Math 1553
Introduction to Linear Algebra

School of Mathematics
Georgia Institute of Technology

Introduction to Linear Algebra

Motivation and Overview

Linear. Algebra.

What is Linear Algebra?

Linear

Algebra

- from al-jebr (Arabic), meaning reunion of broken parts
- 9^{th} century Abu Ja'far Muhammad ibn Muso al-Khwarizmi

Why a whole course?

But these are the easiest kind of equations! I learned how to solve them in 7th grade!

Ah, but engineers need to solve *lots* of equations in *lots* of variables.

$$\begin{array}{rcrcrcrcrcrcr}
3x_1 & + & 4x_2 & + & 10x_3 & + & 19x_4 & - & 2x_5 & - & 3x_6 & = & 141 \\
7x_1 & + & 2x_2 & - & 13x_3 & - & 7x_4 & + & 21x_5 & + & 8x_6 & = & 2567 \\
-x_1 & + & 9x_2 & + & \frac{3}{2}x_3 & + & x_4 & + & 14x_5 & + & 27x_6 & = & 26 \\
\frac{1}{2}x_1 & + & 4x_2 & + & 10x_3 & + & 11x_4 & + & 2x_5 & + & x_6 & = & -15
\end{array}$$

Often, it's enough to know some information about the set of solutions without having to solve the equations at all!

Also, what if one of the coefficients of the x_i is itself a parameter— like an unknown real number t?

In real life, the difficult part is often in recognizing that a problem can be solved using linear algebra in the first place: need *conceptual* understanding.

Linear Algebra in Engineering

Large classes of engineering problems, no matter how huge, can be reduced to linear algebra:

$$Ax = b \quad \text{or}$$
$$Ax = \lambda x$$

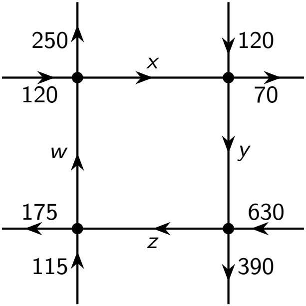
"...and now it's just linear algebra"

Applications of Linear Algebra

Civil Engineering: How much traffic flows through the four labeled segments?

⇝ system of linear equations:

Traffic flow (cars/hr)



Applications of Linear Algebra

Chemistry: Balancing reaction equations

$$\underline{\ x\ }\ C_2H_6 + \underline{\ y\ }\ O_2 \to \underline{\ z\ }\ CO_2 + \underline{\ w\ }\ H_2O$$

$\rightsquigarrow$ system of linear equations, one equation for each element.

Applications of Linear Algebra

Biology: In a population of rabbits. . .

- half of the new born rabbits survive their first year
- of those, half survive their second year
- the maximum life span is three years
- rabbits produce 0, 6, 8 rabbits in their first, second, and third years

If I know the population in 2016 (in terms of the number of first, second, and third year rabbits), then what is the population in 2017?

$\rightsquigarrow$ system of linear equations:

Question

Does the rabbit population have an asymptotic behavior? Is this even a linear algebra question? Yes, it is!

Applications of Linear Algebra

Geometry and Astronomy: Find the equation of a circle passing through 3 given points, say $(1, 0)$, $(0, 1)$, and $(1, 1)$. The general form of a circle is $a(x^2 + y^2) + bx + cy + d = 0$.

$\rightsquigarrow$ system of linear equations:

Very similar to: compute the orbit of a planet:

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

Applications of Linear Algebra

Google: "The 25 billion dollar eigenvector." Each web page has some importance, which it shares via outgoing links to other pages
$\rightsquigarrow$ system of linear equations (in gazillions of variables).

Larry Page flies around in a private 747 because he paid attention in his linear algebra class!

Stay tuned!

Overview of the Course

- Solve the matrix equation $Ax = b$
  - Solve systems of linear equations using matrices, row reduction, and inverses.
  - Solve systems of linear equations with varying parameters using parametric forms for solutions, the geometry of linear transformations, the characterizations of invertible matrices, and determinants.

- Solve the matrix equation $Ax = \lambda x$
  - Solve eigenvalue problems through the use of the characteristic polynomial.
  - Understand the dynamics of a linear transformation via the computation of eigenvalues, eigenvectors, and diagonalization.

- Almost solve the equation $Ax = b$
  - Find best-fit solutions to systems of linear equations that have no actual solution using least squares approximations.

What to Expect This Semester

Your previous math courses probably focused on how to do (sometimes rather involved) computations.

This is important, **but** Wolfram Alpha can do all these problems better than any of us can. Nobody is going to hire you to do something a computer can do better.

> If a computer can do the problem better than you can, then it's just an algorithm: this is not real problem solving.

So what are we going to do?

- About half the material focuses on how to do linear algebra computations—that is still important.
- The other half is on *conceptual* understanding of linear algebra. This is much more subtle: it's about figuring out *what question* to ask the computer, or whether you actually need to do any computations at all.

Resources

Everything is on the course web page.

Including these slides. There's a link from T-Square.

On the webpage you'll find:

- Course administration: the names of your TAs, their office hours, your recitation location, etc.
- Course organization: grading policies, details about homework and exams, etc.
- Help and advice: how to succeed in this course, resources available to you.
- Calendar: what will happen on which day, links to daily slides, quizzes, practice exams, solutions, etc.

T-Square: your grades, link to WeBWorK.

Piazza: this is where to ask questions, and where I'll post announcements.

Chapter 1

Linear Equations

Section 1.1

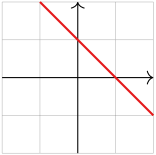
Systems of Linear Equations

One Linear Equation

What does the solution set of a linear equation look like?

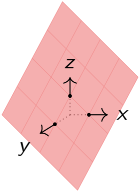
- $x + y = 1$

  $\rightsquigarrow$ a line in the plane: $y = 1 - x$



- $x + y + z = 1$

  $\rightsquigarrow$ a plane in space: $z = 1 - x - y$



- $x + y + z + w = 1$

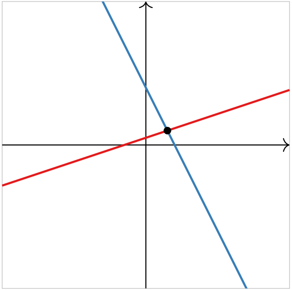
  $\rightsquigarrow$ a “3-plane” in “4-space”. . .

[not pictured here]

Systems of Linear Equations

What does the solution set of a *system* of more than one linear equation look like?

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



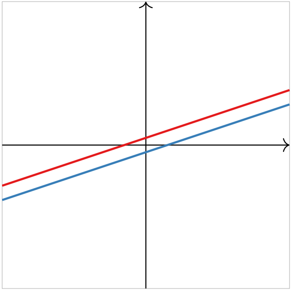
In general it's an intersection of lines, planes, etc.

Kinds of Solution Sets

In what other ways can two lines intersect?

$x - 3y = -3$

$x - 3y = 3$

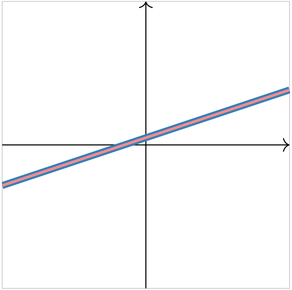


A system of equations with no solutions is called **inconsistent**.

Kinds of Solution Sets

In what other ways can two lines intersect?

$$\begin{aligned} x - 3y &= -3 \\ 2x - 6y &= -6 \end{aligned}$$



Note that multiplying an equation by a nonzero number gives the *same solution set*. In other words, they are *equivalent* (systems of) equations.

Poll

What about in three variables?

Solving Systems of Equations

Example

Solve the system of equations

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned}$$

This is the kind of problem we'll talk about for the first half of the course.

- A **solution** is a list of numbers $x, y, z, \ldots$ that make *all* of the equations true.
- The **solution set** is the collection of all solutions.
- **Solving** the system means finding the solution set.

What is a *systematic* way to solve a system of equations?

Solving Systems of Equations

Example

Solve the system of equations

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned}$$

What strategies do you know?

Solving Systems of Equations

Example

Solve the system of equations

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned}$$

Elimination method: in what ways can you manipulate the equations?

Solving Systems of Equations

Example

Solve the system of equations

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned}$$

Now I've eliminated x from the last equation!

. . . but there's a long way to go still. Can we make our lives easier?

Solving Systems of Equations

Better notation

It sure is a pain to have to write x, y, z, and $=$ over and over again.

Matrix notation: write just the numbers, in a box, instead!

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned} \quad \xrightarrow{\text{becomes}} \quad \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right)$$

This is called an **(augmented) matrix**. Our equation manipulations become **elementary row operations**:

- Multiply all entries in a row by a nonzero number. **(scale)**
- Add a multiple of each entry of one row to the corresponding entry in another. **(row replacement)**
- Swap two rows. **(swap)**

Row Operations

Example

Solve the system of equations

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned}$$

Start:

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right)$$

Goal: we want our elimination method to eventually produce a system of equations like

$$\begin{aligned} x \quad\quad\quad &= A \\ y \quad\quad &= B \\ z &= C \end{aligned}$$

or in matrix form,

So we need to do row operations that make the start matrix look like the end one.

Strategy: fiddle with it so we only have ones and zeros.

Row Operations

Continued

$$\left(\begin{array}{rrr|r} 1 & 2 & 3 & 6 \\ \boxed{2} & -3 & 2 & 14 \\ \boxed{3} & 1 & -1 & -2 \end{array}\right)$$

We want these to be zero.
So we subract multiples of the first row.

$$\left(\begin{array}{rrr|r} 1 & \boxed{2} & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 0 & \boxed{-5} & -10 & -20 \end{array}\right)$$

We want these to be zero.

It would be nice if this were a 1.
We could divide by -7, but that
would produce ugly fractions.

Let's swap the last two rows first.

Row Operations

Continued

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{array}\right)$$

We want these to be zero.

Let's make this a 1 first.

Success!

Check:

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x - 3y + 2z &= 14 \\ 3x + y - z &= -2 \end{aligned} \quad \xrightarrow{\text{substitute solution}}$$

Row Equivalence

Important

The process of doing row operations to a matrix does not change the solution set of the corresponding linear equations!

Definition

Two matrices are called **row equivalent** if one can be obtained from the other by doing some number of elementary row operations.

So the linear equations of row-equivalent matrices have the *same solution set*.

A Bad Example

Example

Solve the system of equations

$$\begin{aligned} x + y &= 2 \\ 3x + 4y &= 5 \\ 4x + 5y &= 9 \end{aligned}$$

Let's try doing row operations:

First clear these by subtracting multiples of the first row.

