obtained by multiplying row k of I_n by $m \neq 0$. That is,

$$E = \begin{bmatrix} \mathbf{e}_1 & \cdots & m\mathbf{e}_k & \cdots & \mathbf{e}_n \end{bmatrix}.$$

Let A be any $n \times n$ matrix. Then,

$$EA = E \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} E\mathbf{a}_1 & \cdots & E\mathbf{a}_n \end{bmatrix}.$$

It is clear that the result of the multiplication $E\mathbf{a}_i$ is the multiply the kth entry of $\mathbf{a}_i$ by m. Thus, the resulting matrix EA is the same as A, except each entry in the kth row is multiplied by m.

The proofs for the other two row operations are similar. $\square$

Lemma 14

Each elementary matrix is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

Proof.

Let E be an elementary matrix and F the elementary matrix that transforms E back into I. Clearly such an F exists because every row operation is reversible. Moreover, E must reverse F. By Lemma **??**, $FE = I$ and $EF = I$, so F is the inverse of E. □

Theorem ??

Suppose A is an invertible $n \times n$ matrix. Any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Proof.

Since A is invertible, then $A \sim I_n$. There exists a series of row operations which transform A into I_n (Theorem **??**). Denote the corresponding elementary matrices by $E_1, \ldots, E_p$. That is, $E_p(E_{p-1} \cdots E_1 A)) = I_n$, or $E_p \cdots E_1 A = I_n$. Since the product of invertible matrices is invertible, then $A = (E_p \cdots E_1)^{-1}$. Since the inverse of an invertible matrix is invertible, it follows that A is invertible and the inverse is $E_p \cdots E_1$. □

Next time

In the next lecture we will:

- Connect the idea of invertible matrices with several other concepts we have discussed so far this semester.
- Study invertible linear transformations through invertible matrices.

Chapter 2: Matrix algebra
§2.3 Characterizations of invertible matrices

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Connect the idea of invertible matrices with several other concepts we have discussed so far this semester.
- Study invertible linear transformations through invertible matrices.

Recalling earlier results

We begin by recalling some earlier results.

Theorem 1

Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Theorem 2

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is 1-1 if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Theorem 3

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A.

(1) T is onto if and only if the columns of A span $\mathbb{R}^m$.

(2) T is 1-1 if and only if the columns of A are linearly independent.

Recalling earlier results

Theorem 4

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n. If A is an invertible $n \times n$ matrix, then any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Theorem 5

Let A be an invertible $n \times n$ matrix. Then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Theorem 6 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following are equivalent.

(1) *A is invertible.*

(2) *A is row equivalent to I_n.*

(3) *A has n pivot positions.*

(4) *The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

(5) *The columns of A form a linearly independent set.*

(6) *The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is 1-1.*

(7) *The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.*

(8) *The columns of A span $\mathbb{R}^n$.*

(9) *The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.*

(10) *There exists an $n \times n$ matrix C such that $CA = I_n$.*

(11) *There exists an $n \times n$ matrix D such that $AD = I_n$.*

(12) *A^T is invertible.*

Theorem 6 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following are equivalent.

(1) *A is invertible.*

(2) *A is row equivalent to I_n.*

(3) *A has n pivot positions.*

(4) *The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

(6) *The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is 1-1.*

(12) *A^T is invertible.*

Proof.

(1) $\Leftrightarrow$ (2) by Theorem 4, (4) $\Leftrightarrow$ (6) by Theorem 2, and (1) $\Leftrightarrow$ (12) by Theorem 5.

(2) $\Leftrightarrow$ (3) is clear because I_n has n pivots and row operations do not change the number of pivots.

(2) $\Leftrightarrow$ (4) because row operations do not change the solution set and the only solution to $\mathbf{x} = I_n\mathbf{x} = \mathbf{0}$ is $\mathbf{0}$.

The invertible matrix theorem

Theorem 6 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following are equivalent.

(3) A has n pivot positions.

(5) The columns of A form a linearly independent set.

(6) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is 1-1.

(7) The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.

(8) The columns of A span $\mathbb{R}^n$.

(9) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.

Proof.

(5) $\Leftrightarrow$ (6) and (8) $\Leftrightarrow$ (9) by Theorem 3.

(3) $\Leftrightarrow$ (7) $\Leftrightarrow$ (8) by Theorem 1.

Thus, (1) $-$ (9) are all equivalent.

Theorem 6 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following are equivalent.

(1) A is invertible.

(10) There exists an $n \times n$ matrix C such that $CA = I_n$.

(11) There exists an $n \times n$ matrix D such that $AD = I_n$.

Proof.

Clearly, (1) implies (10) and (11). If (10) holds, then multiplying both sides of the matrix equation $A\mathbf{x} = \mathbf{0}$ (on the left) by C

$$(CA)\mathbf{x} = C\mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}.$$

So $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Thus, (10) $\Rightarrow$ (4).

Suppose (11) holds and let $\mathbf{b} \in \mathbb{R}^n$. Then

$$A(D\mathbf{b}) = (AD)\mathbf{b} = I_n\mathbf{b} = \mathbf{b},$$

so the matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^n$. Thus, (11) $\Rightarrow$ (7). $\square$

Definition 7

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *invertible* if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $S(T(\mathbf{x})) = \mathbf{x} = T(S(\mathbf{x}))$. We call $S = T^{-1}$ the *inverse* of T.

Theorem 8

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A. Then T is invertible if and only if A is an invertible matrix. In that case, $S : \mathbb{R}^n \to \mathbb{R}^n$ given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique linear transformation such that $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof.

Suppose T is invertible with inverse S. We claim A is invertible. Let $\mathbf{b} \in \mathbb{R}^n$ and set $\mathbf{x} = S(\mathbf{b})$. Then $T(\mathbf{x}) = T(S(\mathbf{b})) = \mathbf{b}$, so T is onto. Hence, by the Invertible Matrix Theorem, A is invertible.

Definition 7

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *invertible* if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $S(T(\mathbf{x})) = \mathbf{x} = T(S(\mathbf{x}))$. We call $S = T^{-1}$ the *inverse* of T.

Theorem 8

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A. Then T is invertible if and only if A is an invertible matrix. In that case, $S : \mathbb{R}^n \to \mathbb{R}^n$ given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique linear transformation such that $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof.

Conversely, suppose A is invertible. Define $S : \mathbb{R}^n \to \mathbb{R}^n$ by $S(\mathbf{x}) = A^{-1}\mathbf{x}$. Then

$$S(T(\mathbf{x})) = S(A\mathbf{x}) = A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}$$
$$T(S(\mathbf{x})) = T(A^{-1}\mathbf{x}) = A(A^{-1}\mathbf{x}) = (AA^{-1})\mathbf{x} = I_n\mathbf{x} = \mathbf{x}.$$

Hence, S is the inverse of T. Uniqueness follows from uniqueness of A^{-1}. □

Next time

In the next lecture we will:

- Define a partition of a matrix and use this to simplify some computations on big matrices.
- Study the Liontief Input-Output model for balancing an economy.

Chapter 2: Matrix algebra
§2.4 Partitioned matrices
§2.6 The Leontief Input-Output Model

MTH 222

Linear Algebra

Goals for today:

- Define a partition of a matrix and use this to simplify some computations on big matrices.
- Study the Liontief Input-Output model for balancing an economy.

Partitioning a matrix

Definition 1

A *partition* of a matrix A is a decomposition of A into rectangular submatrices $A_1, \ldots, A_n$ such that each entry in A lies in some submatrix.

Example 2

The matrix

$$A = \begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 1 & 0 & 7 \\ 4 & 5 & 2 & 1 \end{bmatrix}$$

can be partitioned as

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

where $A_{11} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 0 \end{bmatrix}$, $A_{12} = \left[\begin{smallmatrix} 4 \\ 7 \end{smallmatrix}\right]$, $A_{21} = \begin{bmatrix} 4 & 5 & 2 \end{bmatrix}$, and $A_{22} = \begin{bmatrix} 1 \end{bmatrix}$.

There are many other partitions of this matrix and in general a matrix will have several partitions. I leave it as an exercise to write another partition of this matrix.

Adding partitioned matrices

If A and B are matrices of the same size and partition in the same way, then one may obtain $A + B$ by adding corresponding partitions. Similarly, we can multiply A and B as partitioned matrices so long as multiplication between the blocks makes sense.

Example 3

Let A be as in Example 2 and

$$B = \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] \text{ where } B_1 = \begin{bmatrix} 1 & 3 \\ 0 & 7 \\ 2 & 0 \end{bmatrix} \text{ and } B_2 = \begin{bmatrix} -1 & -3 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix} = \begin{bmatrix} -5 & 12 \\ -5 & -8 \\ 7 & 44 \end{bmatrix}.$$

Horizontal and vertical partitions

One way to partition a matrix is with only vertical or only horizontal lines. In this way we can denote

$$A = \begin{bmatrix} \text{col}_1(A) & \text{col}_2(A) & \cdots & \text{col}_n(A) \end{bmatrix} \text{ or } B = \begin{bmatrix} \text{row}_1(B) \\ \text{row}_2(B) \\ \cdots \\ \text{row}_n(B) \end{bmatrix}.$$

Theorem 4 (Column-row expansion of AB)

If A is $m \times n$ and B is $n \times p$ then

$$AB = \text{col}_1(A)\,\text{row}_1(B) + \cdots + \text{col}_n(A)\,\text{row}_n(B).$$

Block diagonal matrices

We say a matrix A is block diagonal if

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & A_n \end{bmatrix}.$$

where each A_i is a square matrix. If each A_i is invertible, then A is invertible and

$$A^{-1} = \begin{bmatrix} A_1^{-1} & 0 & \cdots & 0 \\ 0 & A_2^{-1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & A_n^{-1} \end{bmatrix}.$$

Example 5

Let $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

We can find A^{-1} without using row reduction. We have

$$A_1 = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad A_1^{-1} = \frac{1}{2}\begin{bmatrix} 0 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1/2 & 1/2 \end{bmatrix}$$
$$A_2 = 3 \quad \text{and} \quad A_2^{-1} = 1/3.$$

Hence,

$$A^{-1} = \begin{bmatrix} A_1^{-1} & 0 \\ 0 & A_2^{-1} \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

The Leontief Input-Output Model

Suppose we divide the economy of the US into *n sectors* that produce or output goods and services. We can then denote the output of each sector as a *production vector* $\mathbf{x} \in \mathbb{R}^n$. Another part of the economy only consumes. We denote this by a *final demand vector* $\mathbf{d}$ that lists the values of the goods and services demanded by the nonproductive part of the economy.

In between these two are the goods and services needed by the producers to create the products that are in demand. Optimally, one would balance all aspects of the production, but this is very complex. A model for finding this balance is due to Wassily Leontief. Ideally, one would find a production level $\mathbf{x}$ such that the amounts produced will balance the total demand, so that

$$\{ \text{ amt produced } \mathbf{x} \ \} = \{ \text{ int demand } \} + \{ \text{ final demand } \mathbf{d} \ \}.$$

The model assumes that for each sector, there is a unit consumption vector in $\mathbb{R}^n$ that lists the inputs needed per unit of output in the sector (in millions of dollars).

Example 6

Suppose the economy consists of three sectors: manufacturing, agriculture, and services, with unit consumption vectors c_1, c_2, and c_3, respectively.

	Inputs consumed per unit of output		
Purchased from:	Manufacturing	Agriculture	Services
Manufacturing	.50	.40	.20
Agriculture	.20	.30	.10
Services	.10	.10	.30
	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$

We have $100\mathbf{c}_1 = \begin{bmatrix} 50 \\ 20 \\ 10 \end{bmatrix}$. Hence, to produce 100 units of output, manufacturing will order (or *demand*) 50 units from other parts of manufacturing, 20 units from agriculture, and 10 units from services.

Example 6

Suppose the economy consists of three sectors: manufacturing, agriculture, and services, with unit consumption vectors c_1, c_2, and c_3, respectively.

	Inputs consumed per unit of output		
Purchased from:	Manufacturing	Agriculture	Services
Manufacturing	.50	.40	.20
Agriculture	.20	.30	.10
Services	.10	.10	.30
	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$

If manufacturing produces $\mathbf{x}_1$ units of output, then $\mathbf{x}_1\mathbf{c}_1$ represents the *intermediate demands* of manufacturing, because the amounts in $\mathbf{x}_1\mathbf{c}_1$ will be consumed in the process of creating $\mathbf{x}_1$ units of output. Similarly, $\mathbf{x}_2\mathbf{c}_2$ and $\mathbf{x}_3\mathbf{c}_3$ list the corresponding intermediate demands of agriculture and services, respectively. Hence,

$$\{\text{ int demand }\} = \mathbf{x}_1\mathbf{c}_1 + \mathbf{x}_2\mathbf{c}_2 + \mathbf{x}_3\mathbf{c}_3 = C\mathbf{x}$$

where $C = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3\end{bmatrix}$ is the *consumption matrix*. Here, $C = \begin{bmatrix} .50 & .40 & .20 \\ .20 & .30 & .10 \\ .10 & .10 & .30 \end{bmatrix}$

Using notation from our example, we can rewrite our model as

$$\underset{\text{amt produced}}{\mathbf{x}} = \underset{\text{int demand}}{C\mathbf{x}} + \underset{\text{final demand}}{\mathbf{d}}$$

This can be rewritten as $(I - C)\mathbf{x} = \mathbf{d}$ where I is the $n \times n$ identity matrix.

Example 6 (continued)

Suppose the final demand is 50 units for manufacturing, 30 units for agriculture, and 20 units for services. We will find the production level $\mathbf{x}$ that satisfies this demand.

This amounts to row reducing the augmented matrix $[I - C \mid \mathbf{d}]$ where $\mathbf{d}$ is given above. We can compute $I - C$ directly and find that

$$[I - C \mid \mathbf{d}] = \begin{bmatrix} .50 & -.40 & -.20 & 50 \\ -.20 & .70 & -.10 & 30 \\ -.10 & -.10 & .70 & 20 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 226 \\ 0 & 1 & 0 & 119 \\ 0 & 0 & 1 & 78 \end{bmatrix}.$$

How else might we solve this problem?

Theorem 7

Let C be the consumption matrix for an economy, and let $\mathbf{d}$ be the final demand. If C and $\mathbf{d}$ have nonnegative entries and if each column sum of C is less than 1, then $(I - C)^{-1}$ exists and the production vector

$$\mathbf{x} = (I - C)^{-1}\mathbf{d}$$

has nonnegative entries and is the unique solution of $\mathbf{x} = C\mathbf{x} + \mathbf{d}$.

In a 3×3 system, finding the inverse is not difficult, though a little tedious and we'd repeat the row reduction above. However, in the context of this theorem there is another way to get the inverse, or at least a close approximation of it.

Suppose the initial demand is set at $\mathbf{d}$. This creates an intermediate demand of $C\mathbf{d}$. To meet this intermediate demand, industries will need *more* input. This creates additional intermediate demands of $C(C\mathbf{d}) = C^2\mathbf{d}$. This process repeats (forever?) and in the next round we have $C(C^2\mathbf{d}) = C^3\mathbf{d}$. Hence, the production level that will meet this demand is

$$\mathbf{d} + C\mathbf{d} + C^2\mathbf{d} + C^3\mathbf{d} + \cdots = (I + C + C^2 + C^3 + \cdots)\mathbf{d}.$$

By our hypotheses on C, $C^m \to 0$ as $m \to \infty$. Hence,

$$(I - C)^{-1} \approx I + C + C^2 + \cdots + C^m$$

where the approximation may be made as close to $(I - C)^{-1}$ as desired by taking m sufficiently large.

Question

Under what conditions will this ever be exact?

The i, j entry in $(I - C)^{-1}$ is the increased amounts that sector i will have to produce to satisfy an increase of 1 unit in the final demand from sector j.

Next time

In the next lecture we will:

- Define subspaces of $\mathbb{R}^n$.
- Study the column space and null space of a matrix and relate to homogenous systems.
- Define a basis of a subspace and practice finding bases for column and null spaces.

Chapter 2: Matrix algebra
§2.8 Subspaces of $\mathbb{R}^n$

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define subspaces of $\mathbb{R}^n$.
- Study the column space and null space of a matrix and relate to homogenous systems.
- Define a basis of a subspace and practice finding bases for column and null spaces.

Subspaces are subsets of $\mathbb{R}^n$ that are closed under the vector operations we discussed in Chapter 1. This discussion moves us closer to being able to discuss abstract vector spaces later in the course.

Definition 1

A *subspace* of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ such that

(1) $\mathbf{0} \in H$;

(2) If $\mathbf{u}, \mathbf{v} \in H$, then $\mathbf{u} + \mathbf{v} \in H$;

(3) If $\mathbf{u} \in H$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in H$.

Example 2

Consider the set $\{\mathbf{0}\}$ (the set containing only the zero vector for some n). Obviously, $\mathbf{0} \in \{\mathbf{0}\}$. Since $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $c\mathbf{0} = \mathbf{0}$ for all $c \in \mathbb{R}$, then $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ (for every n.) We call this the *trivial subspace* of $\mathbb{R}^n$.

Example 3

$\mathbb{R}^n$ is a subspace of itself. We call any subspace of $\mathbb{R}^n$ that is not $\mathbb{R}^n$ itself a *proper subspace*.

Example 4

Let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ and let $H = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. Then $\mathbf{0} = 0\mathbf{v}_1 \in H$, as is any linear combination of the $\mathbf{v}_i$. Hence, H is a subspace of $\mathbb{R}^n$.

The first condition of a subspace actually be be replaced by the weaker condition that $H \neq \emptyset$. Suppose that $H \neq \emptyset$ is closed under taking scalar multiples and $\mathbf{v} \in H$. Then $\mathbf{0} = 0\mathbf{v} \in H$. However, checking that $\mathbf{0} \in H$ is a convenient way to verify that a set is a subspace.

Example 5

Let L be a line in $\mathbb{R}^n$ *not* through the origin. Then L is not a subspace of $\mathbb{R}^n$ since it does not contain $\mathbf{0}$.

Example 6

Define a set H in $\mathbb{R}^2$ by the following property: $\mathbf{v} \in H$ if $\mathbf{v}$ has exactly one nonzero entry. Then H is not a subspace because it does not contain $\mathbf{0}$.