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 3 & 4 & 5 \\ 4 & 5 & 9 \end{array}\right)$$

Now clear this by subtracting the second row.

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

A Bad Example

Continued

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{array}\right) \quad \xrightarrow{\text{translates into}}$$

In other words, the original equations

$$\begin{aligned} x + y &= 2 \\ 3x + 4y &= 5 \\ 4x + 5y &= 9 \end{aligned} \qquad \text{have the same solutions as} \qquad \begin{aligned} x + y &= 2 \\ y &= -1 \\ 0 &= 2 \end{aligned}$$

But the latter system obviously has no solutions (there is no way to make them all true), so our original system has no solutions either.

Definition

A system of equations is called **inconsistent** if it has no solution. It is **consistent** otherwise.

Section 1.2

Row Reduction and Echelon Forms

Row Echelon Form

Let's come up with an *algorithm* for turning an arbitrary matrix into a "solved" matrix. What do we mean by "solved"?

A matrix is in **row echelon form** if

1. All zero rows are at the bottom.
2. Each leading nonzero entry of a row is to the *right* of the leading entry of the row above.
3. Below a leading entry of a row, all entries are *zero*.

Picture:

$$\begin{pmatrix} \boxed{\star} & \star & \star & \star & \star \\ 0 & \boxed{\star} & \star & \star & \star \\ 0 & 0 & 0 & \boxed{\star} & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\star =$ any number

$\boxed{\star} =$ any nonzero number

Definition

A **pivot** $\boxed{\star}$ is the first nonzero entry of a row of a matrix in row echelon form.

Reduced Row Echelon Form

A matrix is in **reduced row echelon form** if it is in row echelon form, and in addition,

4. The pivot in each nonzero row is equal to 1.
5. Each pivot is the only nonzero entry in its column.

Picture:

$$\begin{pmatrix} \color{red}{1} & 0 & \star & 0 & \star \\ 0 & \color{red}{1} & \star & 0 & \star \\ 0 & 0 & 0 & \color{red}{1} & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\star$ = any number

$\color{red}{1}$ = pivot

Note: Echelon forms do not care whether or not a column is augmented. Just ignore the vertical line.

Question

Can every matrix be put into reduced row echelon form only using row operations?

Answer: Yes! Stay tuned.

Reduced Row Echelon Form

Continued

Why is this the “solved” version of the matrix?

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

is in reduced row echelon form. It translates into

which is clearly the solution.

But what happens if there are fewer pivots than rows? . . . parametrized solution set (later).

Poll

Reduced Row Echelon Form

Theorem

Every matrix is row equivalent to one and only one matrix in reduced row echelon form.

We'll give an algorithm, called **row reduction**, which demonstrates that every matrix is row equivalent to *at least one* matrix in reduced row echelon form.

Note: Like echelon forms, the row reduction algorithm does not care if a column is augmented: ignore the vertical line when row reducing.

The uniqueness statement is interesting—it means that, nomatter *how* you row reduce, you *always* get the same matrix in reduced row echelon form. (Assuming you only do the three legal row operations.) (And you don't make any arithmetic errors.)

Maybe you can figure out why it's true!

Row Reduction Algorithm

Step 1a	Swap the 1st row with a lower one so a leftmost nonzero entry is in 1st row (if necessary).
Step 1b	Scale 1st row so that its leading entry is equal to 1.
Step 1c	Use row replacement so all entries above and below this 1 are 0.
Step 2a	Cover the first row, swap the 2nd row with a lower one so that the leftmost nonzero (uncovered) entry is in 2nd row; uncover 1st row.
Step 2b	Scale 2nd row so that its leading entry is equal to 1.
Step 2c	Use row replacement so all entries above and below this 1 are 0.
Step 3a	Cover the first two rows, swap the 3rd row with a lower one so that the leftmost nonzero (uncovered) entry is in 3rd row; uncover first two rows.
etc.	

Example

$$\left(\begin{array}{ccc|c} 0 & -7 & -4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & -1 & -2 \end{array}\right)$$

Row Reduction

Example

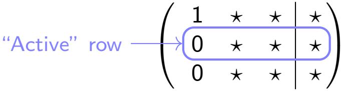
$$\left(\begin{array}{ccc|c} 0 & -7 & -4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & -1 & -2 \end{array}\right)$$

Row Reduction

Example, continued

$$\left(\begin{array}{rrr|r} 1 & 2 & 3 & 6 \\ 0 & -5 & -10 & -20 \\ 0 & -7 & -4 & 2 \end{array}\right)$$

Note: Step 2 never messes up the first (nonzero) column of the matrix, because it looks like this:



$$\text{“Active” row} \longrightarrow \left(\begin{array}{ccc|c} 1 & \star & \star & \star \\ 0 & \star & \star & \star \\ 0 & \star & \star & \star \end{array}\right)$$

Row Reduction

Example, continued

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{array}\right)$$

Note: Step 3 never messes up the columns to the left.

Success! The reduced row echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right) \implies \left\{\begin{array}{rcl} x & = & 1 \\ y & = & -2 \\ z & = & 3 \end{array}\right.$$

Step 4: profit?

Row Reduction

Another example

The linear system

$$\begin{aligned} 2x + 10y &= -1 \\ 3x + 15y &= 2 \end{aligned}$$

gives rise to the matrix

Let's row reduce it:

$$\left(\begin{array}{cc|c} 2 & 10 & -1 \\ 3 & 15 & 2 \end{array}\right)$$

The row reduced matrix

$$\left(\begin{array}{cc|c} 1 & 5 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

corresponds to the *inconsistent* system

$$\begin{aligned} x + 5y &= 0 \\ 0 &= 1. \end{aligned}$$

Inconsistent Matrices

Question

What does an augmented matrix in reduced row echelon form look like, if its system of linear equations is inconsistent?

Answer:

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

An augmented matrix corresponds to an inconsistent system of equations if and only if *the last* (i.e., the augmented) *column is a pivot column.*

Another Example

The linear system

$$\begin{aligned} 2x + y + 12z &= 1 \\ x + 2y + 9z &= -1 \end{aligned} \quad \text{gives rise to the matrix} \quad \left(\begin{array}{ccc|c} 2 & 1 & 12 & 1 \\ 1 & 2 & 9 & -1 \end{array}\right).$$

Let's row reduce it:

$$\left(\begin{array}{ccc|c} 2 & 1 & 12 & 1 \\ 1 & 2 & 9 & -1 \end{array}\right)$$

The row reduced matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 1 & 2 & -1 \end{array}\right) \quad \text{corresponds to the linear system} \quad \begin{cases} x + 5z = 1 \\ y + 2z = -1 \end{cases}$$

Another Example

Continued

The system

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

comes from a matrix in reduced row echelon form. Are we done? Is the system solved?

Yes! Rewrite:

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

For any value of z, there is exactly one value of x and y that makes the equations true. But z can be *anything we want*!

So we have found the solution set: it is all values x, y, z where

$$\begin{aligned} x &= 1 - 5z \\ y &= -1 - 2z \qquad \text{for } z \text{ any real number.} \\ (z &= z) \end{aligned}$$

This is called the **parametric form** for the solution.

Free Variables

Definition

Consider a *consistent* linear system of equations in the variables $x_1, \ldots, x_n$. Let A be a row echelon form of the matrix for this system.

We say that x_i is a **free variable** if its corresponding column in A is *not* a pivot column.

Important

1. You can choose *any value* for the free variables in a (consistent) linear system.
2. Free variables come from *columns without pivots* in a matrix in row echelon form.

In the previous example, z was free because the reduced row echelon form matrix was

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 4 \\ 0 & 1 & 2 & -1 \end{array}\right).$$

In this matrix:

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

the free variables are x_2 and x_4. (What about the last column?)

One More Example

The reduced row echelon form of the matrix for a linear system in x_1, x_2, x_3, x_4 is

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

The free variables are x_2 and x_4: they are the ones whose columns are *not* pivot columns.

This translates into the system of equations

$$\begin{cases} x_1 + 3x_4 = 2 \\ x_3 + x_4 = -1 \end{cases} \quad \Longrightarrow \quad \boxed{\begin{aligned} x_1 &= 2 - 3x_4 \\ x_3 &= -1 - 4x_4 \end{aligned}}$$

What happened to x_2? What is it allowed to be? Anything! The general solution is

for any values of x_2 and x_4.

The boxed equation is called the **parametric form** of the general solution to the system of equations. It is obtained by moving all free variables to the right-hand side of the $=$.

Poll

Summary

There are *three possibilities* for the reduced row echelon form of the augmented matrix of a linear system.

1. The last column is a pivot column.
   In this case, the system is *inconsistent*. There are *zero* solutions, i.e. the solution set is *empty*. Picture:
   $$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$$
2. Every column except the last column is a pivot column.
   In this case, the system has a *unique solution*. Picture:
   $$\left(\begin{array}{ccc|c} 1 & 0 & 0 & \star \\ 0 & 1 & 0 & \star \\ 0 & 0 & 1 & \star \end{array}\right)$$
3. The last column is not a pivot column, and some other column isn't either.
   In this case, the system has *infinitely many* solutions, corresponding to the infinitely many possible values of the free variable(s). Picture:
   $$\left(\begin{array}{cccc|c} 1 & \star & 0 & \star & \star \\ 0 & 0 & 1 & \star & \star \end{array}\right)$$

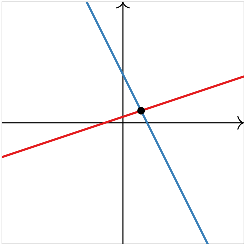
Section 1.3

Vector Equations

Motivation

We want to think about the *algebra* in linear algebra (systems of equations and their solution sets) in terms of *geometry* (points, lines, planes, etc).

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



This will give us better insight into the properties of systems of equations and their solution sets.

To do this, we need to introduce *n*-dimensional space $\mathbf{R}^n$, and **vectors** inside it.

Line, Plane, Space, . . .

Recall that **R** denotes the collection of all real numbers, i.e. the number line.

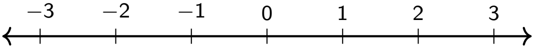
Definition

Let n be a positive whole number. We define

$$\mathbf{R}^n = \text{all ordered } n\text{-tuples of real numbers } (x_1, x_2, x_3, \ldots, x_n).$$

Example

When $n = 1$, we just get **R** back: $\mathbf{R}^1 = \mathbf{R}$. Geometrically, this is the *number line*.

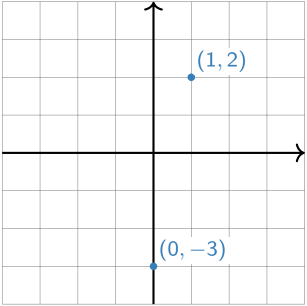


Line, Plane, Space, . . .

Continued

Example

When $n = 2$, we can think of $\mathbf{R}^2$ as the *plane*. This is because every point on the plane can be represented by an ordered pair of real numbers, namely, its x- and y-coordinates.



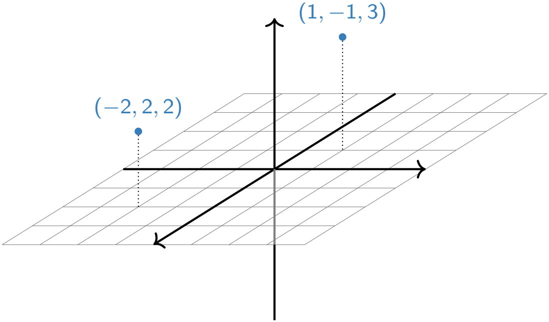
We can use the elements of $\mathbf{R}^2$ to *label* points on the plane, but $\mathbf{R}^2$ is not defined to be the plane!

Line, Plane, Space, . . .

Continued

Example

When $n = 3$, we can think of $\mathbf{R}^3$ as the *space* we (appear to) live in. This is because every point in space can be represented by an ordered triple of real numbers, namely, its x-, y-, and z-coordinates.



Again, we can use the elements of $\mathbf{R}^3$ to *label* points in space, but $\mathbf{R}^3$ is not defined to be space!

Line, Plane, Space, . . .

Continued

So what is $\mathbf{R}^4$? or $\mathbf{R}^5$? or $\mathbf{R}^n$?

. . . go back to the *definition*: ordered n-tuples of real numbers

$$(x_1, x_2, x_3, \dots, x_n).$$

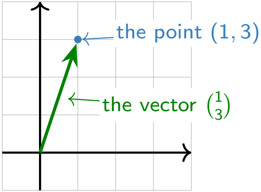
They're still "geometric" spaces, in the sense that our intuition for $\mathbf{R}^2$ and $\mathbf{R}^3$ sometimes extends to $\mathbf{R}^n$, but they're harder to visualize.

We'll make definitions and state theorems that apply to any $\mathbf{R}^n$, but we'll only draw pictures for $\mathbf{R}^2$ and $\mathbf{R}^3$.

Vectors

In the previous slides, we were thinking of elements of $\mathbf{R}^n$ as **points**: in the line, plane, space, etc.

We can also think of them as **vectors**: arrows with a given length and direction.



So the vector points *horizontally* in the amount of its x-coordinate, and *vertically* in the amount of its y-coordinate.

When we think of an element of $\mathbf{R}^n$ as a vector, we write it as a matrix with n rows and one column:

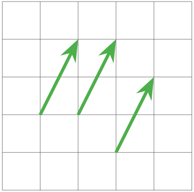
$$v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

We'll see why this is useful later.

Points and Vectors

So what is the difference between a point and a vector?

A vector need not start at the origin: *it can be located anywhere*! In other words, an arrow is determined by its length and its direction, not by its location.



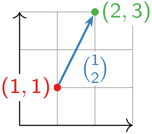
These arrows all represent the vector $\begin{pmatrix}1\\2\end{pmatrix}$.

However, unless otherwise specified, we'll assume a vector starts at the origin: we'll usually be sloppy and identify the vector $\binom{1}{2}$ with the point $(1, 2)$.

This makes sense in the real world: many physical quantities, such as velocity, are represented as vectors. But it makes more sense to think of the velocity of a car as being located at the car.

Another way to think about it: a vector is a *difference* between two points, or the arrow from one point to another.

For instance, $\begin{pmatrix}1\\2\end{pmatrix}$ is the arrow from $(1, 1)$ to $(2, 3)$.



Vector Algebra

Definition

- We can add two vectors together:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a+x \\ b+y \\ c+z \end{pmatrix}.$$

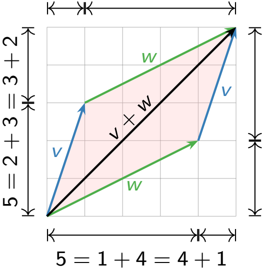
- We can multiply, or **scale**, a vector by a real number c:

$$c \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c \cdot x \\ c \cdot y \\ c \cdot z \end{pmatrix}.$$

  We call c a **scalar** to distinguish it from a vector. If v is a vector and c is a scalar, cv is called a **scalar multiple** of v.

(And likewise for vectors of length n.) For instance,

Vector Addition and Subtraction: Geometry

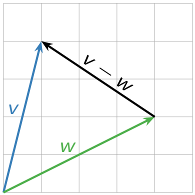


The parallelogram law for vector addition

Geometrically, the sum of two vectors v, w is obtained as follows: place the tail of w at the head of v. Then $v + w$ is the vector whose tail is the tail of v and whose head is the head of w. Doing this both ways creates a parallelogram. For example,

$$\begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}5\\5\end{pmatrix}.$$

Why? The width of $v + w$ is the sum of the widths, and likewise with the heights.



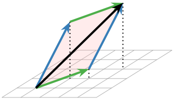
Vector subtraction

Geometrically, the difference of two vectors v, w is obtained as follows: place the tail of v and w at the same point. Then $v - w$ is the vector from the head of v to the head of w. For example,

$$\begin{pmatrix}1\\4\end{pmatrix} - \begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}-3\\2\end{pmatrix}.$$

Why? If you add $v - w$ to w, you get v.

This works in higher dimensions too!

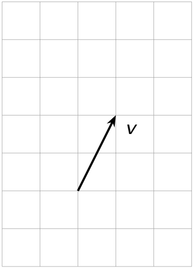


Scalar Multiplication: Geometry

Scalar multiples of a vector

These have the same *direction* but a different *length*.

Some multiples of v.



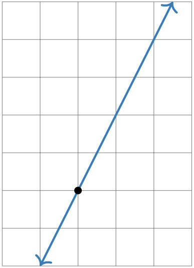
$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$2v = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$-\frac{1}{2}v = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$$

$$0v = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

All multiples of v.



So the scalar multiples of v form a *line*.

Linear Combinations

We can add and scalar multiply in the same equation:

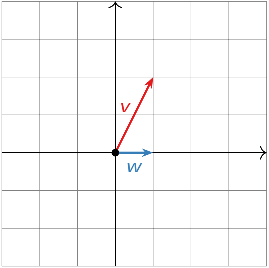
$$w = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p$$

where $c_1, c_2, \ldots, c_p$ are scalars, $v_1, v_2, \ldots, v_p$ are vectors in $\mathbf{R}^n$, and w is a vector in $\mathbf{R}^n$.

Definition

We call w a **linear combination** of the vectors $v_1, v_2, \ldots, v_p$. The scalars $c_1, c_2, \ldots, c_p$ are called the **weights** or **coefficients**.

Example



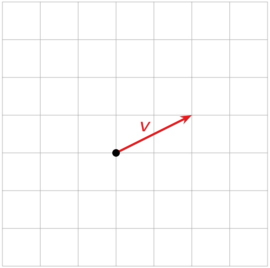
Let $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

What are some linear combinations of v and w?

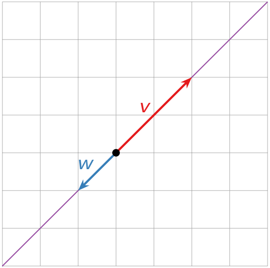
- $v + w$
- $v - w$
- $2v + 0w$
- $2w$
- $-v$

Poll

More Examples



What are some linear combinations of $v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$?



Question

What are all linear combinations of

$$v = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} -1 \\ -1 \end{pmatrix}?$$

Answer: The line which contains both vectors.

What's different about this example and the one on the poll?

Systems of Linear Equations

Question

Is $\begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$ a linear combination of $\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix}$?

Systems of Linear Equations

Continued

$$\begin{aligned} x - y &= 8 \\ 2x - 2y &= 16 \\ 6x - y &= 3 \end{aligned} \quad \xrightsquigarrow{\text{matrix form}} \quad \left(\begin{array}{cc|c} 1 & -1 & 8 \\ 2 & -2 & 16 \\ 6 & -1 & 3 \end{array}\right)$$

$$\xrightsquigarrow{\text{row reduce}} \quad \left(\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -9 \\ 0 & 0 & 0 \end{array}\right)$$

$$\xrightsquigarrow{\text{solution}} \quad \begin{aligned} x &= -1 \\ y &= -9 \end{aligned}$$

Conclusion:

$$-\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} - 9\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$$

What is the relationship between the original vectors and the matrix form of the linear equation? They have the same columns!

Shortcut: You can make the augmented matrix without writing down the system of linear equations first.

Vector Equations and Linear Equations

Summary

The vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = b,$$

where $v_1, v_2, \ldots, v_p, b$ are vectors in $\mathbf{R}^n$ and $x_1, x_2, \ldots, x_p$ are scalars, has the same solution set as the linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_p & b \\ | & | & & | & | \end{array}\right),$$

where the v_i's and b are the columns of the matrix.

So we now have (at least) *two* equivalent ways of thinking about linear systems of equations:

1. Augmented matrices.
2. Linear combinations of vectors (vector equations).

The last one is more geometric in nature.

Span

It is important to know what are *all* linear combinations of a set of vectors $v_1, v_2, \ldots, v_p$ in $\mathbf{R}^n$: it's exactly the collection of all b in $\mathbf{R}^n$ such that the vector equation (in the unknowns $x_1, x_2, \ldots, x_p$)

$$x_1v_1 + x_2v_2 + \cdots + x_pv_p = b$$

has a solution (i.e., is consistent).

Definition

Let $v_1, v_2, \ldots, v_p$ be vectors in $\mathbf{R}^n$. The **span** of $v_1, v_2, \ldots, v_p$ is the collection of all linear combinations of $v_1, v_2, \ldots, v_p$, and is denoted $\mathrm{Span}\{v_1, v_2, \ldots, v_p\}$. In symbols:

$$\boxed{\mathrm{Span}\{v_1, v_2, \ldots, v_p\} = \overset{\text{“the set of”}}{\Big\{}\, x_1v_1 + x_2v_2 + \cdots + x_pv_p \;\overset{\text{“such that”}}{\Big|}\; x_1, x_2, \ldots, x_p \text{ in } \mathbf{R} \,\Big\}}.$$

Synonyms: $\mathrm{Span}\{v_1, v_2, \ldots, v_p\}$ is the subset **spanned by** or **generated by** $v_1, v_2, \ldots, v_p$.