Next we'll define two subspaces related to matrices (and hence also linear transformations). Though they are defined in different ways, we'll see next soon how they are connected.

Definition 7

Let A be an $m \times n$ matrix. The *column space* of a matrix A is the set $\mathrm{Col}(A)$ of all linear combinations of the columns of A. The *null space* is the set $\mathrm{Nul}(A)$ of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

If $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ with $\mathbf{a}_i \in \mathbb{R}^m$, then $\mathrm{Col}(A) = \mathrm{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ so $\mathrm{Col}(A)$ is a subspace of $\mathbb{R}^m$. We will show soon that $\mathrm{Nul}(A)$ is a subspace of $\mathbb{R}^n$.

Example 8

Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Is $\mathbf{b} \in \mathrm{Col}(A)$? What is $\mathrm{Nul}(A)$?

This is the equivalent to asking whether **b** is in the span of the columns of A. So we form the corresponding augmented matrix and row reduce:

$$\begin{bmatrix} 1 & -3 & -4 & 3 \\ -4 & 6 & -2 & 3 \\ -3 & 7 & 6 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 5 & -9/2 \\ 0 & 1 & 3 & -5/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the system is consistent, then $\mathbf{b} \in \mathrm{Col}(A)$.

For $\mathrm{Nul}(A)$, we consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. We row reduce, but it is similar as above (but with zeros in the last column). This means that the solution in parametric vector form is

$$\mathbf{x} = \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix} x_3.$$

It follows that $\mathrm{Nul}(A) = \mathrm{Span}\left\{ \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix} \right\}$.

Null space

Theorem 9

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$.

Proof.

By definition, $\text{Nul}(A)$ is a set in $\mathbb{R}^n$. Clearly $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in \text{Nul}(A)$.

Let $\mathbf{x}, \mathbf{y} \in \text{Nul}(A)$ so $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. Then $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus, $\mathbf{x} + \mathbf{y} \in \text{Nul}(A)$.

Finally, suppose $\mathbf{x} \in \text{Nul}(A)$ and $c \in \mathbb{R}$. Then $A(c\mathbf{x}) = c(A\mathbf{x}) = c\mathbf{0} = \mathbf{0}$, so $c\mathbf{x} \in \text{Nul}(A)$.

Thus, $\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$. $\square$

Definition 10

A *basis* for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.

If A is $n \times n$ and invertible then the columns of A are a basis for $\mathbb{R}^n$ because they are linearly independent and span $\mathbb{R}^n$.

Recall that $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are called the standard basis vectors for $\mathbb{R}^n$ (now you know why).

Theorem 11

Any basis of $\mathbb{R}^n$ has size n.

Proof.

Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ be a basis for $\mathbb{R}^n$ and let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix}$. Since the columns of A span $\mathbb{R}^n$, then there must be a pivot in every row. Since the columns of A are linearly independent, then there must be a pivot in every column. This implies that A has n rows and columns, so $p = n$. $\square$

Example 12

Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & 4 \end{bmatrix}$. Then A row reduces to

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

Hence $\mathrm{Col}(A) = \mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ because those are pivot columns. It turns out that $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ is a basis for Col(A) (why?). On the other hand, the solution space of the matrix equation $A\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} x_4.$$

Thus, $\mathrm{Nul}(A) = \mathrm{Span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right\}$. The set $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for Nul(A) (why?).

Theorem 13

The pivot columns of a matrix A form a basis for the column space of A.

Proof.

Row reduction does not change the linear relations between the columns. Each pivot column corresponds to a standard basis vector in RREF(A) and hence the pivot columns are linearly independent.

Let $\mathbf{b}$ be a non-pivot column of A. Since $\mathbf{b}$ does not have a pivot, then all of its nonzero entries lie to the right of the pivots. Consequently, it is a linear combination of the pivot columns to its left. Thus, the pivot columns span Col(A). $\square$

Next time

In the next lecture we will:

- Define dimension for subspaces.
- Determine the dimension of Col(A) and Nul(A). Use this to state the Rank-Nullity Theorem.
- Define the coordinates of a vector relative to a basis and state the Basis Theorem.

Chapter 2: Matrix algebra
§2.9 Dimension and Rank

MTH 222

Linear Algebra

Goals for today:

- Define dimension for subspaces.
- Determine the dimension of $\mathrm{Col}(A)$ and $\mathrm{Nul}(A)$. Use this to state the Rank-Nullity Theorem.
- Define the coordinates of a vector relative to a basis and state the Basis Theorem.

Definition 1

The *dimension* of a nonzero subspace H of $\mathbb{R}^n$, denoted $\dim H$, is the number of vectors in any basis of H. The dimension of the zero subspace is defined to be zero.

At the end of this section we will prove that dimension is well-defined (that is, all bases have the same dimension). We have already proved this for $\mathbb{R}^n$.
We now specialize the above to the column and null space of a matrix.

Definition 2

The *rank* of a matrix A is the dimension of the column space of A:

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

The *nullity* of a matrix A is the dimension of the null space of A:

$$\operatorname{nul} A = \dim \operatorname{Nul}(A).$$

Example 3

Let $A = \begin{bmatrix} 1 & 3 & 2 & -6 \\ 3 & 9 & 1 & 5 \\ 2 & 6 & -1 & 9 \\ 5 & 15 & 0 & 14 \end{bmatrix}$.

Then A row reduces to the matrix

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are three pivot colums, so rank $A = 3$. On the other hand, there is one free variable (x_2), and so nul $A = 1$.

Theorem 4 (Rank-Nullity Theorem)

If a matrix A has n columns, then $\operatorname{rank} A + \operatorname{nul} A = n$.

Proof.

The rank of A ($\operatorname{rank} A$) is the number of pivot columns in A and the nullity of A ($\operatorname{nul} A$) is the number of columns corresponding to free variables. Since each columns is exactly one of these, the result follows. $\square$

Theorem 5 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following are equivalent to the statement that A is invertible.

- *The columns of A form a basis of $\mathbb{R}^n$.*
- $\operatorname{Col}(A) = \mathbb{R}^n$.
- $\dim \operatorname{Col}(A) = n$.
- $\operatorname{rank} A = n$.
- $\operatorname{Nul}(A) = \{\mathbf{0}\}$.
- $\operatorname{nul} A = 0$.

Unique expressions in terms of bases

Theorem 6

Let H be a subspace of $\mathbb{R}^n$ with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$. If $\mathbf{x} \in H$, then $\mathbf{x}$ can be written in exactly one way as a linear combination of basis vectors in $\mathcal{B}$.

Proof.

Suppose $\mathbf{x}$ can be written in two ways:

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p = d_1\mathbf{b}_1 + \cdots + d_p\mathbf{b}_p,$$

with $c_i, d_i \in \mathbb{R}$. Taking differences we find

$$\mathbf{0} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_p - d_p)\mathbf{b}_p.$$

By linear independence, $c_i - d_i = 0$ for all $i = 1, \ldots, p$. That is, $c_i = d_i$ for $i = 1, \ldots, p$. □

We now want to set up the theorem that says the notion of dimension is well-defined.

Definition 7

Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ is a basis for a subspace H of $\mathbb{R}^n$. For each $\mathbf{x} \in H$, the *coordinates of* $\mathbf{x}$ *relative to* B are the weights $c_1, \ldots, c_p$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p.$$

The *coordinate vector of* $\mathbf{x}$ *relative to* $\mathcal{B}$ (or the $\mathcal{B}$-coordinate vector) is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

The map $T_{\mathcal{B}} : H \to \mathbb{R}^p$ given by $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is called a *coordinate mapping of* $\mathcal{B}$.

Example 8

Let $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$.

To find $[\mathbf{x}]_{\mathcal{B}}$, we just need to know how to write $\mathbf{x}$ as a linear combination of the vectors in $\mathcal{B}$. But we already know how to do this:

$$\begin{bmatrix} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

If we write the vectors of $\mathcal{B}$ as $\mathbf{b}_1$ and $\mathbf{b}_2$, respectively, the $\mathbf{x} = 2\mathbf{b}_1 + 3\mathbf{b}_2$. In terms of coordinates,

$$[\mathbf{x}]_{\mathcal{B}} = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right].$$

Theorem 9

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis for a subspace H of $\mathbb{R}^n$. The coordinate mapping $T_\mathcal{B} : H \to \mathbb{R}^p$ is a 1-1 and onto (bijective) linear transformation.

Proof.

By Theorem 6, $T_\mathcal{B}$ is well-defined. We show first that $T_\mathcal{B}$ is linear. Let $\mathbf{x}, \mathbf{y} \in H$. Then

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p, \quad \text{and} \quad \mathbf{y} = d_1\mathbf{b}_1 + \cdots + d_p\mathbf{b}_p,$$

for some scalars $c_i, d_i \in \mathbb{R}$. So, $\mathbf{x} + \mathbf{y} = (c_1 + d_1)\mathbf{b}_1 + \cdots + (c_p + d_p)\mathbf{b}_p$. Now,

$$T_\mathcal{B}(\mathbf{x} + \mathbf{y}) = \begin{bmatrix} c_1+d_1 \\ \vdots \\ c_p+d_p \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_p \end{bmatrix} = T_\mathcal{B}(\mathbf{x}) + T_\mathcal{B}(\mathbf{y}).$$

Similarly, if $\lambda \in \mathbb{R}$, then

$$T_\mathcal{B}(\lambda\mathbf{x}) = \begin{bmatrix} \lambda c_1 \\ \vdots \\ \lambda c_p \end{bmatrix} = \lambda \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = \lambda T_\mathcal{B}(\mathbf{x}).$$

Theorem 9

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis for a subspace H of $\mathbb{R}^n$. The coordinate mapping $T_{\mathcal{B}} : H \to \mathbb{R}^p$ is a 1-1 and onto (bijective) linear transformation.

Proof.

To show injectivity, let $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p \in H$ such that $T_{\mathcal{B}}(\mathbf{x}) = \mathbf{0}$. Then

$$\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = [\mathbf{x}]_{\mathcal{B}} = \mathbf{0},$$

so $c_i = 0$ for all i. Hence, $\mathbf{x} = \mathbf{0}$ and $T_{\mathcal{B}}$ is 1-1.

To prove surjectivity, assume

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} \in \mathbb{R}^p.$$

Then $\mathbf{y} = a_1\mathbf{b}_1 + \cdots + a_p\mathbf{b}_p \in H$ because $\mathcal{B}$ spans H and $T_{\mathcal{B}}(\mathbf{y}) = [\mathbf{y}]_{\mathcal{B}} = \mathbf{a}$. Hence $T_{\mathcal{B}}$ is onto. $\square$

Theorem 10

Let H be a subspace of $\mathbb{R}^n$ with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$. Then every basis of H has p elements.

Proof.

Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset H$ with $m > p$. We claim S is linearly dependent (and hence cannot be a basis). Note that the set $\{[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_m]_{\mathcal{B}}\}$ is linearly dependent in $\mathbb{R}^p$ because $m > p$. Hence, there exist scalars $c_1, \ldots, c_m$ (not all zero) such that

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_m[\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

Because the coordinate mapping is a linear transformation, this implies

$$[c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

The coordinate mapping is 1-1 and so $c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0}$. Hence, the set is linearly dependent and so any basis of H can have at most p elements.

Suppose $\mathcal{B}'$ is another basis with k elements. By the above, $k \leq p$. Switching the roles of $\mathcal{B}'$ and $\mathcal{B}$ we see that $p \leq k$. This implies $p = k$ so every basis has p elements. $\square$

Theorem 11 (The Basis Theorem)

Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of exactly p elements in H is automatically a basis for H. Also, any set of p elements of H that spans H is automatically a basis for H.

Proof.

Since H has dimension p, then there exists a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$.

Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be a set of p linearly independent elements in H. If $\mathcal{U}$ spans H, then $\mathcal{U}$ is a basis. Otherwise, there exists some element $\mathbf{u}_{p+1} \notin \operatorname{Span} \mathcal{U}$. Hence, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{u}_{p+1}\}$ is linearly independent. We can continue this process, but not indefinitely. The number of linearly independent elements in the set cannot exceed p. Thus, our set will eventually span all of H, and hence will be a basis. But a basis for H cannot have more than p elements by Theorem 10. Hence, $\mathcal{U}$ must have been a basis to begin with.

Theorem 11 (The Basis Theorem)

Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of exactly p elements in H is automatically a basis for H. Also, any set of p elements of H that spans H is automatically a basis for H.

Proof.

Now suppose $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a set of p elements in H that span H. If the set is not linearly independent, then one of the vectors is a linear combination of the others. Hence, we can remove that element, say $\mathbf{v}_1$ and the span of $\{\mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is the same as the span of $\mathcal{V}$. Continue in this way until the remaining elements are linearly independent. But then the set is linearly independent and spans H, hence is a basis. But by Theorem 11, every basis has exactly p elements, a contradiction unless $\mathcal{V}$ was a basis to begin with. □

Another way of viewing the previous theorem is the following: The first part says that any linearly independent set in H can be extended to a basis. The second part says that any spanning set contains a basis.

Next time

In the next lecture we will:

- Determine the change of basis matrix that transforms coordinates in one basis to coordinates in another.

Chapter 2: Matrix algebra
§Change of Basis

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Determine the change of basis matrix that transforms coordinates in one basis to coordinates in another.

First we recall our definition of coordinates and the coordinate transformation.

Definition 1

Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ is a basis for a subspace H of $\mathbb{R}^n$. For each $\mathbf{x} \in H$, the *coordinates of* $\mathbf{x}$ *relative to* B are the weights $c_1, \ldots, c_p$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p.$$

The *coordinate vector of* $\mathbf{x}$ *relative to* $\mathcal{B}$ (or the $\mathcal{B}$-coordinate vector) is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

The map $T_{\mathcal{B}} : H \to \mathbb{R}^p$ given by $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is called a *coordinate mapping of* $\mathcal{B}$.

Example 2

Let $\mathcal{B} = \{\left[\begin{smallmatrix}1\\5\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\4\end{smallmatrix}\right]\}$ and $\mathbf{x} = \left[\begin{smallmatrix}3\\-1\end{smallmatrix}\right]$. We will compute the standard matrix of $T_{\mathcal{B}}$ and use this to find the $\mathcal{B}$-coordinates of $\mathbf{x}$.

We need to compute $T_{\mathcal{B}}(\mathbf{e}_1)$ and $T_{\mathcal{B}}(\mathbf{e}_2)$. We know how to do this problem in general but here we'll just observe that $\mathbf{e}_1 = -4\mathbf{b}_1 + 5\mathbf{b}_2$ and $\mathbf{e}_2 = \mathbf{b}_1 - \mathbf{b}_2$. Thus,

$$T_{\mathcal{B}}(\mathbf{e}_1) = [\mathbf{e}_1]_{\mathcal{B}} = \left[\begin{smallmatrix}-4\\5\end{smallmatrix}\right] \quad \text{and} \quad T_{\mathcal{B}}(\mathbf{e}_2) = [\mathbf{e}_2]_{\mathcal{B}} = \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right].$$

Then the standard matrix of $T_{\mathcal{B}}$, denoted $P_{\mathcal{B}}$, is

$$P_{\mathcal{B}} = \begin{bmatrix} T_{\mathcal{B}}(\mathbf{e}_1) & T_{\mathcal{B}}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}, \quad \text{so} \quad T_{\mathcal{B}}(\mathbf{x}) = P_{\mathcal{B}}\mathbf{x} = \left[\begin{smallmatrix}-13\\16\end{smallmatrix}\right].$$

We can check this: $\mathbf{x} = -13\mathbf{b}_1 + 16\mathbf{b}_2 = \left[\begin{smallmatrix}3\\-1\end{smallmatrix}\right]$.

There is another (simpler!) way to do the previous example.

In the last section we proved that $T_{\mathcal{B}}$ is a 1-1 and onto (bijective) linear map and hence it is invertible with inverse $T_{\mathcal{B}}^{-1}$. The standard matrix of $T_{\mathcal{B}}^{-1}$ is $P_{\mathcal{B}}^{-1}$.

It's easy to see that $T_{\mathcal{B}}(\mathbf{b}_i) = \mathbf{e}_i$. Hence, $T_{\mathcal{B}}^{-1}(\mathbf{e}_i) = \mathbf{b}_i$ and so

$$P_{\mathcal{B}}^{-1} = \begin{bmatrix} T_{\mathcal{B}}^{-1}(\mathbf{e}_1) & \cdots & T_{\mathcal{B}}^{-1}(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}.$$

This is just the matrix whose columns are the vectors in the basis $\mathcal{B}$!

Example 3

Again let $\mathcal{B} = \{\left[\begin{smallmatrix}1\\5\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\4\end{smallmatrix}\right]\}$.

Since $P_{\mathcal{B}}^{-1} = \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix}$, then $P_{\mathcal{B}} = (P_{\mathcal{B}}^{-1})^{-1} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}$.

Change of basis

We've seen how to switch between the standard basis ($\mathcal{E}$) and a given basis. Now say we have two bases of $\mathbb{R}^n$: $\mathcal{B}$ and $\mathcal{C}$. We have the tools we need to change from the basis $\mathcal{B}$ to the basis $\mathcal{C}$. Consider the following diagram.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathcal{B}} & & \mathbb{R}^n_{\mathcal{C}} \\
& \mathbb{R}^n_{\mathcal{E}} \arrow[ul, "T_{\mathcal{B}}"'] \arrow[ur, "T_{\mathcal{C}}"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{R}^n_{\mathcal{B}} \arrow[rr, dashed, "T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}"] \arrow[dr, dashed, bend right, "T_{\mathcal{B}}^{-1}"'] & & \mathbb{R}^n_{\mathcal{C}} \\
& \mathbb{R}^n_{\mathcal{E}} \arrow[ul, "T_{\mathcal{B}}"'] \arrow[ur, "T_{\mathcal{C}}"] &
\end{tikzcd}$$

If we want to go from $\mathcal{B}$ to $\mathcal{C}$, that is, convert coordinates in $\mathcal{B}$ to coordinates in $\mathcal{C}$, we need to first go from $\mathcal{B}$ to $\mathcal{E}$ and then from $\mathcal{E}$ to $\mathcal{C}$.