This is the first of several definitions in this class that you simply **must learn**. I will give you other ways to think about Span, and ways to draw pictures, but *this is the definition*. Having a vague idea what Span means will not help you solve any exam problems!

Span

Continued

Now we have several equivalent ways of making the same statement:

1. A vector b is in the span of $v_1, v_2, \ldots, v_p$.
2. The linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_p & b \\ | & | & & | & | \end{array}\right)$$

   is consistent.

3. The vector equation

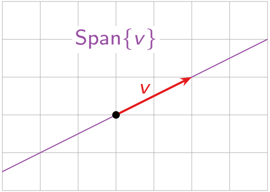
$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = b$$

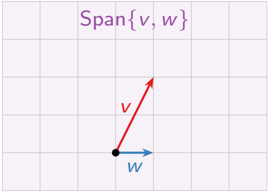
   has a solution.

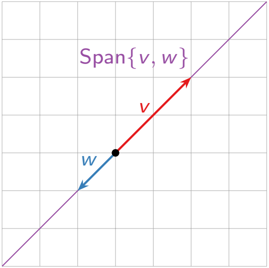
Note: **equivalent** means that, for any given list of vectors $v_1, v_2, \ldots, v_p, b$, *either* all three statements are true, *or* all three statements are false.

Pictures of Span

Drawing a picture of Span$\{v_1, v_2, \ldots, v_p\}$ is the same as drawing a picture of all linear combinations of $v_1, v_2, \ldots, v_p$.

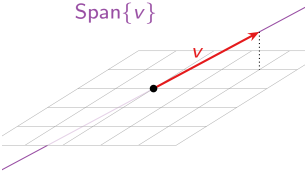




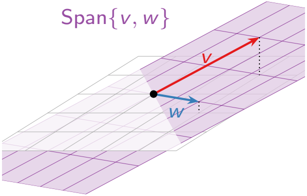


Pictures of Span

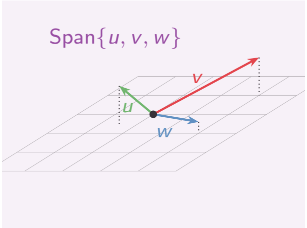
In $\mathbf{R}^3$



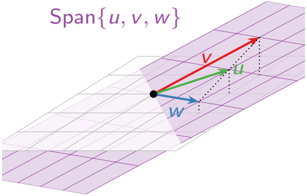
Span$\{v\}$



Span$\{v, w\}$



Span$\{u, v, w\}$



Span$\{u, v, w\}$

Poll

Section 1.4

The Matrix Equation $Ax = b$

Matrix × Vector

the first number is
the number of rows

the second number is
the number of columns

Let A be an $m \times n$ matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \qquad \text{with columns } v_1, v_2, \ldots, v_n$$

Definition

The **product** of A with a vector x in $\mathbf{R}^n$ is the linear combination

this means the equality
is a *definition*

$$Ax = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \boxed{\overset{\text{def}}{=}} x_1 v_1 + x_2 v_2 + \cdots + x_n v_n.$$

these must be equal

The output is a vector in $\mathbf{R}^m$.

Note that the number of columns of A has to equal the number of rows of x.

Example

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} =$$

Matrix Equations

An example

Question

Let v_1, v_2, v_3 be vectors in $\mathbf{R}^3$. How can you write the vector equation

$$2v_1 + 3v_2 - 4v_3 = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}$$

in terms of matrix multiplication?

Matrix Equations

In general

Let $v_1, v_2, \dots, v_n$, and b be vectors in $\mathbf{R}^m$. Consider the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b.$$

It is equivalent to the **matrix equation**

$$Ax = b$$

where

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Conversely, if A is any $m \times n$ matrix, then

$$Ax = b \qquad \text{is equivalent to the vector equation} \qquad x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b$$

where $v_1, \dots, v_n$ are the columns of A, and $x_1, \dots, x_n$ are the entries of x.

Linear Systems, Vector Equations, Matrix Equations, . . .

We now have *four* equivalent ways of writing (and thinking about) linear systems:

1. As a system of equations:

$$\begin{aligned} 2x_1 + 3x_2 &= 7 \\ x_1 - x_2 &= 5 \end{aligned}$$

2. As an augmented matrix:

$$\left(\begin{array}{cc|c} 2 & 3 & 7 \\ 1 & -1 & 5 \end{array}\right)$$

3. As a vector equation ($x_1 v_1 + \cdots + x_n v_n = b$):

$$x_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

4. As a matrix equation ($Ax = b$):

$$\begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

We will move back and forth freely between these over and over again, for the rest of the semester. Get comfortable with them now!

In particular, *all four have the same solution set.*

Matrix × Vector

Another way

Definition

A **row vector** is a matrix with one row. The product of a row vector of length n and a (column) vector of length n is

$$\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \overset{\text{def}}{=} a_1x_1 + \cdots + a_nx_n.$$

This is a scalar.

If A is an $m \times n$ matrix with rows $r_1, r_2, \ldots, r_m$, and x is a vector in $\mathbf{R}^n$, then

$$Ax = \begin{pmatrix} — r_1 — \\ — r_2 — \\ \vdots \\ — r_m — \end{pmatrix} x = \begin{pmatrix} r_1x \\ r_2x \\ \vdots \\ r_mx \end{pmatrix}$$

This is a vector in $\mathbf{R}^m$ (again).

Matrix × Vector

Both ways

Example

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} =$$

Note this is the same as before:

Now you have *two* ways of computing Ax.

In the second, you calculate Ax one entry at a time.

The second way is usually the most convenient, but we'll use both.

Spans and Solutions to Equations

Let A be a matrix with columns $v_1, v_2, \ldots, v_n$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

Very Important Fact That Will Appear on Every Midterm and the Final

$Ax = b$ has a solution

$\iff$ there exist $x_1, \ldots, x_n$ such that $A\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = b$

"if and only if"

$\iff$ there exist $x_1, \ldots, x_n$ such that $x_1 v_1 + \cdots + x_n v_n = b$

$\iff$ b is a linear combination of $v_1, \ldots, v_n$

$\iff$ b is in the span of the columns of A.

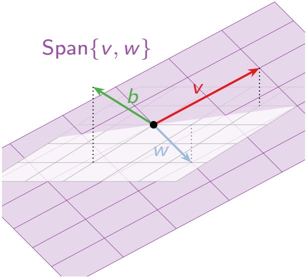
The last condition is geometric.

Spans and Solutions to Equations

Example

Question

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$ have a solution?



Columns of A:

$$v = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \qquad w = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Output vector:

$$b = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$$

Is b contained in the span of the columns of A? It sure doesn't look like it.

Conclusion: $Ax = b$ is *inconsistent*.

Spans and Solutions to Equations

Example, continued

Question

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$ have a solution?

Answer: Let's check by solving the matrix equation using row reduction.

In other words, the matrix equation

$$\begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$$

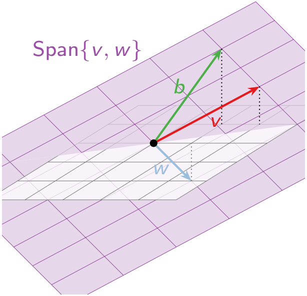
has no solution, as the picture shows.

Spans and Solutions to Equations

Example

Question

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ have a solution?



Columns of A:

$$v = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \qquad w = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Solution vector:

$$b = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

Is b contained in the span of the columns of A? It looks like it: in fact,

$$b = 1v + (-1)w \implies x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Spans and Solutions to Equations

Example, continued

Question

Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ have a solution?

Answer: Let's do this systematically using row reduction.

This is consistent with the picture on the previous slide:

$$1\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} - 1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \qquad \text{or} \qquad A\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Poll

When Solutions Always Exist

Here are criteria for a linear system to always have a solution.

Theorem

Let A be an $m \times n$ (non-augmented) matrix. The following are equivalent

1. $Ax = b$ has a solution for all b in $\mathbf{R}^m$.
2. The span of the columns of A is all of $\mathbf{R}^m$.
3. A has a pivot in each row.

recall that this means that for given A, either they're all true, or they're all false

Why is (1) the same as (2)? This was the $\boxed{\text{Very Important}}$ box from before.

Why is (1) the same as (3)? If A has a pivot in each row then its reduced row echelon form looks like this:

$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 1 & \star \end{pmatrix} \quad \text{and } (\,A \mid b\,) \text{ reduces to this:} \quad \left(\begin{array}{ccccc|c} 1 & 0 & \star & 0 & \star & \star \\ 0 & 1 & \star & 0 & \star & \star \\ 0 & 0 & 0 & 1 & \star & \star \end{array}\right).$$

There's no b that makes it inconsistent, so there's always a solution. If A doesn't have a pivot in each row, then its reduced form looks like this:

$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and this can be made inconsistent:} \quad \left(\begin{array}{ccccc|c} 1 & 0 & \star & 0 & \star & 0 \\ 0 & 1 & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{array}\right).$$

Properties of the Matrix–Vector Product

Let c be a scalar, u, v be vectors, and A a matrix.

- $A(u + v) = Au + Av$
- $A(cv) = cAv$

See Lay, §1.4, Theorem 5.

Consequence: If u and v are solutions to $Ax = 0$, then so is every vector in $\text{Span}\{u, v\}$. Why?

Important

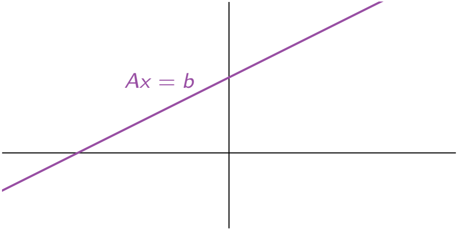
The set of solutions to $Ax = 0$ is a span.

Section 1.5

Solution Sets of Linear Systems

Plan For Today

Today we will learn to describe and draw the solution set of an arbitrary system of linear equations $Ax = b$, using spans.



Recall: the **solution set** is the collection of all vectors x such that $Ax = b$ is true.

Last time we discussed the set of vectors b for which $Ax = b$ has a solution.

We also described this set using spans, but it was a *different problem*.

Homogeneous Systems

Everything is easier when $b = 0$, so we start with this case.

Definition

A system of linear equations of the form $Ax = 0$ is called **homogeneous.**

These are linear equations where everything to the right of the $=$ is zero. The opposite is:

Definition

A system of linear equations of the form $Ax = b$ with $b \neq 0$ is called **nonhomogeneous** or **inhomogeneous.**

A homogeneous system always has the solution $x = 0$. This is called the **trivial solution.** The nonzero solutions are called **nontrivial.**

Observation

$Ax = 0$ has a nontrivial solution

$\iff$ there is a free variable

$\iff$ A has a column with no pivot.

Homogeneous Systems

Example

Question

What is the solution set of $Ax = 0$, where

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

We know how to do this: first form an augmented matrix and row reduce.

The only solution is the trivial solution $x = 0$.

Observation

Since the last column (everything to the right of the =) was zero to begin, it will always stay zero! So it's not really necessary to write augmented matrices in the homogeneous case.

Homogeneous Systems

Example

Question

What is the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}?$$

This last equation is called the **parametric vector form** of the solution.

It is obtained by listing equations for all the variables, in order, including the free ones, and making a vector equation.

Homogeneous Systems

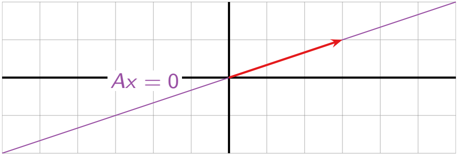
Example, continued

Question

What is the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}?$$

Answer: $x = x_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ for any x_2 in $\mathbf{R}$. The solution set is $\mathrm{Span}\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$.



Note: *one* free variable means the solution set is a *line* in $\mathbf{R}^2$ ($2 = \#$ variables $= \#$ columns).

Homogeneous Systems

Example

Question

What is the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & -1 & -5 \\ 1 & 0 & -2 \end{pmatrix}?$$

Homogeneous Systems

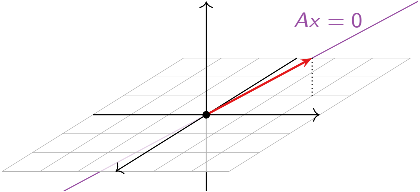
Example, continued

Question

What is the solution set of $Ax = 0$, where

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & -1 & -5 \\ 1 & 0 & -2 \end{pmatrix}?$$

Answer: $\operatorname{Span}\left\{ \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \right\}.$



Note: *one* free variable means the solution set is a *line* in $\mathbf{R}^3$ ($3 = \#$ variables $= \#$ columns).

Homogeneous Systems

Example

Question

What is the solution set of $Ax = 0$, where $A =$

$$\begin{pmatrix} 1 & 2 & 0 & -1 \\ -2 & -3 & 4 & 5 \\ 2 & 4 & 0 & -2 \end{pmatrix} \xrightarrow{\text{row reduce}} \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{\text{equations}} \left\{ \begin{array}{rrrrl} x_1 & & -\ 8x_3 & -\ 7x_4 & = 0 \\ & x_2 & +\ 4x_3 & +\ 3x_4 & = 0 \end{array} \right.$$

$$\xrightarrow{\text{parametric form}} \left\{ \begin{array}{rcrr} x_1 & = & 8x_3 & +\ 7x_4 \\ x_2 & = & -4x_3 & -\ 3x_4 \\ x_3 & = & x_3 & \\ x_4 & = & & x_4 \end{array} \right.$$

$$\xrightarrow{\text{parametric vector form}} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix}.$$

Homogeneous Systems

Example, continued

Question

What is the solution set of $Ax = 0$, where

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

Answer: $\text{Span}\left\{ \begin{pmatrix} 8 \\ -4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 \\ -3 \\ 0 \\ 1 \end{pmatrix} \right\}.$

[not pictured here]

Note: *two* free variables means the solution set is a *plane* in $\mathbf{R}^4$ ($4 = \#$ variables $= \#$ columns).

Parametric Vector Form

Homogeneous systems

Let A be an $m \times n$ matrix. Suppose that the free variables in the homogeneous equation $Ax = 0$ are $x_i, x_j, x_k, \ldots$

Then the solutions to $Ax = 0$ can be written in the form

$$x = x_i v_i + x_j v_j + x_k v_k + \cdots$$

for some vectors $v_i, v_j, v_k, \ldots$ in $\mathbf{R}^n$, and any scalars $x_i, x_j, x_k, \ldots$

The solution set is

$$\text{Span}\{v_i,\ v_j,\ v_k, \ldots\}.$$

The equation above is called the **parametric vector form** of the solution.

Poll

Nonhomogeneous Systems

Example

Question

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -3 \\ -6 \end{pmatrix}?$$

The only difference from the homogeneous case is the constant vector $p = \binom{-3}{0}$.

Note that p is itself a solution: take $x_2 = 0$.

Nonhomogeneous Systems

Example, continued

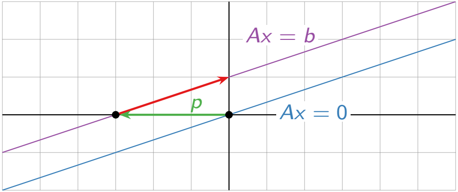
Question

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -3 \\ -6 \end{pmatrix}?$$

Answer: $x = x_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \end{pmatrix}$ for any x_2 in $\mathbf{R}$.

This is a *translate* of $\text{Span}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\}$: it is the parallel line through $p = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$.



It can be written

$$\text{Span}\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\} + \begin{pmatrix} -3 \\ 0 \end{pmatrix}.$$

Nonhomogeneous Systems

Example

Question

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & -1 & -5 \\ 1 & 0 & -2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -5 \\ -3 \\ -2 \end{pmatrix}?$$

Nonhomogeneous Systems

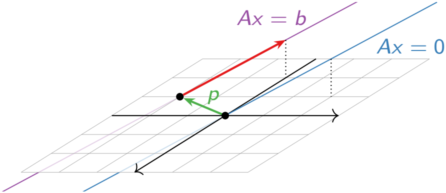
Example, continued

Question

What is the solution set of $Ax = b$, where

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & -1 & -5 \\ 1 & 0 & -2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} -5 \\ -3 \\ -2 \end{pmatrix}?$$

Answer: $\operatorname{Span}\left\{\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\right\} + \begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix}.$



The solution set is a *translate* of

$$\operatorname{Span}\left\{\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\right\}:$$

it is the parallel line through

$$p = \begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix}.$$

Homogeneous vs. Nonhomogeneous Systems

Key Observation

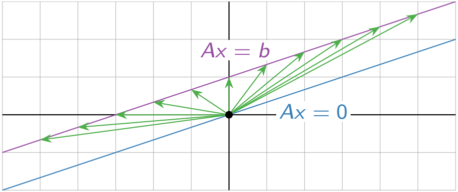
The set of solutions to $Ax = b$, if it is nonempty, is obtained by taking one **specific** or **particular solution** p to $Ax = b$, and adding all solutions to $Ax = 0$.

Why? If $Ap = b$ and $Ax = 0$, then

$$A(p + x) = Ap + Ax = b + 0 = b,$$

so $p + x$ is also a solution to $Ax = b$.

We know the solution set of $Ax = 0$ is a span. So the solution set of $Ax = b$ is a *translate* of a span: it is *parallel* to a span. (Or it is empty.)



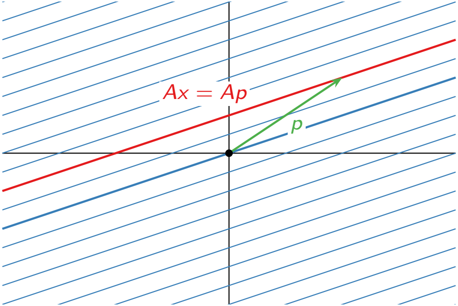
This works for *any* specific solution p: it doesn't have to be the one produced by finding the parametric vector form and setting the free variables all to zero, as we did before.

Homogeneous vs. Nonhomogeneous Systems

Varying b

If we understand the solution set of $Ax = 0$, then we understand the solution set of $Ax = b$ for all b: they are all translates (or empty).

For instance, if $A = \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix}$, then the solution sets for varying b look like this:



Which b gives the solution set $Ax = b$ in red in the picture?

Choose p on the red line, and set $b = Ap$. Then p is a specific solution to $Ax = b$, so the solution set of $Ax = b$ is the red line.

Note the cool optical illusion!

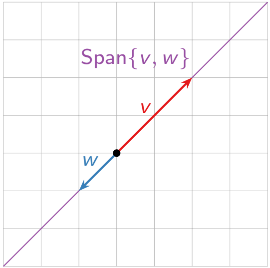
For a matrix equation $Ax = b$, you now know how to find which b's are possible, and what the solution set looks like for all b, both using spans.

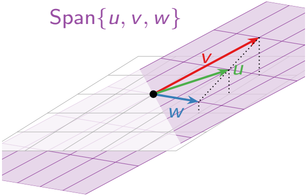
Section 1.7

Linear Independence

Motivation

Sometimes the span of a set of vectors is "smaller" than you expect from the number of vectors.





This can mean many things. For example, it can mean you're using too many vectors to write your solution set.

Notice in each case that one vector in the set is already in the span of the others—so it doesn't make the span bigger.

Today we will formalize this idea in the concept of *linear (in)dependence*.

Linear Independence

Definition

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ in $\mathbf{R}^n$ is **linearly independent** if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

has only the trivial solution $x_1 = x_2 = \cdots = x_p = 0$. The set $\{v_1, v_2, \ldots, v_p\}$ is **linearly dependent** otherwise.

In other words, $\{v_1, v_2, \ldots, v_p\}$ is linearly dependent if there exist numbers $x_1, x_2, \ldots, x_p$, not all equal to zero, such that

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0.$$

This is called a **linear dependence relation**.

Like span, linear (in)dependence is another one of those big vocabulary words that you absolutely need to learn. Much of the rest of the course will be built on these concepts, and you need to know exactly what they mean in order to be able to answer questions on quizzes and exams (and solve real-world problems later on).

Linear Independence

Definition

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ in $\mathbf{R}^n$ is **linearly independent** if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

has only the trivial solution $x_1 = x_2 = \cdots = x_p = 0$. The set $\{v_1, v_2, \ldots, v_p\}$ is **linearly dependent** otherwise.

Note that linear (in)dependence is a notion that applies to a *collection of vectors*, not to a single vector, or to one vector in the presence of some others.

Checking Linear Independence

Question: Is $\left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\2\end{pmatrix}, \begin{pmatrix}3\\1\\4\end{pmatrix} \right\}$ linearly independent?

Checking Linear Independence

Question: Is $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \right\}$ linearly independent?