Definition 4

Suppose we have bases $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ of $\mathbb{R}^n$. Then the change of basis transformation is

$$\begin{aligned} T_{\mathcal{B}\to\mathcal{C}} : \mathbb{R}^n &\to \mathbb{R}^n \\ [\mathbf{x}]_{\mathcal{B}} &\mapsto [\mathbf{x}]_{\mathcal{C}} \end{aligned}$$

is linear and is the composition $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}$ where $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ are the coordinate mappings of $\mathcal{B}$ and $\mathcal{C}$, respectively. Its standard matrix is $P_{\mathcal{B}\to\mathcal{C}} = P_{\mathcal{C}} P_{\mathcal{B}}^{-1}$ and is called the change of basis matrix.

Example 5

Let $\mathcal{B} = \{[\begin{smallmatrix}1\\5\end{smallmatrix}], [\begin{smallmatrix}1\\4\end{smallmatrix}]\}$ be as before and $\mathcal{C} = \left\{ \left[\begin{smallmatrix}-1\\3\end{smallmatrix}\right], \left[\begin{smallmatrix}-2\\5\end{smallmatrix}\right] \right\}$. We will find the change of basis matrix $P_{\mathcal{B}\to\mathcal{C}}$.

We previously computed $P_{\mathcal{B}}^{-1}$ so it is left to compute $P_{\mathcal{C}}$. We know that

$$P_{\mathcal{C}}^{-1} = \begin{bmatrix} -1 & -2 \\ 3 & 5 \end{bmatrix} \quad \text{so} \quad P_{\mathcal{C}} = (P_{\mathcal{C}}^{-1})^{-1} = \begin{bmatrix} 5 & 2 \\ -3 & -1 \end{bmatrix}.$$

Hence,

$$P_{\mathcal{B}\to\mathcal{C}} = P_{\mathcal{C}} P_{\mathcal{B}}^{-1} = \begin{bmatrix} 5 & 2 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 15 & 13 \\ -8 & -7 \end{bmatrix}.$$

Suppose $\mathbf{x} \in \mathbb{R}^2$ with $[\mathbf{x}]_{\mathcal{B}} = \left[\begin{smallmatrix}1\\-2\end{smallmatrix}\right]$. To find $[\mathbf{x}]_{\mathcal{C}}$, we apply $P_{\mathcal{B}\to\mathcal{C}}$ to $[\mathbf{x}]_{\mathcal{B}}$,

$$P_{\mathcal{B}\to\mathcal{C}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 15 & 13 \\ -8 & -7 \end{bmatrix} \left[\begin{smallmatrix}1\\-2\end{smallmatrix}\right] = \left[\begin{smallmatrix}-11\\6\end{smallmatrix}\right].$$

We can check by converting both to the standard basis,

$$\mathbf{x} = (1)\mathbf{b}_1 + (-2)\mathbf{b}_2 = \left[\begin{smallmatrix}-1\\-3\end{smallmatrix}\right] \quad \text{and} \quad \mathbf{x} = (-11)\mathbf{c}_1 + (6)\mathbf{c}_2 = \left[\begin{smallmatrix}-1\\-3\end{smallmatrix}\right].$$

Example 6

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix} \right\} \quad \text{and} \quad \mathcal{C} = \left\{ \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ 3 \\ 2 \end{bmatrix} \right\}.$$

Both $\mathcal{B}$ and $\mathcal{C}$ are bases for $\mathbb{R}^3$ (check this!). Let $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ be their respective coordinate mappings with standard matrices $P_{\mathcal{B}}$ and $P_{\mathcal{C}}$, respectively. Then

$$P_{\mathcal{B}}^{-1} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix}, P_{\mathcal{B}} = \begin{bmatrix} 15 & 2 & -6 \\ 5 & 1 & -2 \\ -7 & -1 & -3 \end{bmatrix},$$

$$P_{\mathcal{C}}^{-1} = \begin{bmatrix} 5 & 0 & 7 \\ 2 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}, P_{\mathcal{C}} = \begin{bmatrix} 3 & -7 & 0 \\ 1 & -3 & 1 \\ -2 & 5 & 0 \end{bmatrix}.$$

Thus,

$$P_{\mathcal{B} \to \mathcal{C}} = P_{\mathcal{C}} P_{\mathcal{B}}^{-1} = \begin{bmatrix} 10 & -21 & 6 \\ 6 & -8 & 7 \\ -7 & 15 & -4 \end{bmatrix}$$

Example 6 (cont.)

Write $\mathbf{x} \in \mathbb{R}^3$ with $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$. Applying our map

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} 10 & -21 & 6 \\ 6 & -8 & 7 \\ -7 & 15 & -4 \end{bmatrix}$$

gives

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{B}\to\mathcal{C}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -18 \\ -5 \\ 13 \end{bmatrix}.$$

We check by writing both in terms of the standard basis,

$$\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2 - \mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = -18\mathbf{c}_1 - 5\mathbf{c}_2 + 13\mathbf{c}_3.$$

Change of basis for transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be any linear transformation and $\mathcal{B}, \mathcal{C}$ bases of $\mathbb{R}^n$. The map $[T]_{\mathcal{B},\mathcal{C}} : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$[\mathbf{x}]_{\mathcal{B}} \mapsto [T(\mathbf{x})]_{\mathcal{C}}$$

is linear.

In fact, it is just the composition $T_{\mathcal{C}} \circ T \circ T_{\mathcal{B}}^{-1}$. If A_T is the standard matrix of T, then the standard matrix of $[T]_{\mathcal{B},\mathcal{C}}$ is $P_{\mathcal{C}} \circ A_T \circ P_{\mathcal{B}}^{-1}$. Hence, we have the diagrams.

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "T_{\mathcal{B}}"] \arrow[d, "T"'] & \mathbb{R}^n_{\mathcal{B}} \arrow[d, "T_{\mathcal{B},\mathcal{C}}"] \\
\mathbb{R}^n \arrow[r, "T_{\mathcal{C}}"'] & \mathbb{R}^n_{\mathcal{C}}
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{R}^n \arrow[r, "P_{\mathcal{B}}"] \arrow[d, "A_T"'] & \mathbb{R}^n_{\mathcal{B}} \arrow[d, "P_{\mathcal{B}\to\mathcal{C}}"] \\
\mathbb{R}^n \arrow[r, "P_{\mathcal{C}}"'] & \mathbb{R}^n_{\mathcal{C}}
\end{tikzcd}$$

When $T = \mathrm{id}_{\mathbb{R}^n}$, then this reduces to the above case.

Next time

In the next lecture we will:

- Define the determinant of general $n \times n$ matrices.
- Compute determinants using cofactor expansion.
- Consider determinants for special cases such as upper triangular matrices.

Chapter 3: Determinants
§3.1 Introduction to Determinants
§3.2 Properties of Determinants

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define the determinant of general $n \times n$ matrices.
- Compute determinants using cofactor expansion.
- Consider determinants for special cases such as upper triangular matrices.
- Connect determinants to invertibility through row reduction.

Small determinants

The determinant of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$. The matrix is invertible if and only if this value is nonzero. In this chapter we'll develop a similar criteria for $n \times n$ matrices.

For an $n \times n$ matrix $A = (a_{ij})$, we denote by A_{ij} the submatrix obtained by deleting the ith row and jth column.

Example 1

Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then $A_{23} = \begin{bmatrix} 2 & -4 \\ 1 & 4 \end{bmatrix}$ and $A_{33} = \begin{bmatrix} 2 & -4 \\ 3 & 1 \end{bmatrix}$.

Small determinants

The definition of determinant is *recursive*. That means, in order to compute the determinant of an $n \times n$ matrix we first need to know how to compute the determinant of an $(n-1) \times (n-1)$ matrix. This is ok because we already know how to compute the determinant of a 2×2 matrix.

We won't fully derive the formula for an $n \times n$ matrix. However, I will try to give you an idea of where the formula comes from in the 2×2 and 3×3 cases.

2×2 **case:** Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be invertible, so $A \sim I_2$. Suppose $a \neq 0$. (If $a = 0$, the just switch row 1 and row 2. If $a = 0$ and $c = 0$, then A is not invertible.)

Row reduction gives

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} 1 & b/a \\ c & d \end{bmatrix} \sim \begin{bmatrix} 1 & b/a \\ 0 & d - cb/a \end{bmatrix}.$$

Thus, $A \sim I_n$ if and only if $d - \frac{cb}{a} \neq 0 \Leftrightarrow ad - bc \neq 0$.

3×3 **case:** Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

We'll assume that $a_{11} \neq 0$ along with some other nonzero assumptions. Again, A is invertible if and only if $A \sim I_3$. Row reduction gives

$$A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} - \frac{a_{21}a_{12}}{a_{11}} & a_{23} - \frac{a_{13}a_{21}}{a_{11}} \\ 0 & a_{32} - \frac{a_{31}a_{12}}{a_{11}} & a_{33} - \frac{a_{31}a_{13}}{a_{11}} \end{bmatrix}$$

$$\sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22}a_{11} - a_{21}a_{12} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{22}a_{11}a_{33} - a_{11}a_{23}a_{32} - a_{21}a_{12}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{bmatrix}$$

We can rewrite this a bit to get

$$\begin{aligned} \det(A) &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{12} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}). \end{aligned}$$

Determinant definition

Definition 2

For $n \geq 2$, the *determinant* of an $n \times n$ matrix $A = (a_{ij})$ is the sum of n terms of the form $\pm a_{1j} \det(A_{1j})$ with alternating $\pm$ signs:

$$\begin{aligned} |A| &= \det(A) \\ &= a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}) \\ &= \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}). \end{aligned}$$

Determinant – example

Example 3

Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then

$$A_{11} = \begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix} \qquad \det(A_{11}) = -9$$

$$A_{12} = \begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix} \qquad \det(A_{12}) = -5$$

$$A_{13} = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \qquad \det(A_{13}) = 11.$$

Hence,

$$\begin{aligned} \det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) \\ &= 2(-9) - (-4)(-5) + (3)(11) \\ &= -18 - 20 + 33 = -5. \end{aligned}$$

The definition of the determinant we gave is one of many (equivalent) definitions. In particular, the choice of using the first row is completely arbitrary.

The (i,j)-cofactor is $C_{ij} = (-1)^{i+j}\det(A_{ij})$. Our earlier formula is then

$$\det(A) = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

Theorem 4 (Cofactor Expansion)

The determinant of any $n \times n$ matrix can be determined by cofactor expansion along any row or column. In particular,

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

(jth col) $$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}C_{ij}.$$

Example 5

Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Expanding on the second column gives

$$A_{12} = \begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix} \qquad \det(A_{12}) = -5$$

$$A_{22} = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} \qquad \det(A_{22}) = -5$$

$$A_{32} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \qquad \det(A_{32}) = -5.$$

Hence,

$$\begin{aligned} \det(A) &= (-1)^{1+2} a_{12} \det(A_{12}) + (-1)^{2+2} a_{22} \det(A_{22}) + (-1)^{3+2} a_{32} \det(A_{32}) \\ &= -(-4)(-5) + (1)(-5) - 4(-5) \\ &= -20 - 5 + 20 = -5. \end{aligned}$$

Determinant facts

We will prove all of the following results eventually. Let A and B be $n \times n$ matrices.

- A is invertible if and only if $\det A \neq 0$.
- If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.
- $\det(A) = \det(A^T)$.
- $\det(AB) = \det(A)\det(B)$.
- If A is invertible, then A^{-1} is invertible and $\det(A^{-1}) = \frac{1}{\det(A)}$.

We will also show how row reduction affects the determinant of a matrix. Using this along with the above rules is a good way to compute determinants without going through cofactor expansion.

Several of these facts can be proved using *mathematical induction*.

The (First) Principle of Mathematical Induction

Let $S(n)$ be a statement about the integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If for all integers k with $k \geq n_0$, $S(k)$ implies that $S(k+1)$ is true, then $S(n)$ is true for all integers greater than or equal to n_0.

You should think of this like a staircase. First we show we can get to the staircase, then we show how to take a step (any step).

Theorem 6

If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.

Proof.

We will prove this for A upper-triangular. The lower-triangular case is similar.

The result for an upper-triangular 2×2 matrix follows from our formula:

$$\det(A) = \det\left(\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}\right) = ad - 0 = ad.$$

Suppose the claim holds for all upper-triangular $k \times k$ matrices, $k \geq 2$. Let A be a $(k+1) \times (k+1)$ upper-triangular matrix. By cofactor expansion along the first column, $\det(A) = a_{11}\det(A_{11})$. The submatrix A_{11} is an upper-triangular $k \times k$ matrix and hence by the inductive hypothesis, $\det(A_{11})$ is a product of its diagonal entries. That is, $\det(A_{11}) = a_{22}a_{33}\cdots a_{(k+1)(k+1)}$ so

$$\det(A) = a_{11}\det(A_{11}) = a_{11}a_{22}a_{33}\cdots a_{(k+1)(k+1)}.$$

Thus, the claim follows by the principle of mathematical induction. $\square$

Theorem 7

If A is a square matrix, then $\det(A) = \det(A^T)$.

Proof.

The claim holds for all 2×2 matrices since

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc = \det\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\right).$$

Suppose the claim holds for all $k \times k$ matrices, $k \geq 2$. Then $C_{ij}^T = C_{ji}$ for all $k \times k$ matrices. Let A be a $(k+1) \times (k+1)$ matrix. We use cofactor expansion along the first column of A^T:

$$\det(A^T) = a_{11} C_{11}^T + a_{12} C_{21}^T + \cdots + a_{1(k+1)} C_{(k+1)1}^T.$$

Since each cofactor C_{ij} is a $k \times k$ matrix, then we apply the inductive hypothesis to get

$$\begin{aligned}\det(A^T) &= a_{11} C_{11}^T + a_{12} C_{21}^T + \cdots + a_{1(k+1)} C_{(k+1)1}^T \\ &= a_{11} C_{11} + a_{12} C_{12} + \cdots + a_{1(k+1)} C_{1(k+1)} = \det(A).\end{aligned}$$

Therefore, the claim follows by the principle of mathematical induction. □

Theorem 8

Let A be a square matrix.

(1) If a multiple of one row of A is added to another row to produce a matrix B, then $\det A = \det B$.

(2) If two rows of A are interchanged to produce B, then $\det A = -\det B$.

(3) If one row of A is multiplied by k to produce B, then $\det B = k \cdot \det A$.

These rules can be proved using elementary matrices. Before attempting a proof of the above theorem, let's look at an example and some of the consequences.

Example 9

Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. We will compute the determinant of A using row reduction.

$$\begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 4 & -1 \\ 3 & 1 & 2 \\ 2 & -4 & 3 \end{bmatrix} \xrightarrow[R_2+(-3)R_1]{R_3+(-2)R_1} \begin{bmatrix} 1 & 4 & -1 \\ 0 & -11 & 5 \\ 0 & -12 & 5 \end{bmatrix}$$

$$\xrightarrow{R_3+(-\frac{12}{11})R_2} \begin{bmatrix} 1 & 4 & -1 \\ 0 & -11 & 5 \\ 0 & 0 & -5/11 \end{bmatrix}$$

This last matrix is upper triangular, so its determinant is just the product of the diagonal entries. Hence, the determinant of the last matrix is 5. From our original matrix, we use row addition (which doesn't change the determinant) and we interchanged once (which scales the determinant by -1). Hence, $\det(A) = (-1)(5) = -5$.

Theorem 10

A square matrix A is invertible if and only if $\det A \neq 0$.

Proof.

Let U be the row echelon form of A obtained by row replacement and row interchange. That is, we have not used scaling. Suppose r row interchanges were used to produce U. By the Invertible Matrix Theorem, U is invertible if and only if A is invertible. Since U is upper triangular, then $\det U$ is just the product of the pivots in A. Hence,

$$\det A = \begin{cases} (-1)^r \cdot (\text{product of pivots in U}) & \text{if } A \text{ is invertible} \\ 0 & \text{if } A \text{ is not invertible.} \end{cases}$$

Thus, $\det A \neq 0$ if and only if $\det U \neq 0$ and $\det U \neq 0$ if and only if U is invertible if and only if A is invertible. $\square$

Example 11

Let $A = \begin{bmatrix} 1 & 2 & -2 \\ -5 & -1 & 9 \\ 1 & -7 & -1 \end{bmatrix}$. We will compute $\det(A)$ to determine whether or not A is invertible. Using cofactor expansion along the first row:

$$\begin{aligned}
\det(A) &= (-1)^{1+1} a_{11} \det(A_{11}) + (-1)^{1+2} a_{12} \det(A_{12}) + (-1) a_{13} \det(A_{13}) \\
&= 1 \det\left(\begin{bmatrix} -1 & 9 \\ -7 & -1 \end{bmatrix}\right) - 2 \det\left(\begin{bmatrix} -5 & 9 \\ 1 & -1 \end{bmatrix}\right) + (-2) \det\left(\begin{bmatrix} -5 & -1 \\ 1 & -7 \end{bmatrix}\right) \\
&= 64 + 8 - 72 = 0.
\end{aligned}$$

Hence, A is not invertible.

Next time

In the next lecture we will:

- Consider how determinants behave with respect to row reduction.
- Learn Cramer's Rule for solving systems of equations.
- Consider applications of determinants to area and volume.

Chapter 3: Determinants
§3.2 Properties of Determinants
§3.3 Cramer's Rule

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Consider how determinants behave with respect to row reduction.
- Learn Cramer's Rule for solving systems of equations.
- Consider applications of determinants to area and volume.

Recall that an elementary matrix is obtained from the identity matrix by performing a single row operation. If E is an elementary matrix, then

$$\det(E) = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r. \end{cases}$$

Example 1

Suppose E is obtained by adding $3R_1$ to R_2. Then

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since E is lower triangular, then $\det(E)$ is the product of the diagonal entries. Hence, $\det(E) = 1$.

Elementary matrices

Recall that an elementary matrix is obtained from the identity matrix by performing a single row operation. If E is an elementary matrix, then

$$\det(E) = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r. \end{cases}$$

Example 2

Suppose E is obtained by switching R_1 with R_2. Then

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

By cofactor expansion along the first row, we have

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

Elementary matrices

Recall that an elementary matrix is obtained from the identity matrix by performing a single row operation. If E is an elementary matrix, then

$$\det(E) = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r. \end{cases}$$

Example 3

Suppose E is obtained by scaling $5R_3$. Then

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

Then E is diagonal and so $\det(E)$ is the product of the diagonal entries. Hence, $\det(E) = 5$.