Linear Independence and Matrix Columns

In general, $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

has only the trivial solution, if and only if the matrix equation

$$Ax = 0$$

has only the trivial solution, where A is the matrix with columns $v_1, v_2, \ldots, v_p$:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_p \\ | & | & & | \end{pmatrix}.$$

This is true if and only if the matrix A has a pivot in each column.

Important

- The vectors $v_1, v_2, \ldots, v_p$ are linearly independent if and only if the matrix with columns $v_1, v_2, \ldots, v_p$ has a pivot in each column.
- Solving the matrix equation $Ax = 0$ will either verify that the columns $v_1, v_2, \ldots, v_p$ of A are linearly independent, or will produce a linear dependence relation.

Linear Independence

Criterion

Suppose that one of the vectors $\{v_1, v_2, \ldots, v_p\}$ is a linear combination of the other ones (that is, it is in the span of the other ones):

$$v_3 = 2v_1 - \frac{1}{2}v_2 + 6v_4$$

Then the vectors are linearly *de*pendent:

Conversely, if the vectors are linearly dependent

$$2v_1 - \frac{1}{2}v_2 + 6v_4 = 0.$$

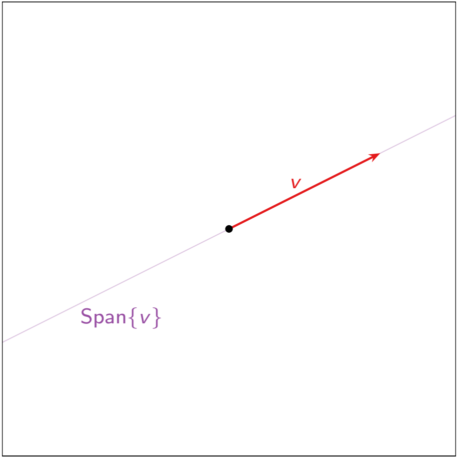
then one vector is a linear combination of (in the span of) the other ones:

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly *de*pendent if and only if one of the vectors is in the span of the other ones.

Linear Independence

Pictures in $\mathbf{R}^2$

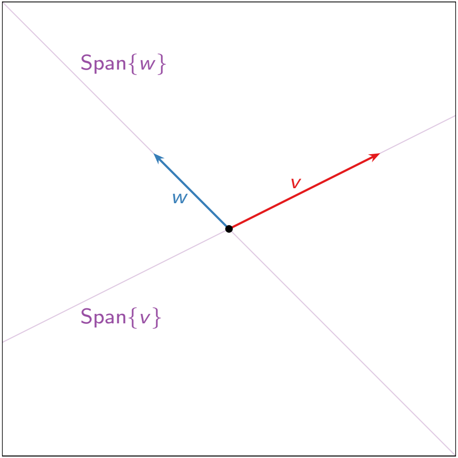


In this picture

One vector $\{v\}$:
Linearly independent if $v \neq 0$.

Linear Independence

Pictures in $\mathbf{R}^2$



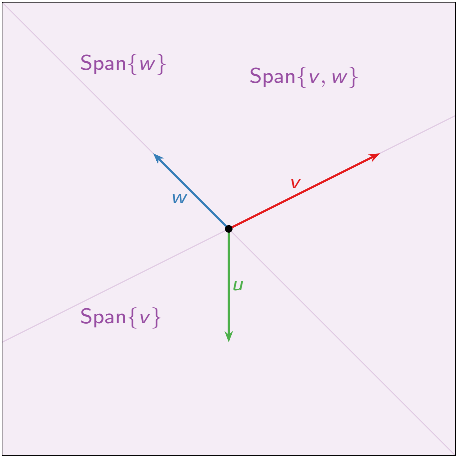
In this picture

One vector $\{v\}$:
Linearly independent if $v \neq 0$.

Two vectors $\{v, w\}$:
Linearly independent: neither is in the span of the other.

Linear Independence

Pictures in $\mathbf{R}^2$



In this picture

One vector $\{v\}$:
Linearly independent if $v \neq 0$.

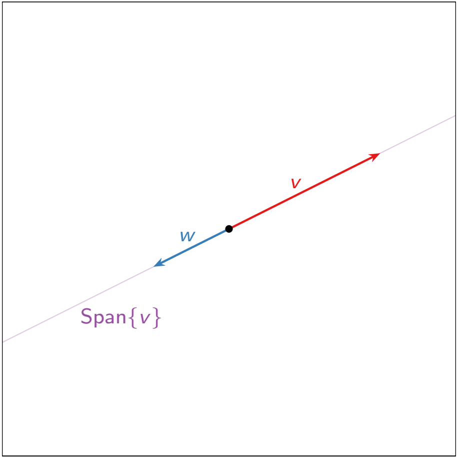
Two vectors $\{v, w\}$:
Linearly independent: neither is in the span of the other.

Three vectors $\{v, w, u\}$:
Linearly dependent: u is in Span$\{v, w\}$.

Also v is in Span$\{u, w\}$ and w is in Span$\{u, v\}$.

Linear Independence

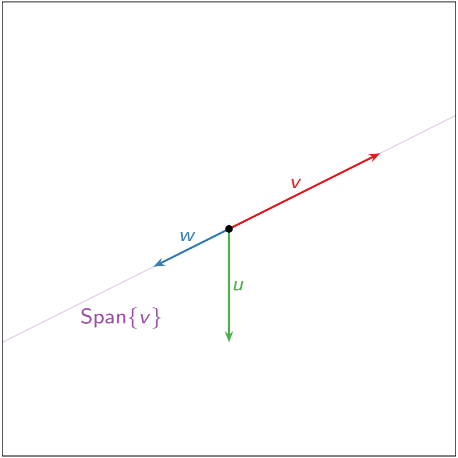
Pictures in $\mathbf{R}^2$



Two collinear vectors $\{v, w\}$:
Linearly dependent: w is in Span$\{v\}$ (and vice-versa).

Linear Independence

Pictures in $\mathbf{R}^2$



Two collinear vectors $\{v, w\}$:
Linearly dependent: w is in Span$\{v\}$ (and vice-versa).

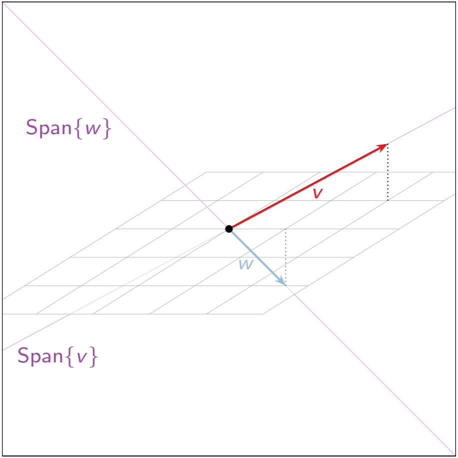
Observe: *Two* vectors are linearly *dependent* if and only if they are *collinear*.

Three vectors $\{v, w, u\}$:
Linearly dependent: w is in Span$\{v\}$ (and vice-versa).

Observe: If a set of vectors is linearly dependent, then so is any larger set of vectors!

Linear Independence

Pictures in $\mathbf{R}^3$

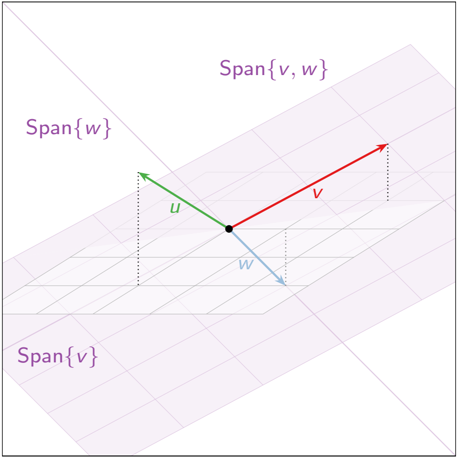


In this picture

Two vectors $\{v, w\}$:
Linearly independent: neither is in the span of the other.

Linear Independence

Pictures in $\mathbf{R}^3$



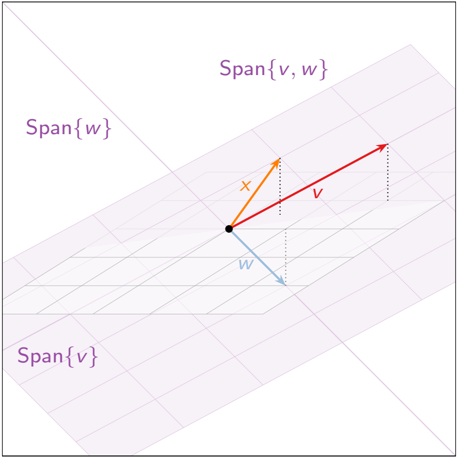
In this picture

Two vectors $\{v, w\}$:
Linearly independent: neither is in the span of the other.

Three vectors $\{v, w, u\}$:
Linearly independent: no one is in the span of the other two.

Linear Independence

Pictures in $\mathbf{R}^3$



In this picture

Two vectors $\{v, w\}$:
Linearly independent: neither is in the span of the other.

Three vectors $\{v, w, x\}$:
Linearly dependent: x is in $\mathrm{Span}\{v, w\}$.

Poll

Linear Independence

Stronger criterion

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly *de*pendent if and only if one of the vectors is in the span of the other ones.

Take the largest j such that v_j is in the span of the others. Then v_j is in the span of $v_1, v_2, \ldots, v_{j-1}$. Why? If not ($j = 3$):

$$v_3 = 2v_1 - \frac{1}{2}v_2 + 6v_4$$

Rearrange:

so v_4 works as well, but v_3 was supposed to be the last one that was in the span of the others.

Better Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly dependent if and only if there is some j such that v_j is in $\mathrm{Span}\{v_1, v_2, \ldots, v_{j-1}\}$.

Linear Independence

Increasing span criterion

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly dependent if and only if there is some j such that v_j is in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$.

Equivalently, $\{v_1, v_2, \ldots, v_p\}$ is linearly *in*dependent if for every j, the vector v_j is not in $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$.

This means $\text{Span}\{v_1, v_2, \ldots, v_j\}$ is *bigger* than $\text{Span}\{v_1, v_2, \ldots, v_{j-1}\}$.

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if, for every j, the span of $v_1, v_2, \ldots, v_j$ is strictly larger than the span of $v_1, v_2, \ldots, v_{j-1}$.

Translation

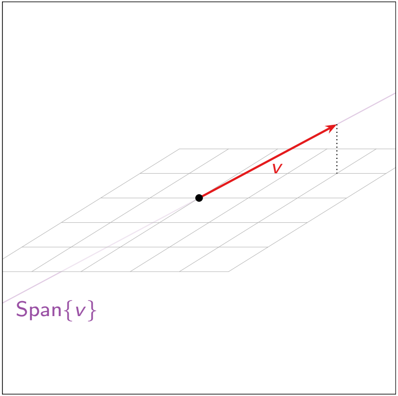
A set of vectors is linearly independent if and only if, every time you add another vector to the set, the span gets bigger.

Linear Independence

Increasing span criterion: pictures

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if, for every j, the span of $v_1, v_2, \ldots, v_j$ is strictly larger than the span of $v_1, v_2, \ldots, v_{j-1}$.



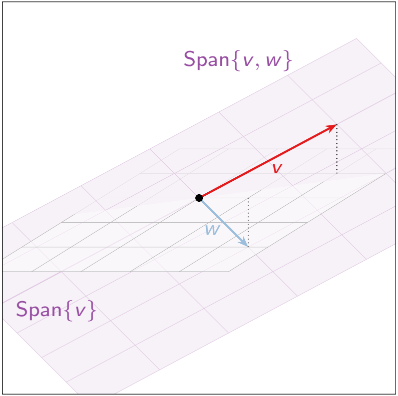
One vector $\{v\}$:
Linearly independent: span got bigger (than $\{0\}$).

Linear Independence

Increasing span criterion: pictures

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if, for every j, the span of $v_1, v_2, \ldots, v_j$ is strictly larger than the span of $v_1, v_2, \ldots, v_{j-1}$.



One vector $\{v\}$:

Linearly independent: span got bigger (than $\{0\}$).

Two vectors $\{v, w\}$:

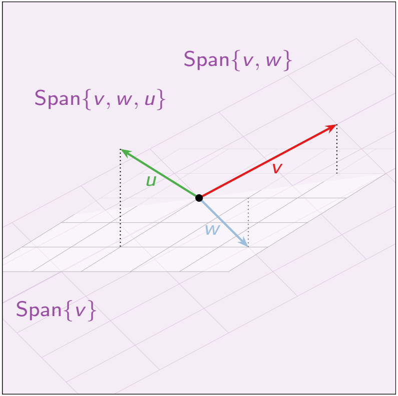
Linearly independent: span got bigger.

Linear Independence

Increasing span criterion: pictures

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if, for every j, the span of $v_1, v_2, \ldots, v_j$ is strictly larger than the span of $v_1, v_2, \ldots, v_{j-1}$.



One vector $\{v\}$:
Linearly independent: span got bigger (than $\{0\}$).

Two vectors $\{v, w\}$:
Linearly independent: span got bigger.

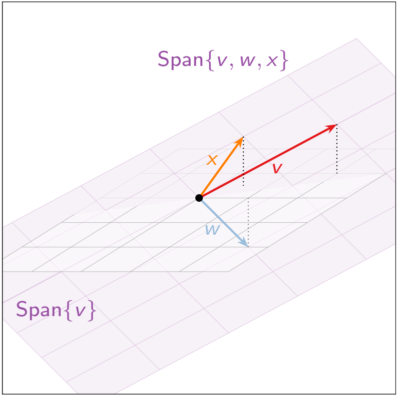
Three vectors $\{v, w, u\}$:
Linearly independent: span got bigger.

Linear Independence

Increasing span criterion: pictures

Theorem

A set of vectors $\{v_1, v_2, \ldots, v_p\}$ is linearly independent if and only if, for every j, the span of $v_1, v_2, \ldots, v_j$ is strictly larger than the span of $v_1, v_2, \ldots, v_{j-1}$.



One vector $\{v\}$:
Linearly independent: span got bigger (than $\{0\}$).

Two vectors $\{v, w\}$:
Linearly independent: span got bigger.

Three vectors $\{v, w, x\}$:
Linearly dependent: span didn't get bigger.

Linear Independence

Two more facts

Fact 1: Say $v_1, v_2, \ldots, v_n$ are in $\mathbf{R}^m$. If $n > m$ then $\{v_1, v_2, \ldots, v_n\}$ is linearly *de*pendent: the matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

cannot have a pivot in each column (it is too wide).

This says you can't have 4 linearly independent vectors in $\mathbf{R}^3$, for instance.

> A wide matrix can't have linearly independent columns.

Fact 2: If one of $v_1, v_2, \ldots, v_n$ is zero, then $\{v_1, v_2, \ldots, v_n\}$ is linearly *de*pendent. For instance, if $v_1 = 0$, then

$$1 \cdot v_1 + 0 \cdot v_2 + 0 \cdot v_3 + \cdots + 0 \cdot v_n = 0$$

is a linear dependence relation.

> A set containing the zero vector is linearly dependent.

Section 1.8

Introduction to Linear Transformations

Motivation

Let A be an $m \times n$ matrix. For the matrix equation $Ax = b$ we have learned to describe

- the solution set: all x in $\mathbf{R}^n$ making the equation true.
- the column span: the set of all b in $\mathbf{R}^m$ making the equation consistent.

It turns out these two sets are very closely related to each other.

In order to understand this relationship, it helps to think of the matrix A as a *transformation* from $\mathbf{R}^n$ to $\mathbf{R}^m$.

It's a special kind of transformation called a *linear transformation*.

This is also a way to understand the *geometry* of *matrices*.

Transformations

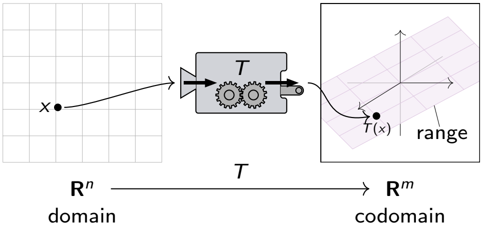
Definition

A **transformation** (or **function** or **map**) from $\mathbf{R}^n$ to $\mathbf{R}^m$ is a rule T that assigns to each vector x in $\mathbf{R}^n$ a vector $T(x)$ in $\mathbf{R}^m$.

- $\mathbf{R}^n$ is called the **domain** of T (the inputs).
- $\mathbf{R}^m$ is called the **codomain** of T (the outputs).
- For x in $\mathbf{R}^n$, the vector $T(x)$ in $\mathbf{R}^m$ is the **image** of x under T.
  Notation: $x \mapsto T(x)$.
- The set of all images $\{T(x) \mid x \text{ in } \mathbf{R}^n\}$ is the **range** of T.

Notation:

$$T\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m \quad \text{means} \quad T \text{ is a transformation from } \mathbf{R}^n \text{ to } \mathbf{R}^m.$$



It may help to think of T as a “machine” that takes x as an input, and gives you $T(x)$ as the output.

Functions from Calculus

Many of the functions you know and love have domain and codomain **R**.

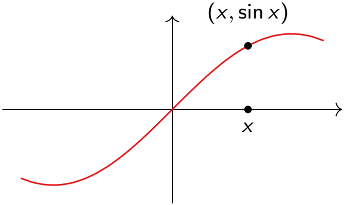
$$\sin\colon \mathbf{R} \longrightarrow \mathbf{R} \qquad \sin(x) = \left(\begin{array}{l}\text{the length of the opposite edge over the} \\ \text{hypotenuse of a right triangle with angle} \\ x \text{ in radians}\end{array}\right)$$

Note how I've written down the *rule* that defines the function sin.

$$f\colon \mathbf{R} \longrightarrow \mathbf{R} \qquad f(x) = x^2$$

Note that "x^2" is sloppy (but common) notation for a function: it doesn't have a name!

You may be used to thinking of a function in terms of its graph.



The horizontal axis is the domain, and the vertical axis is the codomain.

This is fine when the domain and codomain are **R**, but it's hard to do when they're $\mathbf{R}^2$ and $\mathbf{R}^3$! You need five dimensions to draw that graph.

Matrix Transformations

Most of the transformations we encounter in this class will come from (surprise) matrices!

Definition

Let A be an $m \times n$ matrix. The **matrix transformation** associated to A is the transformation

$$T\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m \qquad \text{defined by} \qquad T(x) = Ax.$$

In other words, T takes the vector x in $\mathbf{R}^n$ to the vector Ax in $\mathbf{R}^m$.

Your life will be much easier
if you just remember these.

- The *domain* of T is $\mathbf{R}^n$, which is the number of *columns* of A.
- The *codomain* of T is $\mathbf{R}^m$, which is the number of *rows* of A.
- The *range* of T is the set of all images of T:

$$T(x) = Ax = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n.$$

This is the *column span* of A. It is a span of vectors in the codomain.

Matrix Transformations

Example

Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and let $T(x) = Ax$, so $T\colon \mathbf{R}^2 \to \mathbf{R}^3$.

- If $u = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ then $T(u) =$
- Let $b = \begin{pmatrix} 7 \\ 5 \\ 7 \end{pmatrix}$. Find v in $\mathbf{R}^2$ such that $T(v) = b$. Is there more than one?

Matrix Transformations

Example, continued

Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and let $T(x) = Ax$, so $T\colon \mathbf{R}^2 \to \mathbf{R}^3$.

- Is there any c in $\mathbf{R}^3$ such that there is more than one v in $\mathbf{R}^2$ with $T(v) = c$?

  Translation: is there any c in $\mathbf{R}^3$ such that the solution set of $Ax = c$ has more than one vector v in it?

  The solution set of $Ax = c$ is a translate of the solution set of $Ax = b$ (from before), which has one vector in it. So the solution set to $Ax = c$ has only one vector. So no!

- Find c such that there is *no* v with $T(v) = c$.

  Translation: Find c such that $Ax = c$ is inconsistent.

  Translation: Find c not in the column span of A (i.e., the range of T).

  We could draw a picture, or notice: $a \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a+b \\ b \\ a+b \end{pmatrix}$. So anything in the column span has the same first and last coordinate. So $c = \left(\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right)$ is not in the column span (for example).