Example 4

Suppose A is 2×2, so $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

If E is scaling row 1 by r, then

$$\det(EA) = \begin{vmatrix} ra & rb \\ c & d \end{vmatrix} = rad - rbc = r(ad - bc) = r\det(A) = \det(E)\det(A).$$

If E is scaling row 2 by r, the argument is similar. If E is switching row 1 and row 2, then

$$\det(EA) = \begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - ad = -(ad - bc) = -\det(A) = \det(E)\det(A).$$

If E is adding r times row 1 to row 2, then

$$\begin{aligned} \det(EA) = \begin{vmatrix} a & b \\ c + ra & d + rb \end{vmatrix} &= a(d + rb) - b(c + ra) = (ad - bc) + r(ab - ba) \\ &= ad - bc = \det(A) = \det(E)\det(A). \end{aligned}$$

Lemma 5

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix. Then $\det(EA) = \det(E)\det(A)$.

Proof.

The proof is by induction on the size of A. The case when A is 2×2 is Example 4. Suppose for all $k \times k$ matrices A, $k \geq 2$, $\det(EA) = \alpha \det(A)$ where $\alpha = 1, -1, r$ depending on E. We will show the claim holds for $(k+1) \times (k+1)$ matrices.

Since $k + 1 \geq 3$, there is some row, say row i of A that is unchanged by E. Let $B = EA$. We will use cofactor expansion along row i of B. The B_{ij} are $k \times k$ matrices obtained from A by applying E. By the inductive hypothesis, $\det(B_{ij}) = \alpha \det(A_{ij})$ where $\alpha = 1, -1, r$ depending on E. So,

$$\det(EA) = \det(B)$$

$$= a_{i1}(-1)^{i+1}\det(B_{i1}) + a_{i2}(-1)^{i+2}\det(B_{i2}) + \cdots + a_{i(k+1)}(-1)^{i+(k+1)}\det(B_{i(k+1)}).$$

$$= \alpha a_{i1}(-1)^{i+1}\det(A_{i1}) + \alpha a_{i2}(-1)^{i+2}\det(A_{i2}) + \cdots + \alpha a_{i(k+1)}(-1)^{i+(k+1)}\det(A_{i(k+1)})$$

$$= \alpha \det(A) = \det(E)\det(A).$$

Thus, the theorem is true by the principle of mathematical induction. □

Multiplicative property

Theorem 6

If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

Proof.

Suppose A is not invertible, then neither is AB (this was in a problem session). Hence, $\det(AB) = \det(A)\det(B)$ because both sides are zero. A similar argument holds if B is not invertible.

Suppose A is invertible. Then $A \sim I_n$ by a series of elementary operations. That is, $A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1$. Thus, by Lemma 5,

$$|AB| = |E_p E_{p-1} \cdots E_1 B| = |E_p||E_{p-1}| \cdots |E_1||B| = |E_p E_{p-1} \cdots E_1||B| = |A||B|. \qquad \square$$

Corollary 7

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

For a $n \times n$ matrix A and any $\mathbf{b} \in \mathbb{R}^n$, denote by $A_i(\mathbf{b})$ the matrix obtained by replacing column i in A with $\mathbf{b}$.

Theorem 8 (Cramer's Rule)

Let A be an invertible $n \times n$ matrix. For any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \quad i = 1, \ldots, n.$$

Proof.

Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. If $A\mathbf{x} = \mathbf{b}$, then

$$\begin{aligned} A \cdot I_i(\mathbf{x}) &= A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}). \end{aligned}$$

Now $\det(A \cdot I_i(\mathbf{x})) = \det A_i(\mathbf{b})$ so

$$\det(A) \det I_i(\mathbf{x}) = \det A_i(\mathbf{b}). \tag{1}$$

But $\det(I_i(\mathbf{x})) = x_i$, so (1) redues to $\det(A)x_i = \det A_i(\mathbf{b})$. The result now follows. $\square$

Example 9

We will use Cramer's Rule to solve the following system:

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

We have $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$ so $\det(A) = -2$. Since $\mathbf{b} = \left[\begin{smallmatrix} 3 \\ 7 \end{smallmatrix}\right]$, then

$$A_1(\mathbf{b}) = \begin{bmatrix} 3 & 4 \\ 7 & 6 \end{bmatrix} \qquad \det(A_1(\mathbf{b})) = -10 \qquad x_1 = \frac{\det(A_1(\mathbf{b}))}{\det(A)} = \frac{-10}{-2} = 5$$

$$A_2(\mathbf{b}) = \begin{bmatrix} 3 & 3 \\ 5 & 7 \end{bmatrix} \qquad \det(A_2(\mathbf{b})) = 6 \qquad x_2 = \frac{\det(A_2(\mathbf{b}))}{\det(A)} = \frac{6}{-2} = -3.$$

Cramer's Inverse Formula

Let A be an $n \times n$ invertible matrix. Cramer's Rule also leads to a formula for the inverse of A.

The jth column of A^{-1} is the unique solution to $A\mathbf{x} = \mathbf{e}_j$. By Cramer's Rule, the ith entry of $\mathbf{x}$ is $x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A}$. Cofactor expansion down column i of $A_i(e_j)$ shows that

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji}.$$

Definition 10

The *adjugate matrix* of A, denoted $\operatorname{adj} A$, is the $n \times n$ matrix with (i, j)-entry C_{ji}.

Theorem 11 (Cramer's Inverse Formula)

Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

The next example illustrates why this isn't a good method for computing the inverse matrix.

Cramer's Inverse Formula

Example 12

We will use Cramer's Inverse Formula to find the inverse of $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$.

$$\begin{aligned} C_{11} &= (-1)^{1+1}\det(A_{11}) = -9 & C_{12} &= (-1)^{1+2}\det(A_{12}) = 5 & C_{13} &= (-1)^{1+3}\det(A_{13}) = 11 \\ C_{21} &= (-1)^{2+1}\det(A_{21}) = 8 & C_{22} &= (-1)^{2+2}\det(A_{22}) = -5 & C_{23} &= (-1)^{2+3}\det(A_{23}) = -12 \\ C_{31} &= (-1)^{3+1}\det(A_{11}) = -11 & C_{32} &= (-1)^{3+2}\det(A_{12}) = 5 & C_{33} &= (-1)^{3+3}\det(A_{13}) = 14 \end{aligned}$$

Hence,

$$\operatorname{adj} A = \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} -9 & 8 & -11 \\ 5 & -5 & 5 \\ 11 & -12 & 14 \end{bmatrix}.$$

Therefore,

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A = -\frac{1}{5}\begin{bmatrix} -9 & 8 & -11 \\ 5 & -5 & 5 \\ 11 & -12 & 14 \end{bmatrix}$$

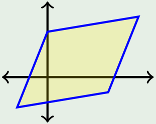
Area of a parallelogram

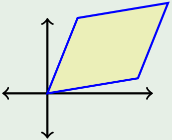
If A is a 2×2 matrix, then $|\det A|$ is the area of the parallelogram determined by the columns of A.

The key point to the above theorem is that if $\mathbf{a}_1, \mathbf{a}_2 \neq \mathbf{0}$, then for any scalar c the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ is equal to that of $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

Example 13

Consider the parallelogram determined by the points $(-2,-2),(0,3),(4,-1),(6,4)$. We first translate the parallelogram by adding $(2,2)$ to each vertex.





Then the points are $(0,0),(2,5),(6,1),(8,6)$. This is the parallelogram determined by the vectors through the points $(2,5)$ and $(6,1)$. Thus,

$$\text{Area } = \begin{vmatrix} 2 & 6 \\ 5 & 1 \end{vmatrix} = |-28| = 28.$$

Volume of a parallelepiped

If A is a 3×3 matrix, then $|\det A|$ is the volume of the parallelepiped determined by the columns of A.

Theorem 14

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation with standard matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\textit{area of } T(S)\} = |\det A| \cdot \{\textit{area of } S.\}.$$

If $T : \mathbb{R}^3 \to \mathbb{R}^3$ and S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\textit{volume of } T(S)\} = |\det A| \cdot \{\textit{volume of } S.\}.$$

Next time

In the next lecture we will:

- Define eigenvectors and eigenvalues of a matrix.
- Find eigenvalues using the characteristic equation.
- Define eigenspaces and find a basis using row-reduction.

Chapter 5: Eigenvectors and Eigenvalues
§5.1 Introduction to Eigenstuff
§5.2 The Characteristic Equation

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define eigenvectors and eigenvalues of a matrix.
- Find eigenvalues using the characteristic equation.
- Define eigenspaces and find a basis using row-reduction.

Eigenvectors

Eigenvectors and eigenvalues are important tools in both pure and applied mathematics. Google's PageRank algorithm relies on the theory of eigenvectors. Those in differential equations may have already come across these.

Example 1

Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

Observe that $A\mathbf{u} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\mathbf{u}$.

Geometrically, we can interpret this as A stretching the vector $\mathbf{u}$. This does not happen with $\mathbf{v}$.

Definition 2

An *eigenvector* of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an *eigenvalue* of A if there is a nontrivial solution of $A\mathbf{x} = \lambda\mathbf{x}$. We say $\mathbf{x}$ is the eigenvector corresponding to λ.

Example 3

Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$. Then $\mathbf{u} = \left[\begin{smallmatrix} 6 \\ -5 \end{smallmatrix}\right]$ is an eigenvector for the matrix A with eigenvalue -4.

Invertible Matrix Theorem (cont.)

Let A be an $n \times n$ matrix. Then A is invertible if and only if 0 is *not* an eigenvalue of A.

Proof.

If 0 is an eigenvalue of A, then there is some nontrivial solution to the matrix equation $A\mathbf{x} = \mathbf{0}$. But this implies that A is not invertible. $\square$

Example 4

Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$. We will show that 7 is an eigenvalue of A.

If 7 is an eigenvalue, then there is some eigenvector $\mathbf{x}$ such that $A\mathbf{x} = 7\mathbf{x}$. This is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or $(A - 7I)\mathbf{x} = \mathbf{0}$. That is, $\mathbf{x}$ is a solution to the homogeneous matrix equation for the matrix $A - 7I$. Hence, we row reduce:

$$A - 7I = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}.$$

The general solution is of the form $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_2$. Hence, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector for A.

We can check this:

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Note that any vector in the null space of $A - 7I$ will also be an eigenvector.

Definition 5

For an eigenvalue λ of a matrix A, the *eigenspace* $E_\lambda(A)$ of A corresponding to λ is the null space of the matrix $A - \lambda I$. The dimension of the eigenspace corresponding to an eigenvalue λ is called the *geometric multiplicity of* λ, denoted $\text{geomult}_\lambda(A)$.

Any eigenspace of a matrix A is automatically a subspace of $\mathbb{R}^n$ because null spaces are subspaces.

Example 6

Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. We will find a basis for the eigenspace $E_2(A)$.

We need a basis for the null space of $A - 2I$ so we set up the corresponding matrix and row reduce:

$$A - 2I = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -1/2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence, a basis of $E_2(A)$ is

$$\left\{ \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Since $E_2(A)$ is 2-dimensional, then $\operatorname{geomult}_2(A) = 2$.

The characteristic equation

Suppose λ is an eigenvalue of an $n \times n$ matrix A, then $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero $\mathbf{v}$. Said another way, the matrix equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. That is, $(A - \lambda I)$ is **not** invertible and so $\det(A - \lambda I) = 0$.

Definition 7

For an $n \times n$ matrix A, the *characteristic polynomial* of A is $\chi_A(t) = \det(A - tI)$. The *characteristic equation* is $\chi_A(t) = 0$.

The roots of the characteristic equation are the eigenvalues of A.

Recall that the *multiplicity* of a root λ in an equation $p(t)$ is the number of times $(\lambda - t)$ appears in the factorization of $p(t)$. That is, $p(t) = (\lambda - t)^\alpha q(t)$ such that $(\lambda - t)$ does not divide $q(t)$, then the multiplicity of λ is α.

Definition 8

If λ is a root of the characteristic equation, then the multiplicity of the root is called the *algebraic multiplicity of* λ, denoted $\text{algmult}_\lambda(A)$.

Example 9

Let $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$. Then

$$\det(A - tI) = \det\left(\begin{bmatrix} 2-t & 3 \\ 3 & -6-t \end{bmatrix}\right) = (2-t)(-6-t) - 9 = t^2 + 4t - 21.$$

This factors as $(t+7)(t-3)$. Thus, $\det(A - tI) = 0$ if $t = 3$ or -7. So, the eigenvalues of A are $3, -7$.

Now we compute a basis of each eigenspace:

$$A - 3I = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix} \qquad E_3(A) : \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$$

$$A + 7I = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1/3 \\ 0 & 0 \end{bmatrix} \qquad E_{-7}(A) : \left\{ \begin{bmatrix} -1/3 \\ 1 \end{bmatrix} \right\}$$

Hence, each of the eigenvalues has algebraic and geometric multiplicity 1.

Example 10

Let $A = \begin{bmatrix} 2 & 1 & 7 & 0 \\ 0 & 5 & 2 & 4 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$. Then $\det(A - tI) = (2-t)(5-t)(-1-t)(2-t)$. Thus, the eigenvalues are $2, 5, -1, 2$, which are exactly the diagonal entries of the matrix.

Theorem 11

The eigenvalues of a triangular matrix are the entries along the main diagonal.

Proof.

If A is triangular, then so is $A - tI$ and hence

$$\det(A - tI) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t).$$

Thus, the roots of the characteristic equation are the diagonal entries of A. $\square$

Definition 12

If A and B are $n \times n$ matrices, then A is said to be *similar* to B if there exists an invertible matrix P such that $B = P^{-1}AP$.

Theorem 13

If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues.

Proof.

If $B = P^{-1}AP$, then

$$B - tI = P^{-1}AP - tP^{-1}P = P^{-1}(A - tI)P.$$

Thus,

$$\det(B - tI) = \det(P^{-1}(A - tI)P) = \det(A - tI).$$

$\square$

Next time

In the next lecture we will:

- Discuss the procedure for diagonalizing matrices.
- Find conditions for when a matrix is diagonalizable.

Chapter 5: Eigenvectors and Eigenvalues
§5.3 Diagonalization

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Discuss the procedure for diagonalizing matrices.
- Find conditions for when a matrix is diagonalizable.

Similar matrices

Definition 1

If A and B are $n \times n$ matrices, then A is said to be *similar* to B if there exists an invertible matrix P such that $B = P^{-1}AP$.

Theorem 2

If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues.

Similarity is an example of an *equivalence relation*. That means it is

- Reflexive: A is similar to itself
$$A = I^{-1}AI$$
- Symmetric: If A is similar to B, then B is similar to A
$$B = P^{-1}AP \Rightarrow A = PBP^{-1}$$
- Transitive: If A is similar to B and B is similar to C, then A is similar to C
$$A = P^{-1}BP \text{ and } B = Q^{-1}CQ \Rightarrow A = P^{-1}(Q^{-1}CQ)P = (QP)^{-1}CQP$$

Definition 3

An $n \times n$ matrix is *diagonalizable* if it is similar to a diagonal matrix. To *diagonalize* a matrix A is to find an invertible matrix P and a diagonal matrix D such that $D = P^{-1}AP$ (equivalently, $A = PDP^{-1}$).

We will see how to use eigenvalues and eigenvectors to diagonalize matrices. The following example is similar to a problem from earlier in the semester.

Example 4

Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Then $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$. Then

$$A^3 = (PDP^{-1})(PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)D(P^{-1}P)DP^{-1} = PD^3P^{-1}.$$

Note that $D^3 = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$. So

$$A^3 = PD^3P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 223 & 98 \\ -196 & -71 \end{bmatrix}.$$

Diagonalizable matrices

Theorem 5

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof.

Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the (linearly independent) eigenvectors of A with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Set $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and D to be the diagonal matrix with entries $\lambda_1, \dots, \lambda_n$. By IMT, P is invertible and

$$AP = A\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} = PD.$$

Hence, $D = P^{-1}AD$ so A is diagonalizable.

Conversely, suppose A is diagonalizable. Then there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. Write $P = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$ and let $c_1, \dots, c_n$ be the diagonal entries of D. As above, $AP = PD$ implies $A\mathbf{x}_i = c_i\mathbf{x}_i$ for each i. But then each $\mathbf{x}_i$ is an eigenvector for A. Because P is invertible, the vectors $\mathbf{x}_i$ are linearly independent. □

Diagonalization

Under the hypotheses of Theorem 5, the eigenvectors of A form a basis of $\mathbb{R}^n$ called an *eigenbasis* of $\mathbb{R}^n$ with respect to A.

By Theorem 5, the following steps diagonalize an $n \times n$ matrix A.

(1) Find the eigenvalues of A using the characteristic polynomial.

(2) Find an eigenbasis for each eigenvalue.

(3) Construct P with columns from eigenvectors (assuming there are n linearly independent eigenvalues).

(4) Construct D from eigenvalues in order corresponding to P.

(5) Check!

Example 6

Let $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$.

Then

$$\begin{aligned}
\det(A - tI) &= \begin{vmatrix} 1-t & 3 & 3 \\ -3 & -5-t & -3 \\ 3 & 3 & 1-t \end{vmatrix} \\
&= (1-t)\begin{vmatrix} -5-t & -3 \\ 3 & 1-t \end{vmatrix} - 3\begin{vmatrix} -3 & -3 \\ 3 & 1-t \end{vmatrix} + 3\begin{vmatrix} -3 & -5-t \\ 3 & 3 \end{vmatrix} \\
&= (1-t)(t^2 + 4t + 4) - 3(3t+6) + 3(3t+6) \\
&= (1-t)(t+2)^2.
\end{aligned}$$

Hence, the eigenvalues for this matrix are 1 and -2.

Example 6

Let $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$. The eigenvalues for this matrix are 1 and -2.

We compute a basis for each eigenspace $E_\lambda(A)$.

$$A - (1)I = \begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \left[\begin{smallmatrix} 1 \\ -1 \\ 1 \end{smallmatrix} \right] \right\},$$

$$A - (-2)I = \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \left[\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix} \right], \left[\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix} \right] \right\}.$$

Thus, A is diagonalizable by the matrices

$$P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

When is a matrix diagonalizable?

Theorem 7

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

Proof.

Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly dependent. Because $\mathbf{v}_1 \neq \mathbf{0}$, then there exists $p \in \{2, \ldots, r\}$ such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent and $\mathbf{v}_{p+1}$ is a linear combination of those vectors. That is, there exist $c_1, \ldots, c_p$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1}. \tag{1}$$

Multiplying both sides of (1) by A gives

$$\begin{aligned}
c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\
c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}\mathbf{v}_{p+1} \\
c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \\
c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p &= \mathbf{0}.
\end{aligned}$$

This contradicts the linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ because none of the $\lambda_i - \lambda_{p+1}$ are zero by hypothesis. □

The next theorem follows almost directly from Theorem 7.

Theorem 8

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Here is the more general case. It says that A is diagonalizable if the dimension of the eigenspace for each eigenvalue matches multiplicity of that eigenvalue as a root of the characteristic polynomial.

Theorem 9

For every eigenvalue λ of an $n \times n$ matrix A, $\operatorname{geomult}_\lambda(A) \leq \operatorname{algmult}_\lambda(A)$.

Proof.

Let λ be an eigenvalue of A and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for $E_\lambda(A)$. We extend $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Note that none of the $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ are eigenvectors for λ (if they were, they would be basis elements of $E_\lambda(A)$). Let $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$, so P is invertible by the IMT. Set $B = P^{-1}AP$. Then B is a block matrix of the form

$$B = \left[\begin{array}{cccc|c} \lambda & 0 & \cdots & 0 & \\ 0 & \lambda & \cdots & 0 & \\ \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \lambda & \\ \hline & & 0 & & \star \end{array}\right]$$

Since B is similar to A, they have the same characteristic equation and hence the same algebraic multiplicity for each eigenvalue. It is clear that the algebraic multiplicity of λ in B is k. $\square$

When is a matrix diagonalizable?

Theorem 10

An $n \times n$ matrix A is diagonalizable if and only if $\operatorname{geomult}_\lambda(A) = \operatorname{algmult}_\lambda(A)$ *for every eigenvalue λ.*

Proof.

The statement $\operatorname{geomult}_\lambda(A) = \operatorname{algmult}_\lambda(A)$ for every eigenvalue λ is equivalent to $\mathbb{R}^n$ having an eigenbasis with respect to A, which is equivalent to A having n linearly independent eigenvectors. $\square$

Example 11

Consider the matrix $A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$.

The eigenvalues of this matrix are 1 and -2. The algebraic multiplicity of -2 is 2. However, the geometric multiplicity of -2 is 1 (check). Hence, A is *not* diagonalizable.

Example 12

Consider the matrix $A = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 1 & 4 & -3 & 0 \\ -1 & -2 & 0 & -3 \end{bmatrix}$.
The eigenvalues of A are 5 and -3. Then a basis of the two eigenspaces is

$$E_5(A) : \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\} \qquad E_{-3}(A) : \left\{ \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

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

$$P = \begin{bmatrix} -16 & 8 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} -3 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

Next time

In the next lecture we will:

- Discuss applications of eigenstuff.
- Solve discrete dynamical systems, including predator-prey systems.
- Solve systems of linear differential equations.

Chapter 5: Eigenvectors and Eigenvalues
§5.6 Discrete Dynamical Systems
§5.7 Applications to Differential Equations

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Discuss applications of eigenstuff.
- Solve discrete dynamical systems, including predator-prey systems.
- Solve systems of linear differential equations.

Example 1

Wood rats are the primary source of food for spotted owls in the Redwood forest. Denote the population of owls and wood rats at time k by

$$\mathbf{x}_k = \left[\begin{smallmatrix} O_k \\ R_k \end{smallmatrix}\right],$$

where k is the time in months, O_k is the number of owls in the region, and R_k is the number of wood rats in the region (in thousands).

The populations at time $k+1$ are related to the populations at time k by the *dynamical system*,

$$\begin{aligned} O_{k+1} &= (.5)O_k + (.4)R_k \\ R_{k+1} &= -p \cdot O_k + (1.1)R_k, \end{aligned}$$

where $1000p$ denotes the average number of rats eaten by one owl in a month. Note that we can denote this system by

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{where} \quad A = \begin{bmatrix} .5 & .4 \\ -p & 1.1 \end{bmatrix}.$$

We will study the long-term behavior (or *evolution*) of this system.

Example 1

We can denote this system by

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{where} \quad A = \begin{bmatrix} .5 & .4 \\ -p & 1.1 \end{bmatrix}.$$

Let $p = .104$. Then the eigenvalues of A are $\lambda_1 = 1.02$ and $\lambda_2 = .58$ with corresponding eigenvectors

$$\mathbf{v}_1 = \left[\begin{smallmatrix} 10 \\ 13 \end{smallmatrix}\right], \qquad \mathbf{v}_2 = \left[\begin{smallmatrix} 5 \\ 1 \end{smallmatrix}\right].$$

Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathbb{R}^2$, then any initial vector $\mathbf{x}_0$ can be written as $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ for scalars c_1, c_2. Then,

$$\mathbf{x}_1 = A\mathbf{x}_0 = A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1(A\mathbf{v}_1) + c_2(A\mathbf{v}_2) = c_1(1.02)\mathbf{v}_1 + c_2(.58)\mathbf{v}_2.$$

In general,

$$\mathbf{x}_k = c_1(1.02)^k\mathbf{v}_1 + c_2(.58)^k\mathbf{v}_2.$$

Example 1

We can denote this system by

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{where} \quad A = \begin{bmatrix} .5 & .4 \\ -p & 1.1 \end{bmatrix}.$$

We have

$$\mathbf{x}_k = c_1(1.02)^k\mathbf{v}_1 + c_2(.58)^k\mathbf{v}_2.$$

Assume $c_1 > 0$. As $k \to \infty$, $(.58)^k \to 0$. Hence, for sufficiently large k we have

$$\mathbf{x}_k \approx c_1(1.02)^k\left[\begin{smallmatrix} 10 \\ 13 \end{smallmatrix}\right].$$

As k increases, the approximation gets better. So for large k,

$$\mathbf{x}_{k+1} \approx c_1(1.02)^{k+1}\left[\begin{smallmatrix} 10 \\ 13 \end{smallmatrix}\right] = (1.02)c_1(1.02)^k\left[\begin{smallmatrix} 10 \\ 13 \end{smallmatrix}\right] = (1.02)\mathbf{x}_k.$$

This says both populations grow by a factor of about 1.02 every month, but the ratio of critters (10 owls to 13 thousand rats) stays the same.

Discrete dynamical systems

In general, assume a dynamical system is described by the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ and that A is diagonalizable with n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues and arrange them so that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$, any initial vector $\mathbf{x}_0$ can be written (uniquely) as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

As in our example, we have

$$\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n.$$

In general we will study the behavior as $k \to \infty$.

Example 2

Let $x(t)$ be a differentiable function of t. We can use calculus to find all solutions to the differential equation

$$x'(t) = 2x(t).$$

Write $x = x(t)$. By separation of variables, we have

$$\begin{aligned} \frac{dx}{dt} &= 2x \\ \frac{dx}{x} &= 2dt \\ \ln x &= 2t + C. \end{aligned}$$

Hence, the general solution is $x = ce^{2t}$ for any constant c.

Definition 3

A system of linear first-order differential equations (DEs) is a set of equations

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

where the x_i are all differentiable functions in t and the a_{ij} are all constants.

We are using shorthand here with x_i in place of $x_i(t)$ and x_i' for the derivative of x_i (with respect to t). We can represent the entire system in matrix form as

$$\mathbf{x}'(t) = A\mathbf{x}(t),$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \qquad A = (a_{ij}).$$

Differential equations

Definition 4

A *solution* to the system is a vector-valued function $\mathbf{x}(t)$ that satisfies $\mathbf{x}'(t) = A\mathbf{x}(t)$.

Definition 5

An *initial value problem* is a system of DEs along with an initial condition $\mathbf{x}_0 = \mathbf{x}(0)$.

Note that $\mathbf{x}'(t) = A\mathbf{x}(t)$ is linear. If $c, d \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v}$ are solutions, then

$$(c\mathbf{u} + d\mathbf{v})' = c\mathbf{u}' + d\mathbf{v}' = c(A\mathbf{u}) + d(A\mathbf{v}) = A(c\mathbf{u} + d\mathbf{v}).$$

Moreover, $\mathbf{0}$ is (trivially) a solution and so the set of solutions is a subspace of $\mathbb{R}^n$.

Example 6

Let $\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ and $A = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$.

It follows that our solutions are $x_1 = c_1 e^{3t}$ and $x_2 = c_2 e^{-5t}$. Thus, the solution space is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-5t} \end{bmatrix} = c_1 \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] e^{3t} + c_2 \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] e^{-5t}.$$

This example suggests that solutions to a system of linear equations might have a simple presentation as a linear combination of functions of the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$.

Suppose a solution to $\mathbf{x}'(t) = A\mathbf{x}(t)$ has the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$. By calculus, $\mathbf{x}'(t) = \lambda \mathbf{v} e^{\lambda t}$. But then $A\mathbf{x} = A\mathbf{v}e^{\lambda t}$ (this uses the fact that $e^{\lambda t} \neq 0$ for all t). Hence, $A\mathbf{v} = \lambda \mathbf{v}$, that is, $\mathbf{v}$ is an eigenvector with eigenvalue λ. Solutions of this form are called *eigenfunctions*.

Example 7

We will solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ given

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \qquad A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}.$$

The eigenvalues of A are $-.5, -2$ with corresponding eigenvectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Thus, the eigenfunctions are

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} \quad \text{and} \quad \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t},$$

and any solution is a linear combination of these two. That is,

$$\mathbf{x} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

and we can use the initial conditions to solve for c_1, c_2.

Example 7

Consider the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ given

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \qquad A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}.$$

We have the general solution

$$\mathbf{x} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

so

$$\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \mathbf{x}_0 = \mathbf{x}(0) = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

To solve for c_1, c_2 we solve the vector equation by row reducing the corresponding augmented matrix,

$$\begin{bmatrix} 1 & -1 & 5 \\ 2 & 1 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \end{bmatrix}.$$

Hence, the solution to the given IVP is,

$$\mathbf{x} = 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} - 2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

Next time

In the next lecture we will:

- Begin a study of orthogonality.
- Define inner products and use it to compute the length of a vector.
- Study orthogonal complements of subspaces.

Chapter 6: Orthogonality
§6.1 Inner Products

MTH 222

Linear Algebra

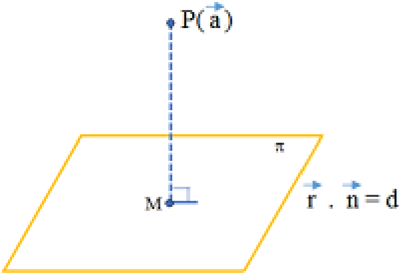
Goals for today:

- Begin a study of orthogonality.
- Define inner products and use it to compute the length of a vector.
- Study orthogonal complements of subspaces.

We now return to a discussion on the geometry of vectors. There are many applications of the notion of orthogonality, some of which we will discuss. Here is a basic geometric question that we will address shortly (those in Calc III may have seen this already).

Question

Given a plane P and a point p (in $\mathbb{R}^3$). What is the distance from p to P? That is, what is the length of the shortest possible line segment that one could draw from p to P?



Definition 1

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The *inner product* of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n.$$

The inner product is also referred to as the *dot product*. Another product, the cross product, will be discussed at a later time.

Example 2

Let $\mathbf{a} = \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Then

$$\mathbf{a} \cdot \mathbf{b} = (-1)(2) + (3)(1) + (5)(-2) = -9.$$

Properties of the inner product

Theorem 3

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *and let* $c \in \mathbb{R}$. *Then we have the following:*

(1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$.

(3) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$.

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$.

Proof.

(1) $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} (u_i \cdot v_i) = \sum_{i=1}^{n} (v_i \cdot u_i) = \mathbf{v} \cdot \mathbf{u}$.

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \sum_{i=1}^{n} (u_i + v_i) \cdot w_i = \sum_{i=1}^{n} (u_i \cdot w_i) + (v_i \cdot w_i) = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$.

(3) $(c\mathbf{u}) \cdot \mathbf{v} = \sum_{i=1}^{n} (cu_i \cdot v_i) = c \sum_{i=1}^{n} (u_i \cdot v_i) = c(\mathbf{u} \cdot \mathbf{v})$.

Theorem 4

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *and let* $c \in \mathbb{R}$. *Then we have the following:*

(1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$.

(3) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$.

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$.

Proof.

(4) Since $u_i \cdot u_i = u_i^2 \geq 0$ for all i, then

$$\mathbf{u} \cdot \mathbf{u} = \sum_{i=1}^{n} (u_i \cdot u_i) = \sum_{i=1}^{n} u_i^2 \geq 0.$$

Moreover, since the u_i^2 are all nonnegative, the sum will be zero if and only if $u_i^2 = 0$ for all i. But $u_i^2 = 0$ if and only if $u_i = 0$. Hence, $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$. $\square$

The inner product is a useful tool to study the geometry of vectors.

Definition 5

The *length* (or *norm*) of $\mathbf{v} \in \mathbb{R}^n$ is the non-negative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \qquad \text{and} \qquad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}.$$

A *unit vector* is a vector of length 1. Note that $\|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$ for $c \in \mathbb{R}$.

If $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unit vector in the same direction as $\mathbf{v}$.

Example 6

Let $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Then

$$\|\mathbf{b}\| = \sqrt{\mathbf{b} \cdot \mathbf{b}} = \sqrt{(2)^2 + (1)^2 + (-2)^2} = \sqrt{9} = 3.$$

The associated unit vector are then

$$\frac{1}{\|\mathbf{b}\|}\mathbf{b} = \frac{1}{3}\mathbf{b} = \begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix}.$$

Distance

Recall that the distance between two points (a_1, a_2) and (b_1, b_2) in $\mathbb{R}^2$ is determined by the well-known distance formula $\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$. We can similarly define distance between vectors.

Definition 7

For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the *distance* between $\mathbf{u}$ and $\mathbf{v}$, written $d(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is, $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Example 8

Let $\mathbf{a} = \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. The distance between $\mathbf{a}$ and $\mathbf{b}$ is

$$\|\mathbf{a} - \mathbf{b}\| = \left\| \begin{bmatrix} -3 \\ 2 \\ 7 \end{bmatrix} \right\| = \sqrt{(-3)^2 + (2)^2 + (7^2)} = \sqrt{62}.$$

Definition 9

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be *orthogonal* (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Orthogonality generalizes the idea of perpendicular lines in $\mathbb{R}^2$. Two lines (represented as vectors $\mathbf{u}, \mathbf{v}$) are perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$.

$$(d(\mathbf{u}, -\mathbf{v}))^2 = \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2$$
$$(d(\mathbf{u}, \mathbf{v}))^2 = \|\mathbf{u} - \mathbf{v}\|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = \|\mathbf{u}\|^2 - 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2 .$$

Hence, these two quantities are equal if and only if $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v}$. Equivalently, $\mathbf{u} \cdot \mathbf{v} = 0$.

The next theorem now follows directly.

Theorem 10 (The Pythagorean Theorem)

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Definition 11

Let $W \subset \mathbb{R}^n$ be a subspace. The set

$$W^\perp = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{z} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}$$

is called the *orthogonal complement of* W.

Part of the problem session this week will be to show that $W^\perp$ is a subspace of $\mathbb{R}^n$.

Example 12

Let W be a plane through the origin in $\mathbb{R}^3$ and let L be a line through $\mathbf{0}$ perpendicular to W. If $\mathbf{z} \in L$ and $\mathbf{w} \in W$ are nonzero, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$. In fact, $L = W^\perp$ and $W = L^\perp$.

Let A be $m \times n$. The *row space* of A, denoted $\operatorname{Row} A$, is the span of the rows of A, thought of as vectors in $\mathbb{R}^n$. Alternatively, we could define the row space by $\operatorname{Row} A = \operatorname{Col} A^T$.

Theorem 13

Let A be an $m \times n$ matrix. Then $(\operatorname{Row} A)^\perp = \operatorname{Nul} A$ and $(\operatorname{Col} A)^\perp = \operatorname{Nul} A^T$.

Proof.

If $\mathbf{x} \in \operatorname{Nul} A$, then $A\mathbf{x} = \mathbf{0}$ by definition. Hence, $\mathbf{x}$ is orthogonal to each row of A. Since the rows of A span $\operatorname{Row} A$, then $\mathbf{x} \in (\operatorname{Row} A)^\perp$.

Conversely, if $\mathbf{x} \in (\operatorname{Row} A)^\perp$, then $\mathbf{x}$ is orthogonal to each row of A and $A\mathbf{x} = 0$, so $\mathbf{x} \in \operatorname{Nul} A$.

For the second statement, we can apply the first part to get

$$(\operatorname{Col} A)^\perp = (\operatorname{Row} A^T)^\perp = \operatorname{Nul} A^T \qquad \square.$$

Orthogonality also gives us a way to determine the angle between vectors in $\mathbb{R}^2$.

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ be nonzero. By the Law of Cosines,

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta.$$

Rearranging gives

$$\begin{aligned}
\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right] \\
&= \frac{1}{2}\left[(u_1^2+u_2^2) + (v_1^2+v_2^2) - \left((u_1-v_1)^2 + (u_2-v_2)^2\right)\right] \\
&= \frac{1}{2}\left[(u_1^2+u_2^2) + (v_1^2+v_2^2) - (u_1^2 - 2u_1v_1 + v_1^2) - (u_2^2 - 2u_2v_2 + v_2^2)\right] \\
&= u_1v_1 + u_2v_2 \\
&= \mathbf{u}\cdot\mathbf{v}.
\end{aligned}$$

Hence,

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

Next time

In the next lecture we will:

- Study orthogonal sets and orthogonal matrices.
- Define an orthonormal basis of a matrix.
- Use orthogonal projections to decompose vectors between a subspace and its complement.

Chapter 6: Orthogonality
§6.2 Orthogonal Sets
§6.3 Orthogonal Projections

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Study orthogonal sets and orthogonal matrices.
- Use orthogonal projections to measure distance from a point to a line.
- Define orthogonal matrices and discuss their properties.
- Use orthogonal projections to decompose vectors between a subspace and its complement.

Orthogonal sets

Definition 1

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an *orthogonal* if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$. If, in addition, each $\mathbf{u}_i$ is a unit vector, then the set is said to be *orthonormal*.

Example 2

Consider the vectors

$$\mathbf{u} = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}.$$

Then $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$, so the set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is orthogonal. The set is *not* orthonormal because the vectors are not unit vectors. However, we could replace each one by its associated unit vector to obtain an orthonormal set with the same span.

Theorem 3

If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent.

Proof.

Write $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$. Then

$$0 = \mathbf{0} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1).$$

Since $\mathbf{u}_1 \cdot \mathbf{u}_1 \neq 0$ (because $\mathbf{u}_1 \neq 0$), then $c_1 = 0$. Repeating this argument with $\mathbf{u}_2, \ldots, \mathbf{u}_p$ gives $c_2 = \cdots = c_p = 0$. Hence, S is linearly independent. □

Orthogonal sets

The following strategy is suggested by the previous theorem:

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Let $\mathbf{u} \in W$ and write $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$. Then $\mathbf{y} \cdot \mathbf{u}_i = c_i(\mathbf{u}_i \cdot \mathbf{u}_i)$, and so

$$c_i = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \qquad i = 1, \ldots, p.$$

Example 4

Define the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}.$$

The vectors in S are orthogonal and hence the set S is linearly independent. Since S consists of 3 linearly independent vectors in $\mathbb{R}^3$, it is a basis for $\mathbb{R}^3$.

Let $\mathbf{x} = \begin{bmatrix} 8 \\ -4 \\ -3 \end{bmatrix}$. Then we can find the coordinates $[\mathbf{x}]_S = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$ using the method above. Denote the vectors in S by $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, respectively. Then

$$c_1 = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} = \frac{5}{2}, \quad c_2 = \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} = -\frac{3}{2}, \quad c_3 = \frac{\mathbf{x} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} = 2.$$

Let p be a point in $\mathbb{R}^2$ represented by the vector $\mathbf{y}$ and let $L = \text{Span}\{\mathbf{u}\}$ be a line through the origin in $\mathbb{R}^2$. Suppose we decompose $\mathbf{y}$ as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where $\hat{\mathbf{y}} \in L$ and $\mathbf{z} \in L^\perp$.

Let $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α. Then $\mathbf{z}$ is orthogonal to $\hat{\mathbf{u}}$ if and only if

$$0 = \mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - (\alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u}).$$

Hence, $\alpha = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$ and so $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. Note that if we replace $\mathbf{u}$ by $c\mathbf{u}$ for any scalar c this definition does not change and thus we have defined the projection for all of L.

Definition 5

Given vectors $\mathbf{y}, \mathbf{u} \in \mathbb{R}^n$, and $L = \text{Span}\{\mathbf{u}\}$, the *orthogonal projection of* $\mathbf{y}$ *onto* L is

$$\hat{\mathbf{y}} = \text{proj}_L\, \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}.$$

Hence, the distance from p to L is the length of $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

Example 6

Let $\mathbf{y} = \left[\begin{smallmatrix} 1 \\ 7 \end{smallmatrix}\right]$ and let $L = \text{Span}\{\mathbf{u}\}$ where $\mathbf{u} = \left[\begin{smallmatrix} -4 \\ 2 \end{smallmatrix}\right]$. Then

$$\hat{\mathbf{y}} = \text{proj}_L\, \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = \frac{1}{2}\mathbf{u} = \left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right].$$

Hence, the the distance from $\mathbf{y}$ to L is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \left\|\left[\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right]\right\| = \sqrt{45} = 3\sqrt{5}.$$

Soon we will generalize this to larger subspaces.

Orthogonal matrices

Definition 7

If W is a subspace of $\mathbb{R}^n$ spanned by an orthonormal set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then we say S is an *orthonormal basis* of W.

The standard basis is an orthonormal basis of $\mathbb{R}^n$.

Theorem 8

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Proof.

Write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \cdots \\ \mathbf{u}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \cdots & \mathbf{u}_1^T \mathbf{u}_n \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \cdots & \mathbf{u}_2^T \mathbf{u}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{u}_n^T \mathbf{u}_1 & \mathbf{u}_n^T \mathbf{u}_2 & \cdots & \mathbf{u}_n^T \mathbf{u}_n \end{bmatrix}.$$

Hence, $U^T U = I$ if and only if $\mathbf{u}_i \cdot \mathbf{u}_i = 1$ for all i and $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$. □

Theorem 9

Let U be an $m \times n$ matrix with orthonormal columns and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

(1) $\|U\mathbf{x}\| = \|\mathbf{x}\|$

(2) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

(3) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ *if and only if* $\mathbf{x} \cdot \mathbf{y} = 0$.

Proof.

We will prove (1). The rest are left as an exercise. Write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$\begin{aligned} \|U\mathbf{x}\|^2 &= U\mathbf{x} \cdot U\mathbf{x} = (\mathbf{u}_1 x_1 + \cdots \mathbf{u}_n x_n) \cdot (\mathbf{u}_1 x_1 + \cdots \mathbf{u}_n x_n) \\ &= \sum_{i,j} (\mathbf{u}_i x_i) \cdot (\mathbf{u}_j x_j) = \sum_{i,j} x_i x_j (\mathbf{u}_i \cdot \mathbf{u}_j) = \sum_i x_i^2 (\mathbf{u}_i \cdot \mathbf{u}_i) = \sum_i x_i^2 = \|\mathbf{x}\|^2 . \quad \square \end{aligned}$$

The first property says that such a matrix (called an *orthogonal matrix*) preserves length.

Projections onto subspaces

The next definition generalizes projections onto lines.

Definition 10

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. For $\mathbf{y} \in \mathbb{R}^n$, the *orthogonal projection of* $\mathbf{y}$ *onto* W is given by

$$\operatorname{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p.$$

This definition matches our previous one when W is 1-dimensional. Note that $\operatorname{proj}_W \mathbf{y} \in W$ because it is a linear combination of basis elements.

Also note that the definition simplifies when the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is orthonormal. In this case, if we let $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$, then $\operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

Theorem 11 (Orthogonal Decomposition Theorem)

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

Proof.

Note that if $W = \{\mathbf{0}\}$, then this theorem is trivial. As noted above, $\operatorname{proj}_W \mathbf{y} \in W$. We claim $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} \in W^\perp$. We have,

$$\mathbf{z} \cdot \mathbf{u}_1 = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \hat{\mathbf{y}} \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \left(\frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)(\mathbf{u}_1 \cdot \mathbf{u}_1) = \mathbf{y} \cdot \mathbf{u}_1 - \mathbf{y} \cdot \mathbf{u}_1 = 0.$$

It is clear that this holds similarly for $\mathbf{u}_2, \cdots, \mathbf{u}_p$. By linearity, $\mathbf{z} \cdot \mathbf{y} = 0$, so $\mathbf{z} \in W^\perp$.

To prove uniqueness, let $\mathbf{y} = \mathbf{w} + \mathbf{x}$ be another decomposition with $\mathbf{w} \in W$ and $\mathbf{x} \in W^\perp$. Then $\mathbf{w} + \mathbf{x} = \mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, so $(\mathbf{w} - \hat{\mathbf{y}}) = (\mathbf{z} - \mathbf{x})$. But $(\mathbf{w} - \hat{\mathbf{y}}) \in W$ and $(\mathbf{z} - \mathbf{x}) \in W^\perp$. Since $W \cap W^\perp = \{\mathbf{0}\}$, then $\mathbf{w} - \hat{\mathbf{y}} = \mathbf{0}$ so $\mathbf{w} = \hat{\mathbf{y}}$. Similarly, $\mathbf{z} = \mathbf{x}$. □

Orthogonal bases

Corollary 12

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Then $\mathbf{y} \in W$ if and only if $\operatorname{proj}_W \mathbf{y} = \mathbf{y}$.

Example 13

Let $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} -3 \\ 5 \\ 0 \end{bmatrix}$. Let $W = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Then

$$\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = (-1)\mathbf{u}_1 - \frac{4}{3}\mathbf{u}_2 = \frac{1}{3}\begin{bmatrix} -13 \\ 7 \\ 2 \end{bmatrix}.$$

We can decompose $\mathbf{y}$ as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \frac{1}{3}\begin{bmatrix} 4 \\ 8 \\ -2 \end{bmatrix}.$$

Theorem 14 (Best Approximation Theorem)

Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. Then $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$ is the closest point in W to $\mathbf{y}$ in the sense that $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$ for all $\mathbf{v} \in W$, $\mathbf{v} \neq \hat{\mathbf{y}}$.

Next time

In the next lecture we will:

- Use the Gram-Schmidt process to construct an orthogonal basis for any subspace.
- Use the method of least squares to find best approximate solutions of inconsistent systems.

Chapter 6: Orthogonality
§6.4 The Gram-Schmidt Process
§6.5 The Least-Squares Problems
§6.6 Applications to Linear Models

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Use the Gram-Schmidt process to construct an orthogonal basis for any subspace.
- Use the method of least squares to find best approximate solutions of inconsistent systems.

Projections onto subspaces

We begin by recalling our (general) definition of orthogonal projections.

Definition 1

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$. For $\mathbf{y} \in \mathbb{R}^n$, the *orthogonal projection of* $\mathbf{y}$ *onto* W is given by

$$\mathrm{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p.$$

Orthogonal projections give us a way to find an orthogonal basis for any subspace W of $\mathbb{R}^n$. Recall that projections provide "best approximations" in the sense of the following theorem.

Theorem 2 (Best Approximation Theorem)

Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. Then $\hat{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$ is the closest point in W to $\mathbf{y}$ in the sense that $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$ for all $\mathbf{v} \in W$, $\mathbf{v} \neq \hat{\mathbf{y}}$.

Constructing an orthogonal basis

Example 3

Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$. We will construct an orthogonal basis for W.

Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\{\mathbf{v}_1\}$. It suffices to find a vector $\mathbf{v}_2 \in W$ orthogonal to W_1.

Let $\mathbf{p} = \text{proj}_{W_1} \mathbf{x}_2 \in W_1$. Then $\mathbf{x}_2 = \mathbf{p} + (\mathbf{x}_2 - \mathbf{p})$ where $\mathbf{x}_2 - \mathbf{p} \in W_1^{\perp}$. Let

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \begin{bmatrix} 9 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}.$$

Now $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and $\mathbf{v}_1, \mathbf{v}_2 \in W$. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W.

Note that if we wanted an orthonormal basis for W then we can just take the unit vectors associated to $\mathbf{v}_1$ and $\mathbf{v}_2$.

This process could continue. Say W was three-dimensional. We could then let $W_2 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and find the projection of $\mathbf{x}_3$ onto W_2. We'll prove the next theorem using this idea.

The Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace $W \subset \mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition,

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$$

for all $1 \leq k \leq p$.

Proof of The Gram-Schmidt Process.

For $1 \leq k \leq p$, set $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and $V_k = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Since $\mathbf{v}_1 = \mathbf{x}_1$, then it holds (trivially) that $W_1 = V_1$ and $\{\mathbf{v}_1\}$ is orthogonal.

Suppose for some k, $1 \leq k < p$, that $W_k = V_k$ and that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \in W_k^{\perp} \subset W_{k+1}.$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Since $\mathbf{x}_{k+1} \in W_{k+1}$, then $\mathbf{v}_{k+1} \in W_{k+1}$. Hence, $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set of $k+1$ nonzero vectors in W_{k+1} and hence a basis of W_{k+1}. Hence, $W_{k+1} = V_{k+1}$. The result now follows by induction. $\square$

Example 4

Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

We will construct an orthogonal basis for W.

Set $\mathbf{v}_1 = \mathbf{x}_1$. Then

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix}.$$

Now,

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} - \frac{5}{2}\begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}.$$

Hence, an orthogonal basis for W is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

In data science, one often wants to be able to approximate a set of data by a curve. One might hope to construct the line that best fits the data. This is known (by one name) as *linear regression*. We will study a linear algebraic approach to this problem.

Suppose the system $A\mathbf{x} = \mathbf{b}$ is inconsistent. Previously, we gave up all hope then of "solving" this system because no solution existed. However, if we give up the idea that we must find an *exact* solution and instead focus on finding an *approximate* solution, then we may have hope of solving.

Definition 5

If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, a *least-squares* solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$.

Geometrically, we think of $A\hat{\mathbf{x}}$ as the projection of $\mathbf{b}$ onto $\operatorname{Col} A$. That is, if $\hat{\mathbf{b}} = \operatorname{proj}_{\operatorname{Col} A} \mathbf{b}$, then the equation $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent. Let $\hat{\mathbf{x}} \in \mathbb{R}^n$ be a solution (there may be several). By the Best Approximation Theorem, $\hat{\mathbf{b}}$ is the point on $\operatorname{Col} A$ closest to $\mathbf{b}$ and so $\hat{\mathbf{x}}$ is a least-squares solution to $A\mathbf{x} = \mathbf{b}$.

By the Orthogonal Decomposition Theorem, $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to $\operatorname{Col} A$. Hence, if $\mathbf{a}_j$ is any column of A, then

$$\mathbf{a}_j \cdot (\mathbf{b} - \hat{\mathbf{b}}) = 0.$$

That is, $\mathbf{a}_j^T(\mathbf{b} - \hat{\mathbf{b}}) = 0$. But $\mathbf{a}_j^T$ is a row of A^T and so

$$A^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}.$$

Replacing $\hat{\mathbf{b}}$ with $A\hat{\mathbf{x}}$ and expanding we get

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

The equations corresponding to this system are the *normal equations* for $A\mathbf{x} = \mathbf{b}$. We have now essentially proven the following theorem.

Theorem 6

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.

Example 7

Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 5 & 1 \\ 4 & 1 \\ 3 & 0 \end{bmatrix} \text{ and } \mathbf{b} = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right].$$

We will use normal equations. First we compute

$$A^T A = \begin{bmatrix} 50 & 9 \\ 9 & 2 \end{bmatrix}, \qquad A^T\mathbf{b} = \left[\begin{smallmatrix} 5 \\ 1 \end{smallmatrix}\right].$$

To solve the equation $A^T A\mathbf{x} = A^T\mathbf{b}$ we invert $A^T A$.

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b} = \frac{1}{19}\begin{bmatrix} 2 & -9 \\ -9 & 50 \end{bmatrix}\left[\begin{smallmatrix} 5 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1/19 \\ 5/19 \end{smallmatrix}\right].$$

Note in general, $A^T A$ need not be invertible.

Theorem 8

Let A be an $m \times n$ matrix. The following statements are equivalent:

(1) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

(2) The columns of A are linearly independent.

(3) The matrix A^TA is invertible.

.

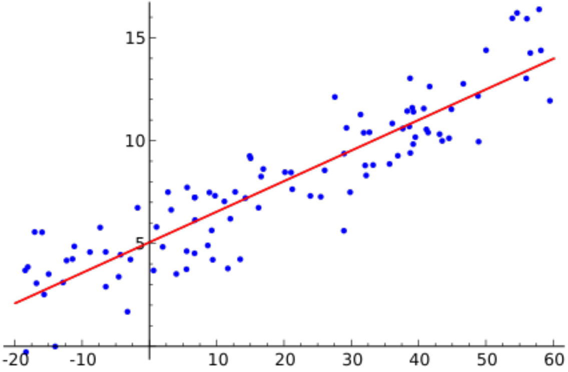
Hence, when A^TA is invertible then the least-squares solution $\hat{\mathbf{x}}$ is unique and

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

Linear regression

We will see how to fit data to a line using least-squares. We denote the equation $A\mathbf{x} = \mathbf{b}$ by $X\beta = \mathbf{y}$. The matrix X is referred to as the *design matrix*, β as the *parameter vector*, and $\mathbf{y}$ as the *observation vector*.

We will model a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ by a line. Denote this line by $y = \beta_0 + \beta_1 x$. The *residual* of a point (x_i, y_i) is the distance from that point to the line. The *least-squares line* minimizes the sum of the squares of the residuals.



Suppose the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ were all on the line. Then they would satisfy,

$$\begin{aligned} \beta_0 + \beta_1 x_1 &= y_1 \\ &\vdots \\ \beta_0 + \beta_1 x_n &= y_n. \end{aligned}$$

We could write this system as $X\beta = \mathbf{y}$ where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

If the data does not lie on the line, then we want the vector β to be the least-squares solution of $X\beta = \mathbf{y}$ that minimizes the distance between $X\beta$ and $\mathbf{y}$.

Example 9

Given the data points $(4,1)$, $(1,2)$, $(3,3)$, $(5,5)$. We will find the equation $y = \beta_0 + \beta_1 x$.

We build the matrix X and vector $\mathbf{y}$ from the data,

$$X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{y} = \left[\begin{smallmatrix} 1 \\ 2 \\ 3 \\ 5 \end{smallmatrix}\right].$$

For the least-squares solution of $X\beta = \mathbf{y}$, we have the normal equation $X^T X\beta = X^T\mathbf{y}$ where

$$X^T X = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}, \quad X^T\mathbf{y} = \left[\begin{smallmatrix} 11 \\ 37 \end{smallmatrix}\right].$$

Hence,

$$\left[\begin{smallmatrix} \beta_0 \\ \beta_1 \end{smallmatrix}\right] = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} 41/35 \\ 17/35 \end{bmatrix}.$$

Next time

In the next lecture we will:

- Study the special properties of symmetric matrices.
- Prove the Spectral Theorem for symmetric matrices.

Chapter 7: Symmetric Matrices
§7.1 Diagonalization of Symmetric Matrices

MTH 222

Linear Algebra

Goals for today:

- Study the special properties of symmetric matrices.
- Prove the Spectral Theorem for symmetric matrices.

Symmetric matrices

We have seen already that it is quite time intensive to determine whether a matrix is diagonalizable. We'll see that there are certain cases when a matrix is always diagonalizable.

Definition 1

A matrix A is *symmetric* if $A^T = A$.

Example 2

The matrix

$$A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

is symmetric.

It is (reasonably) easy to show that the product of symmetric matrices is symmetric, and the inverse of a symmetric matrix is symmetric.

Example 3

Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$.

The eigenvalues of A are -2 and 7. The eigenspaces have bases,

$$\lambda = 7, \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} \right\}, \qquad \lambda = -2, \left\{ \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix} \right\}.$$

Hence, A is diagonalizable. We use Gram-Schmidt to find an orthogonal basis for $\mathbb{R}^3$. Note that the eigenvector for $\lambda = -2$ is already orthogonal to both eigenvectors for $\lambda = 7$.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Finally, we normalize each vector,

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Now the matrix $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is orthogonal and so $U^T U = I$.

The next theorem is stronger than a previous result for general matrices.

Theorem 4

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof.

Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors for A with corresponding eigenvalues λ_1, λ_2, $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1^T(\lambda_2 \mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

Hence, $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, then we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. $\square$

Based on the previous theorem, we say that the eigenspaces of A are *mutually orthogonal*.

Definition 5

An $n \times n$ matrix A is *orthogonally diagonalizable* if there exists an orthogonal $n \times n$ matrix P and a diagonal matrix D such that $A = PDP^T$.

Theorem 6

If A is orthogonally diagonalizable, then A is symmetric.

Proof.

Since A is orthogonally diagonalizable, then $A = PDP^T$ for some orthogonal matrix P and diagonal matrix D. A is symmetric because

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A. \qquad \square$$

It turns out the converse of the above theorem is also true!

The set of eigenvalues of a matrix A is called the *spectrum of A* and is denoted σ_A.

Theorem 7 (The Spectral Theorem for symmetric matrices)

Let A be a (real) $n \times n$ symmetric matrix. Then the following hold.

(1) A has n real eigenvalues, counting multiplicities.

(2) For each eigenvalue λ of A, $\mathrm{geomult}_\lambda(A) = \mathrm{algmult}_\lambda(A)$.

(3) The eigenspaces are mutually orthogonal.

(4) A is orthogonally diagonalizable.

Proof.

Every eigenvalue of a symmetric matrix is real. The second part of (1) as well as (2) are immediate consequences of (4). We proved (3) previously. Note that (4) is trivial when A has *n distinct* eigenvalues by (3).

We prove (4) by induction. Clearly the result holds when A is 1×1. Assume $(n-1) \times (n-1)$ symmetric matrices are orthogonally diagonalizable.

Let A be $n \times n$ and let λ_1 be an eigenvalue of A and $\mathbf{u}_1$ a (unit) eigenvector for λ_1. Set $W = \mathrm{Span}\{\mathbf{u}_1\}$. Then by the Gram-Schmidt process we may extend $\mathbf{u}_1$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ where $\{\mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for $W^\perp$.

Theorem 8 (The Spectral Theorem for symmetric matrices)

Let A be a (real) $n \times n$ symmetric matrix. Then the following hold.

(4) A is orthogonally diagonalizable.

Proof.

Set $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$U^T A U = \begin{bmatrix} \mathbf{u}_1^T A \mathbf{u}_1 & \cdots & \mathbf{u}_1^T A \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n^T A \mathbf{u}_1 & \cdots & \mathbf{u}_n^T A \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}.$$

The first column is as indicated because $\mathbf{u}_i^T A \mathbf{u}_1 = \mathbf{u}_i^T(\lambda \mathbf{u}_1) = \lambda(\mathbf{u}_i \cdot \mathbf{u}_1) = \lambda \delta_{ij}$. As $U^T A U$ is symmetric, $* = 0$ and B is a symmetric $(n-1) \times (n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \ldots, \lambda_n$ (by the inductive hypothesis). Because A and $U^T A U$ are similar, then the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

Theorem 9 (The Spectral Theorem for symmetric matrices)

Let A be a (real) $n \times n$ symmetric matrix. Then the following hold.

(4) A is orthogonally diagonalizable.

Proof.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q such that $Q^T BQ = D$, where the diagonal entries of D are $\lambda_2, \ldots, \lambda_n$. Now

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}^T \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q^T BQ \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D \end{bmatrix}.$$

Note that $\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal. Set $V = U \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$. As the product of orthogonal matrices is orthogonal, V is itself orthogonal and $V^T AV$ is diagonal. □

The Spectral Theorem

Let A be orthogonally diagonalizable. Then $A = UDU^T$ where

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$$

and D is the diagonal matrix whose diagonal entries are the eigenvalues of A: $\lambda_1, \ldots, \lambda_n$. Then

$$A = UDU^T = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T.$$

This is known as the *spectral decomposition of A*. Each $\mathbf{u}_i \mathbf{u}_i^T$ is called a *projection matrix* because $(\mathbf{u}_i \mathbf{u}_i^T)\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\text{Span}\{\mathbf{u}_i\}$.

Example 10

Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$. The orthonormal basis of Col(A) was

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Setting $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ gives $U^TAU = D = \text{diag}(7, 7, -2)$. The projection matrices are

$$\mathbf{u}_1\mathbf{u}_1^T = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \mathbf{u}_2\mathbf{u}_2^T = \frac{1}{18}\begin{bmatrix} 1 & -4 & -1 \\ -4 & 16 & 4 \\ -1 & 4 & 1 \end{bmatrix}, \mathbf{u}_3\mathbf{u}_3^T = \frac{1}{9}\begin{bmatrix} 4 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 4 \end{bmatrix}.$$

The spectral decomposition is

$$7\mathbf{u}_1\mathbf{u}_1^T + 7\mathbf{u}_2\mathbf{u}_2^T - 2\mathbf{u}_3\mathbf{u}_3^T = A.$$

Next time

In the next lecture we will:

- Begin a study of abstract vector spaces.
- Consider examples of vector spaces beyond $\mathbb{R}^n$.
- Define a subspace (in general) and find examples of subspaces.

Chapter 4: Vector Spaces
§4.1 Vector spaces and subspaces

MTH 222

Linear Algebra

Goals for today:

- Begin a study of abstract vector spaces.
- Consider examples of vector spaces beyond $\mathbb{R}^n$.
- Define a subspace (in general) and find examples of subspaces.

Though we are studying "abstract" vector spaces, we will see very quickly that the examples we will study are very much concrete (and familiar)! The point of abstractifying our notion of a vector space is so that we can apply the tools we have developed this semester to other "spaces" that exhibit linear properties.

We will continue to study *real* vector spaces, but this is different than saying $\mathbb{R}^n$. Recall that $\mathbb{R}^n$ is the set of n-dimensional vectors with entries in $\mathbb{R}$. It comes with two operations, addition and scalar multiplications. This is a real vector space because the scalars are real numbers. It turns out that $\mathbb{C}^n$ is a real vector space too!

However, a key takeaway from this unit should be that these (real) vector spaces behave just like $\mathbb{R}^n$. In fact, they behave so much like $\mathbb{R}^n$ that we will be able to show they are *isomorphic* to some $\mathbb{R}^n$.

Definition 1 (Vector Space)

A *vector space* (over $\mathbb{R}$) is a nonempty set V of objects (called *vectors*) along with two operations: addition and multiplication by scalars in $\mathbb{R}$, subject to the following axioms:

(1) If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} \in V$;

(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in V$;

(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$;

(4) There exists an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$;

(5) For all $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;

(6) For all $\mathbf{v} \in V$ and $c \in \mathbb{R}$, $c\mathbf{v} \in V$;

(7) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in V$, $c \in \mathbb{R}$;

(8) $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$ for all $c, d \in \mathbb{R}$, $\mathbf{v} \in V$;

(9) $c(d\mathbf{v}) = (cd)\mathbf{v}$ for all $c, d \in \mathbb{R}$, $\mathbf{v} \in V$;

(10) $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.

Whew!

Example 2

The following are vector spaces under the standard operations:

(1) $\mathbb{R}^n$;

(2) $\mathbb{C}^n$;

(3) $\mathcal{M}_n$, $n \times n$ matrices (with entries in $\mathbb{R}$);

(4) $\mathcal{P}_n$, polynomials (with coefficients in $\mathbb{R}$) of degree at most n;

(5) $\mathbb{S}$, the space of doubly infinite sequences:

$$\{y_k\} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)$$

(6) For $[a, b] \subset \mathbb{R}$, the set $\mathcal{C}([a, b])$ of continuous real-valued functions with domain $[a, b]$.

Checking the vector space axioms is an easy (but tiresome) exercise in most cases. We will check two important examples. For the others we will accept that they are vector spaces. Our primary focus will be on *subspaces*.

Example 3

We have shown previously that $\mathcal{M}_n$ is vector space (we just didn't use that terminology). Here we review what that means.

First note that $\mathcal{M}_n$ is closed under matrix addition (1) and scalar multiplication (6) by definition. Moreover, we checked previously that matrix addition is commutative (2) and associative (3). The $\mathbf{0}$ element is the $n \times n$ zero matrix. It is easy to check that for all $M \in \mathcal{M}_n$, $M + \mathbf{0} = M$ (4) and $M + (-1)M = \mathbf{0}$ (5). We further checked that matrix addition and scalar multiplication distribute over one another (7,8,9). Finally, (10) is clear.

It is (or should be) reasonably clear that the above applies equally to non-square matrices.

Example 4

We claim $\mathcal{P}_n$ is a vector space. Note that $\mathcal{P}_n$ consists of polynomials of the form

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n,$$

with $a_i \in \mathbb{R}$. If all coefficients are zero, then $p(t)$ is the *zero polynomial*.

Let $p(t), q(t) \in \mathcal{P}_n$. Write $p(t)$ as above and $q(t) = b_0 + b_1 t + \cdots + b_n t^n$, $b_i \in \mathbb{R}$. Then

$$(p+q)(t) = p(t) + q(t) = (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n \in \mathcal{P}_n.$$

For $c \in \mathbb{R}$,

$$(cp)(t) = cp(t) = (ca_0) + (ca_1)t + \cdots + (ca_n)t \in \mathcal{P}_n.$$

Thus, axioms (1) and (6) are satisfied by definition. Axioms (2), (3), and (4) are clear. The additive inverse of $p(t)$ is $-p(t) = (-a_0) + (-a_1)t + \cdots + (-a_n)t^n$, so axiom (5) is satisfied. Axioms (7)-(10) are also easy to check.

Theorem 5

Let V be a vector space. For each $\mathbf{u} \in V$ and $c \in \mathbb{R}$,

(a) $0\mathbf{u} = \mathbf{0}$,

(b) $c\mathbf{0} = \mathbf{0}$,

(c) $-\mathbf{u} = (-1)\mathbf{u}$.

Proof.

For (a), let $\mathbf{y} = 0\mathbf{u}$. Then by axiom (8),

$$\mathbf{y} = 0\mathbf{u} = (0+0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u} = \mathbf{y} + \mathbf{y}.$$

Subtracting $\mathbf{y}$ on both sides gives $\mathbf{y} = \mathbf{0}$. Part (b) is similar. For part (c), we have by part (a),

$$\mathbf{0} = 0\mathbf{u} = (1 + (-1))\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}.$$

Subtracting $\mathbf{u}$ on both sides gives $-\mathbf{u} = (-1)\mathbf{u}$. □

Definition 6

A *subspace* of a vector space V is a subset H of V that has three properties:

(1) The zero vector of V is in H,

(2) H is closed under addition (if $\mathbf{u}, \mathbf{v} \in H$ then $\mathbf{u} + \mathbf{v} \in H$);

(3) closed under scalar multiplication (if $\mathbf{u} \in H$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in H$).

As before, we could replace the first condition with the equivalent condition that $H \neq \emptyset$.

Note that subspaces are vector spaces themselves. Another way to formulate the definition of a subspace is that H is a nonempty subset of V that is a vector space under the operations of V.

Example 7

Let V be a vector space. Then $\{\mathbf{0}\}$ is a subspace of V (called the *trivial subspace*). Also, V is a subspace of itself. Any subspace H of V such that $H \neq V$ is called a *proper subspace* of V.

Example 8

Recall that for an $m \times n$ matrix A, $\operatorname{Col} A$ is a subspace of $\mathbb{R}^n$ and $\operatorname{Nul} A$ is a subspace of $\mathbb{R}^m$.

Example 9

Let $m \leq n$, then $\mathcal{P}_m$ is a subspace of $\mathcal{P}_n$.

Definition 10

The *trace* of an $n \times n$ matrix M is defined as the sum of the diagonal entries, denoted $\text{tr}(M)$.

Example 11

Let $S \subset \mathcal{M}_n$ denote the set of $n \times n$ matrices of trace zero. We will show that S is a subspace of $\mathcal{M}_n$.

Clearly, the zero matrix has trace zero so it is an element of $\mathcal{S}$. Let $A, B \in S$ with $A = (a_{ij})$ and $B = (b_{ij})$. Then

$$\text{tr}(A + B) = \sum_{i=1}^{n} a_{ii} + b_{ii} = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = 0 + 0 = 0.$$

Thus, S is closed under addition. Now let $c \in \mathbb{R}$, then

$$\text{tr}(cA) = \sum_{i=1}^{n} (ca_{ii}) = c \sum_{i=1}^{n} a_{ii} = c \cdot 0 = 0.$$

Thus, S is closed under scalar multiplication and is therefore a subspace.

Example 12

Let $H \subset \mathcal{M}_n$ denote the set of $n \times n$ matrices of determinant zero. Clearly, H contains the zero matrix. Moreover, by our determinant rules, if $M \in H$ and $c \in \mathbb{R}$, then

$$\det(cM) = c^n \det(M) = 0.$$

Thus, $cM \in H$. However, if $M, N \in H$, then $M + N$ need not be an element of H. Consider the following example in $n = 2$.

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then $M + N = I_2 \notin H$. Hence, H is not a subspace of $\mathcal{M}_n$.

Next time

In the next lecture we will:

- Define a basis for a subspace of a vector space.
- Define linear transformations for vector spaces.
- Compute the kernel and image of linear transformations.
- Use a coordinate transformation to show that finite dimensional vector spaces are isomorphic to some $\mathbb{R}^n$.

Chapter 4: Vector Spaces
§4.2 Linear Transformations
§4.3 Linearly independent sets; Bases
§4.4 Coordinate systems

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define a basis for a subspace of a vector space.
- Define linear transformations for vector spaces.
- Compute the kernel and image of linear transformations.
- Use a coordinate transformation to show that finite dimensional vector spaces are isomorphic to some $\mathbb{R}^n$.

Old terms new again

Definition 1

Let V be a vector space and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subset V$.

- A *linear combination* in V is a sum of scalar multiples of elements of V.
- The *span* of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the set of all linear combinations of elements of V, denoted $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.
- The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is *linearly independent* if $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ for $c_i \in \mathbb{R}$ implies $c_i = 0$ for all i. When there exist scalars c_i not all zero such that $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$, then set is *linearly dependent*.

Theorem 2

Given elements $\mathbf{v}_1, \ldots, \mathbf{v}_p$ *of a vector space* V, $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ *is a subspace of* V.

Definition 3

Let H be a subspace of a vector space V. A *basis* of H is a linearly independent set that spans H.

Example 4

Let $S = \{1, t, t^2, \ldots, t^n\} \subset \mathcal{P}_n$. We will show that S is a basis of $\mathcal{P}_n$.

Let $p(t) \in \mathcal{P}_n$, so $p(t) = a_0 + a_1 t + \cdots + a_n t^n$ for some scalars a_i. Hence, $p(t) \in S$ so S spans $\mathcal{P}_n$.

For linear independence, suppose $q(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n = \mathbf{0}(t)$ (the zero polynomial). The only polynomial in $\mathcal{P}_n$ with n zeros is the zero polynomial itself (Fundamental Theorem of Algebra), so $c_1 = \cdots = c_n = 0$. Hence, S is linearly independent and thus a basis.

Definition 5

A *linear transformation* $T : V \to W$ between vector spaces V and W is a rule that assigns to each vector $\mathbf{x} \in V$ a unique vector $T(\mathbf{x}) \in W$ such that

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$;

(2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u} \in V$, $c \in \mathbb{R}$.

Every linear transformations over $\mathbb{R}^n \to \mathbb{R}^m$ can be described by some matrix. We will see that the same holds for linear transformations over vector spaces.

Example 6

Let $[a, b] \subset \mathbb{R}$ be an interval and set $V = \mathcal{C}([a, b])$. Let $D : V \to V$ be the map defined by $D(f) = f'$, the derivative of f.

For all $f, g \in V$ and $c \in \mathbb{R}$, we have (by elementary calculus)

$$D(f + g) = (f + g)' = f' + g' = D(f) + D(g)$$
$$D(cf) = (cf)' = cf' = cD(f).$$

Thus, the differentiation operator D is a linear transformation on V.

Definition 7

Let $T : V \to W$ be a linear transformation. The *kernel* of T is the set

$$\text{Ker } T = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$$

and the *image* of T is the set

$$\text{Im } T = \{\mathbf{w} \in W : T(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\}.$$

(The image is also called the *range* in the text.) A one-to-one and onto linear transformation is called an *isomorphism* and in this case we say V and W are *isomorphic*.

Example 8

Again let $V = \mathcal{C}([a, b])$ for an interval $[a, b] \subset \mathbb{R}$ and let $D : V \to V$ be differentiation. Then $\text{Ker } D$ is the set of all constant functions. On the other hand, $\text{Im } D = V$.

Example 9

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let A be the $(m \times n)$ standard matrix.

Since $T(\mathbf{v}) = \mathbf{0}$ if and only if $A\mathbf{v} = \mathbf{0}$, the it is easy to see that $\text{Ker } T = \text{Nul } A$.

Since $T(\mathbf{e}_i) = \mathbf{a}_i$ for $i = 1, \dots, n$, then $\text{Col } A \subset \text{Im } T$. On the other hand, if $\mathbf{y} \in \text{Im } T$, then there exists $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{y}$. But $\mathbf{x} = x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n$ so

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n) = x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \in \text{Col } A.$$

Hence, $\text{Im } T \subset \text{Col } A$ so in fact $\text{Im } T = \text{Col } A$.

Theorem 10

Let V and W be vector spaces and let $T : V \to W$ be a linear transformation.

(1) $T(\mathbf{0}) = \mathbf{0}$.

(2) Ker T *is a subspace of* V.

(3) Im T *is a subspace of* W.

Proof.

(1) By linearity, $T(\mathbf{0}) + T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0})$. Hence, $T(\mathbf{0}) = \mathbf{0}$.

(2) By (1), $\mathbf{0} \in \text{Ker } T$. Now suppose $\mathbf{v}_1, \mathbf{v}_2 \in \text{Ker } T$. Then

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

so $\mathbf{v}_1 + \mathbf{v}_2 \in \text{Ker } T$. Finally, let $\mathbf{v} \in \text{Ker } T$ and $c \in \mathbb{R}$. Then

$$T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{0} = \mathbf{0},$$

so $c\mathbf{v} \in V$. Hence, Ker T is a subspace of V.

Theorem 11

Let V and W be vector spaces and let $T : V \to W$ be a linear transformation.

(1) $T(\mathbf{0}) = \mathbf{0}$.

(2) $\operatorname{Ker} T$ *is a subspace of* V.

(3) $\operatorname{Im} T$ *is a subspace of* W.

Proof.

(3) Again by (1), $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in \operatorname{Im} T$. Let $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{Im} T$. Then there exists $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Thus,

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2,$$

so $\mathbf{w}_1 + \mathbf{w}_2 \in \operatorname{Im} T$. Finally, let $\mathbf{w} \in \operatorname{Im} T$ and $c \in \mathbb{R}$. Then there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$ so

$$T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{w}.$$

Thus, $c\mathbf{w} \in \operatorname{Im} T$ and $\operatorname{Im} T$ is a subspace of W. □

Coordinates

We will be particularly interested in the coordinate mapping $T : H \to \mathbb{R}^p$. This will be used to show that ever finite-dimensional (real) vector spaces are *isomorphic* to some $\mathbb{R}^p$. The following result we know for $\mathbb{R}^n$ and its subspaces.

Theorem 12

Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a basis for a vector space V and let $\mathbf{y} \in V$. There exist unique scalars $c_1, \ldots, c_p \in \mathbb{R}$ such that $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{y}$.

Proof.

Existence follows from the fact that S spans V. Suppose there exist two sets of scalars, $c_1, \ldots, c_p \in \mathbb{R}$ and $d_1, \ldots, d_p \in \mathbb{R}$ such that

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p \quad \text{and}$$
$$\mathbf{y} = d_1\mathbf{v}_1 + \cdots + d_p\mathbf{v}_p.$$

Subtracting these two equations gives

$$\mathbf{0} = (c_1 - d_1)\mathbf{v}_1 + \cdots + (c_p - d_p)\mathbf{v}_p.$$

By linear independence, $c_i = d_i$ for $i = 1, \ldots, p$. □

Definition 13

Let H be a subspace of a vector space V and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis of H. For $\mathbf{y} \in H$, the *coordinates* of $\mathbf{y}$ are the unique scalars $c_1, \ldots, c_p$ such that

$$c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p = \mathbf{y}.$$

The *coordinate vector* is

$$[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

The map $T_{\mathcal{B}} : H \to \mathbb{R}^p$ defined by $\mathbf{y} \mapsto [\mathbf{y}]_{\mathcal{B}}$ is the *coordinate mapping* (determined by $\mathcal{B}$).

Example 14

The set $\mathcal{B} = \{1, 1+t, 1+t+t^2\}$ is a basis of $\mathcal{P}_2$. Consider $p(t) = 2 - 3t + 5t^2$. Then

$$p(t) = 5(1) - 8(1+t) + 5(1+t+t^2).$$

Hence

$$T_{\mathcal{B}}(p(t)) = [p(t)]_{\mathcal{B}} = \begin{bmatrix} 5 \\ -8 \\ 5 \end{bmatrix}.$$

Coordinate Mapping

Theorem 15

Let H be a subspace of a vector space V and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis of H. The coordinate mapping $T_{\mathcal{B}} : H \to \mathbb{R}^p$ is an isomorphism (1-1 and onto linear transformation).

Example 16

Since $\{1, t, \ldots, t^n\}$ is a basis of $\mathcal{P}_n$, then by the theorem $\mathcal{P}_n$ is isomorphic to $\mathbb{R}^{n+1}$.

Example 17

A basis of $\mathcal{M}_2$ is

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Hence, $\mathcal{M}_2$ is isomorphic to $\mathbb{R}^4$.

We can now leverage this to define dimension for vector subspaces.

The proof of these facts are almost identical to those in the $\mathbb{R}^n$ case.

Theorem 18

Let V be a (finite-dimensional) vector space with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$. Then every basis of V has p elements.

Definition 19

The *dimension* of a subspace H of a vector space V is the number of vectors in any basis of H.

The previous theorem implies that this definition is well-defined.

Theorem 20 (The Basis Theorem)

Let V be a p-dimensional vector subspace, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Also, any set of p elements that spans V is automatically a basis for V.

Here is a generalization of a result from earlier in the semester.

Definition 21

Let $T : V \to W$ be a linear transformation. The *rank* of T is $\operatorname{rank} T = \dim \operatorname{Im} T$ and the *nullity* of T is $\operatorname{nul} T = \dim \operatorname{Ker} T$.

Theorem 22 (Rank-Nullity Theorem)

Let $T : V \to W$ be a linear transformation and let V be finite-dimensional. Then

$$\operatorname{rank} T + \operatorname{nul} T = \dim V.$$

Next time

In the next lecture we will:

- Define inner product spaces and consider several examples.
- Study length, distance, and orthogonality in a more general context.
- Discuss an application of inner product spaces to Fourier analysis.

Chapter 6: Orthogonality
§6.7 Inner Product Spaces
§6.8 Applications of Inner Product Spaces

MTH 222

Linear Algebra

It's good to have goals

Goals for today:

- Define inner product spaces and consider several examples.
- Study length, distance, and orthogonality in a more general context.
- Discuss an application of inner product spaces to Fourier analysis.

The following definition generalizes the inner (dot) product on $\mathbb{R}^n$.

Definition 1

An *inner product* on a vector space V is a function that, to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $c \in \mathbb{R}$:

(1) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

(2) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(3) $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

(4) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

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

We could also say that the inner product is a map $V \times V \to \mathbb{R}$ satisfying Axioms (1)-(4).

Example 2

Let $\mathbf{u} = \left[\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right]$ and $\mathbf{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ be vectors in $\mathbb{R}^2$. We define an map $\langle -, - \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2.$$

We will show that this defines an inner product on $\mathbb{R}^2$.

Since $4u_1v_1 + 5u_2v_2 = 4v_1u_1 + 5v_2u_2$, then $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$, so Axiom (1) is satisfied. If $\mathbf{w} = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$, then

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 \\ &= 4u_1w_1 + 4v_1w_1 + 5u_2w_2 + 5v_2w_2 \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle, \end{aligned}$$

so Axiom (2) is satisfied. Let $c \in \mathbb{R}$, then

$$\langle c\mathbf{u}, \mathbf{v} \rangle = 4(cu_1)v_1 + 5(cu_2)v_2 = c(4u_1v_1 + 5u_2v_2) = c\langle \mathbf{u}, \mathbf{v} \rangle,$$

so Axiom (3) is satisfied. Finally, $\langle \mathbf{u}, \mathbf{u} \rangle = 4u_1^2 + 5u_2^2 \geq 0$. Moreover, since $4u_1^2$ and $5u_2^2$ are nonnegative, then $4u_1^2 + 5u_2^2 = 0$ if and only if $u_1 = u_2 = 0$.

Example 3

Let $t_0, t_1, \ldots, t_n$ be distinct real numbers. For polynomials $p, q \in \mathcal{P}_n$, define

$$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n).$$

We will show that this defines an inner product on $\mathcal{P}_n$.

Since $p(t_i)$ and $q(t_i)$ are real numbers, then $p(t_i)q(t_i) = q(t_i)p(t_i)$ for all i. Hence $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$, so Axiom (1) is satisfied. If $r \in \mathcal{P}_n$

$$\begin{aligned}\langle p+q, r \rangle &= (p+q)(t_0)r(t_0) + (p+q)(t_1)r(t_1) + \cdots + (p+q)(t_n)r(t_n)\\ &= (p(t_0)+q(t_0))r(t_0) + (p(t_1)+q(t_1))r(t_1) + \cdots + (p(t_n)+q(t_n))r(t_n)\\ &= p(t_0)r(t_0) + q(t_0)r(t_0) + \cdots + p(t_n)r(t_n) + q(t_n)r(t_n)\\ &= \langle p, r \rangle + \langle q, r \rangle,\end{aligned}$$

so Axiom (2) is satisfied. Axiom (3) is checked similarly. Finally,

$$\langle p, p \rangle = p(t_0)^2 + p(t_1)^2 + \cdots + p(t_n)^2 \geq 0.$$

Hence $\langle p, p \rangle = 0$ if and only if $p(t_i) = 0$ for each i. Then p is a polynomial of degree (at most) n with $n+1$ zeros, so $p = \mathbf{0}$ (the zero polynomial).

Length, distance, and orthogonality

Definition 4

Let V be an inner product space with inner product denoted $\langle -, - \rangle$.

- The *length* of a vector $\mathbf{v} \in V$ is
$$\|v\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$
Equivalently, $\|v\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle$.
- A *unit vector* is one whose length is 1.
- The *distance* between $\mathbf{u}$ and $\mathbf{v}$ in V is $\|\mathbf{u} - \mathbf{v}\|$.
- Vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* is $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example 5

Let $t_0, t_1, \ldots, t_n$ be distinct real numbers. For polynomials $p, q \in \mathcal{P}_n$, define

$$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n).$$

Suppose $n = 2$. Set $t_0 = 0$, $t_1 = 1/2$, and $t_2 = 1$.

If $p(t) = 12t^2$ and $q(t) = 2t - 1$, then

$$\begin{aligned}
\langle p, q \rangle &= p(t_0)q(t_0) + p(t_1)q(t_1) + p(t_2)q(t_2) \\
&= p(0)q(0) + p(1/2)q(1/2) + p(1)q(1) \\
&= (0)(-1) + (3)(0) + (12)(1) = 12.
\end{aligned}$$

We also have

$$\|q\| = \sqrt{\langle q, q \rangle} = \sqrt{q(0)q(0) + q(1/2)q(1/2) + q(1)q(1)} = \sqrt{2}.$$

Given an orthogonal basis for a subspace W of a vector space V, the projection $\text{proj}_W \mathbf{v}$ is defined in the same way as in $\mathbb{R}^n$. One can also define a version of the Gram-Schmidt process for inner product spaces. The Best Approximation Theorem in this case says that $\text{proj}_W \mathbf{v}$ is the vector in W closest to $\mathbf{v}$ with regards to the distance defined relative to $\langle -, - \rangle$.

Example 6

Let $p(t) = 12t^2$ and $q(t) = 2t - 1$ be polynomials in $\mathcal{P}_2$. We keep our inner product from the previous example (so $t_0 = 0$, $t_1 = 1/2$, and $t_2 = 1$). Let $W = \text{Span}\{q\}$. We will compute $\text{proj}_W p$:

$$\hat{p} = \text{proj}_W p = \frac{\langle p, q \rangle}{\langle q, q \rangle} q = \frac{12}{2}(2t - 1) = 12t - 6.$$

The geometric interpretation of $\hat{p}$ is that it is the polynomial in $\mathcal{P}_2$ closest to q when measured only at the given points t_0, t_1, t_2.

Theorem 7 (The Cauchy-Schwarz Inequality)

For all $\mathbf{u}, \mathbf{v} \in V$,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| .$$

Proof.

If $\mathbf{u} = \mathbf{0}$ then both sides are zero and the inequality is true.

Suppose $\mathbf{u} \neq \mathbf{0}$ and let W be the subspace spanned by $\mathbf{u}$. Since $\|c\mathbf{u}\| = |c| \, \|\mathbf{u}\|$, then

$$\|\text{proj}_W \, \mathbf{v}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{|\langle \mathbf{u}, \mathbf{u} \rangle|} \|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|^2} \|\mathbf{u}\| = \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{\|\mathbf{u}\|}.$$

Since $\|\text{proj}_W \, \mathbf{v}\| \leq \|\mathbf{v}\|$ (Orthogonal Decomposition Theorem), then $\frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{\|\mathbf{u}\|} \leq \|\mathbf{v}\|$. $\square$

One major application of Cauchy-Schwarz is to the Triangle Inequality. For the inner product on $\mathbb{R}^n$ we showed this using direct computation.

Theorem 8 (Triangle Inequality)

For all $\mathbf{u}, \mathbf{v} \in V$,

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Proof.

We have

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + 2|\langle \mathbf{u}, \mathbf{v} \rangle| + \langle \mathbf{v}, \mathbf{v} \rangle \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + 2\,\|\mathbf{u}\|\,\|\mathbf{v}\| + \langle \mathbf{v}, \mathbf{v} \rangle \quad \text{(Cauchy-Schwarz)} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2.
\end{aligned}$$

Taking square roots gives the result. $\square$

Example 9

Let $[a,b] \subset \mathbb{R}$ and let $V = \mathcal{C}([a,b])$. For $f, g \in V$, set

$$\langle f, g \rangle = \int_a^b f(t)g(t)\ dt.$$

We will show that this defines an inner product on V.

Since $f(t)g(t) = g(t)f(t)$ for all t, then (1) follows easily. If $h \in V$, then

$$\begin{aligned} \langle f+g, h \rangle &= \int_a^b (f+g)(t)h(t)\ dt = \int_a^b (f(t)+g(t))h(t)\ dt \\ &= \int_a^b f(t)h(t) + g(t)h(t)\ dt \\ &= \int_a^b f(t)h(t)\ dt + \int_a^b g(t)h(t)\ dt = \langle f, h \rangle + \langle g, h \rangle, \end{aligned}$$

so (2) is satisfied.

Example 9

Let $[a,b] \subset \mathbb{R}$ and let $V = \mathcal{C}([a,b])$. For $f, g \in V$, set

$$\langle f, g \rangle = \int_a^b f(t)g(t)\ dt.$$

We will show that this defines an inner product on V.

If $c \in \mathbb{R}$, then

$$\langle cf, g \rangle = \int_a^b (cf)(t)g(t)\ dt = \int_a^b c(f(t)g(t))\ dt = c\int_a^b f(t)g(t)\ dt = c\langle f, g \rangle,$$

so (3) is satisfied. Note that

$$\langle f, f \rangle = \int_a^b [f(t)]^2\ dt \geq 0.$$

If f is the zero function, then clearly the integral is zero. On the other hand, $[f(t)]^2$ is continuous and nonnegative. Hence, if integral is zero then f is the zero function.

This inner product is essential to *Fourier analysis*.

Example 10

Let $V = \mathcal{C}([0, 2\pi])$. For all positive integers m and n, $\sin mt, \cos nt \in V$. If $m \neq n$, then

$$\langle \sin mt, \cos nt \rangle = \int_0^{2\pi} (\sin mt)(\cos nt)\ dt = \frac{1}{2} \int_0^{2\pi} [\sin(mt + nt) + \sin(mt - nt)]\ dt$$

$$= \frac{1}{2} \left[-\frac{\cos(mt + nt)}{m + n} - \frac{\cos(mt - nt)}{m - n} \right]_0^{2\pi} = 0.$$

If $m = n$, then

$$\langle \sin nt, \cos nt \rangle = \int_0^{2\pi} (\sin nt)(\cos nt)\ dt = \frac{1}{2} \int_0^{2\pi} [\sin(mt + nt)]\ dt$$

$$= \frac{1}{2} \left[-\frac{\cos(mt + nt)}{m + n} \right]_0^{2\pi} = 0.$$

A *Fourier series* is an approximation of a continuous function by a linear combination of sine and cosine functions.

Any function in $V = \mathcal{C}([0, 2\pi])$ can be approximated by a function of the form

$$\frac{a_0}{2} + a_1 \cos t + \cdots + a_n \cos nt + b_1 \sin t + \cdots + b_n \sin nt$$

as closely as desired by taking sufficiently large n. Note that the set

$$\{1, \cos t, \cos 2t, \ldots, \cos nt, \sin t, \sin 2t, \ldots, \sin nt\}$$

is orthogonal in V with the above inner product. Above we verified that $\sin mt$ and $\cos nt$ are orthogonal for all m, n. In your book it is verified that $\cos mt$ and $\cos nt$ are orthogonal for $m \neq n$. Part of your homework is to do the same verification for $\sin mt$ and $\sin nt$.

Let V and let W be the subspace of V spanned by the (orthogonal) set

$$\{1, \cos t, \cos 2t, \ldots, \cos nt, \sin t, \sin 2t, \ldots, \sin nt\}$$

An *nth-order Fourier approximation* of a continuous function $f \in V$ is the best approximation to f in W. That is, we compute the Fourier approximation as $\text{proj}_W f$.

Suppose we write

$$f \approx \frac{a_0}{2} + a_1 \cos t + \cdots + a_n \cos nt + b_1 \sin t + \cdots + b_n \sin nt.$$

The coefficients are the *Fourier coefficients* of f and the standard formula for orthogonal projections gives

$$a_k = \frac{\langle f, \cos kt \rangle}{\langle \cos kt, \cos kt \rangle}, \quad b_k = \frac{\langle f, \sin kt \rangle}{\langle \sin kt, \sin kt \rangle},$$

for $k \geq 1$. Since $\langle \cos kt, \cos kt \rangle = \langle \sin kt, \sin kt \rangle = \pi$, then

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos kt \ dt, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin kt \ dt.$$

Since $\langle 1, 1 \rangle = 2\pi$, then

$$\frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_0^{2\pi} f(t) \cdot 1 \ dt = \frac{1}{2} \left[\frac{1}{\pi} \int_0^{2\pi} f(t) \cos(0t) \ dt \right] = \frac{a_0}{2}.$$

Example 11

We will compute the third-order Fourier approximation to $f(t) = t - 1$.

We have

$$\frac{a_0}{2} = \frac{1}{2} \cdot \frac{1}{\pi} \int_0^{2\pi} (t-1)\ dt = \frac{1}{2\pi} \left[\frac{1}{2} t^2 - t \right]_0^{2\pi} = \frac{1}{2\pi} \left[\frac{1}{2} (4\pi)^2 - 2\pi \right]_0^{2\pi} = \pi - 1.$$

For $k > 0$, using integration-by-parts,

$$a_k = \frac{1}{\pi} \int_0^{2\pi} (t-1) \cos kt\ dt = \frac{1}{\pi} \left[\frac{t-1}{k} \sin kt + \frac{1}{k^2} \cos kt \right]_0^{2\pi} = 0,$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} (t-1) \sin kt\ dt = \frac{1}{\pi} \left[\frac{1}{k^2} \sin kt - \frac{t-1}{k} \cos kt \right]_0^{2\pi} = -\frac{2}{k}.$$

Thus, the third-order Fourier approximation to $f(t) = t - 1$ is

$$(\pi - 1) - 2 \sin t - \sin 2t - \frac{2}{3} \sin 3t.$$

Next time

In the next lecture we will:

- Begin Exam Review.