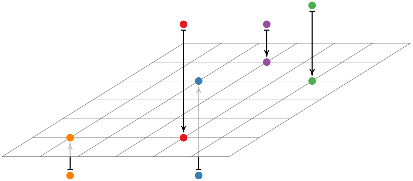
Matrix Transformations

Geometric example

Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and let $T(x) = Ax$, so $T\colon \mathbf{R}^3 \to \mathbf{R}^3$. Then

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

This is *projection onto the xy-axis*. Picture:



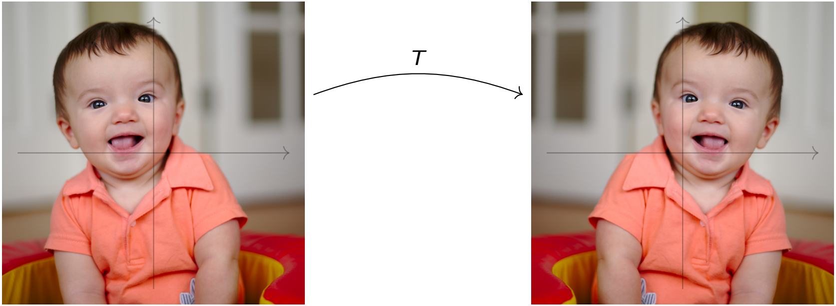
Matrix Transformations

Geometric example

Let $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ and let $T(x) = Ax$, so $T\colon \mathbf{R}^2 \to \mathbf{R}^2$. Then

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}.$$

This is *reflection over the y-axis*. Picture:



Poll

Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and let $T(x) = Ax$, so $T\colon \mathbf{R}^2 \to \mathbf{R}^2$. ($T$ is called a **shear**.)

Linear Transformations

Recall: If A is a matrix, u, v are vectors, and c is a scalar, then

$$A(u+v) = Au + Av \qquad A(cv) = cAv.$$

So if $T(x) = Ax$ is a matrix transformation then,

$$T(u+v) = T(u) + T(v) \qquad T(cv) = cT(v).$$

This property is so special that it has its own name.

Definition

A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is **linear** if it satisfies the above equations for all vectors u, v in $\mathbf{R}^n$ and all scalars c.

In other words, T "respects" addition and scalar multiplication.

Check: if T is linear, then

$$T(0) = 0 \qquad T(cu + dv) = cT(u) + dT(v)$$

for all vectors u, v and scalars c, d. More generally,

$$T\left(c_1 v_1 + c_2 v_2 + \cdots + c_n v_n\right) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n).$$

In engineering this is called **superposition**.

Linear Transformations

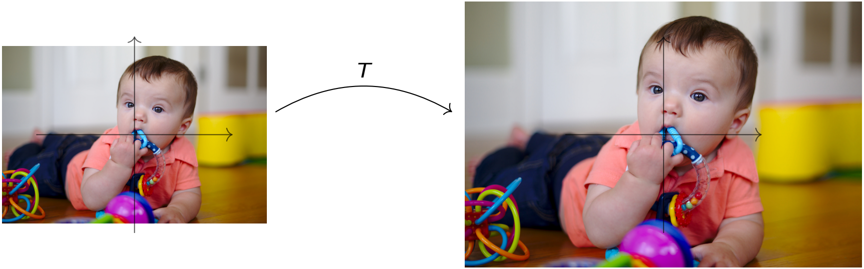
Dilation

Define $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ by $T(x) = 1.5x$. Is T linear? Check:

$$\begin{aligned} T(u+v) &= \\ T(cv) &= \end{aligned}$$

So T satisfies the two equations, hence T is linear.

This is called **dilation** or **scaling** (by a factor of 1.5). Picture:



Linear Transformations

Rotation

Define $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

Is T linear? Check:

$$T\left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right) =$$

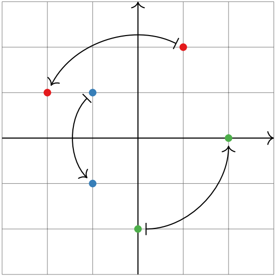
$$T\left(c\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right) =$$

So T satisfies the two equations, hence T is linear. This is called **rotation** (by 90°). Picture:

$$T\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$T\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$T\begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$



Section 1.9

The Matrix of a Linear Transformation

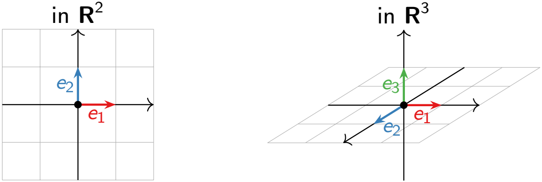
Unit Coordinate Vectors

Definition

The **unit coordinate vectors** in $\mathbf{R}^n$ are

This is what $e_1, e_2, \ldots$ mean, *for the rest of the class.*

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \ldots, \quad e_{n-1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$



Note: if A is an $m \times n$ matrix with columns $v_1, v_2, \ldots, v_n$, then $Ae_i = v_i$ for $i = 1, 2, \ldots, n$: multiplying a matrix by e_i gives you the ith column.

Linear Transformations are Matrix Transformations

Recall: A matrix A defines a linear transformation T by $T(x) = Ax$.

Theorem

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Let

$$A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix}.$$

This is an $m \times n$ matrix, and T is the matrix transformation for A: $T(x) = Ax$.

The matrix A is called the **standard matrix** for T.

Take-Away

Linear transformations are the same as matrix transformations.

Dictionary

Linear transformation $T\colon \mathbf{R}^n \to \mathbf{R}^m$ $\rightsquigarrow$ $m \times n$ matrix $A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix}$

$T(x) = Ax$, $T\colon \mathbf{R}^n \to \mathbf{R}^m$ $\leftsquigarrow$ $m \times n$ matrix A

Linear Transformations are Matrix Transformations

Continued

Why is a linear transformation a matrix transformation?

Suppose for simplicity that $T\colon \mathbf{R}^3 \to \mathbf{R}^2$.

Linear Transformations are Matrix Transformations

Example

Before, we defined a **dilation** transformation $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ by $T(x) = 1.5x$. What is its standard matrix?

Check:

$$\begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.5x \\ 1.5y \end{pmatrix} = 1.5 \begin{pmatrix} x \\ y \end{pmatrix} = T \begin{pmatrix} x \\ y \end{pmatrix}.$$

Linear Transformations are Matrix Transformations

Example

Question

What is the matrix for the linear transformation $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ defined by

$$T(x) = x \text{ rotated counterclockwise by an angle } \theta?$$

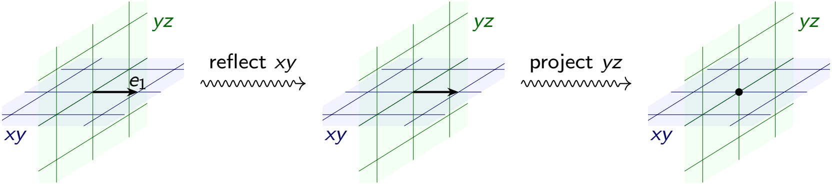
(Check linearity. . .)

Linear Transformations are Matrix Transformations

Example

Question

What is the matrix for the linear transformation $T\colon \mathbf{R}^3 \to \mathbf{R}^3$ that reflects through the *xy*-plane and then projects onto the *yz*-plane?



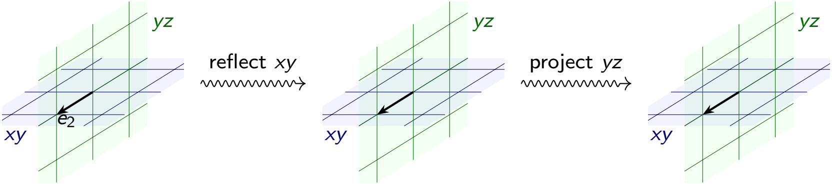
$$T(e_1) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Linear Transformations are Matrix Transformations

Example, continued

Question

What is the matrix for the linear transformation $T\colon \mathbf{R}^3 \to \mathbf{R}^3$ that reflects through the *xy*-plane and then projects onto the *yz*-plane?



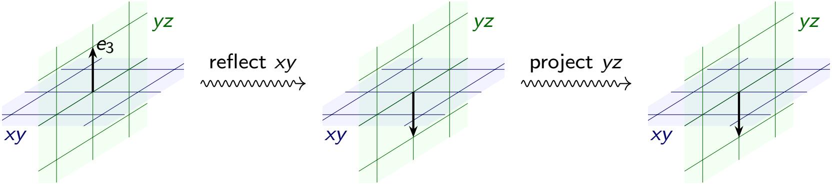
$$T(e_2) = e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Linear Transformations are Matrix Transformations

Example, continued

Question

What is the matrix for the linear transformation $T\colon \mathbf{R}^3 \to \mathbf{R}^3$ that reflects through the xy-plane and then projects onto the yz-plane?



$$T(e_3) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Linear Transformations are Matrix Transformations

Example, continued

Question

What is the matrix for the linear transformation $T\colon \mathbf{R}^3 \to \mathbf{R}^3$ that reflects through the *xy*-plane and then projects onto the *yz*-plane?

$$\left.\begin{aligned} T(e_1) &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ T(e_2) &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ T(e_1) &= \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}\right\} \implies A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

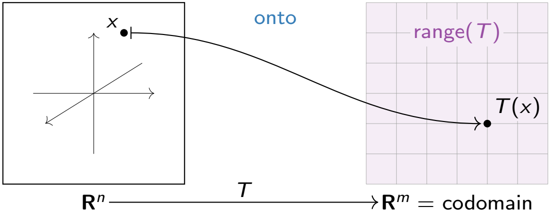
Other Geometric Transformations

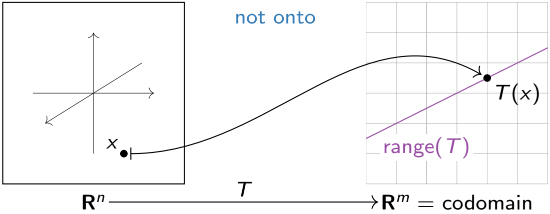
There is a long list of geometric transformations of $\mathbf{R}^2$ in §1.9 of Lay. (Reflections over the diagonal, contractions and expansions along different axes, shears, projections, . . .) Please look them over.

Onto Transformations

Definition

A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is **onto** (or **surjective**) if the range of T is equal to $\mathbf{R}^m$ (its codomain). In other words, each b in $\mathbf{R}^m$ is the image of *at least one* x in $\mathbf{R}^n$: every possible output has an input. Note that *not* onto means there is some b in $\mathbf{R}^m$ which is not the image of any x in $\mathbf{R}^n$.





Characterization of Onto Transformations

Theorem

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation with matrix A. Then the following are equivalent:

- T is onto
- $T(x) = b$ has a solution for every b in $\mathbf{R}^m$
- $Ax = b$ is consistent for every b in $\mathbf{R}^m$
- The columns of A span $\mathbf{R}^m$
- A has a pivot in every row

Question

If $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is onto, what can we say about the relative sizes of n and m?

Answer: T corresponds to an $m \times n$ matrix A. In order for A to have a pivot in every row, it must have *at least as many* columns as rows: $m \leq n$.

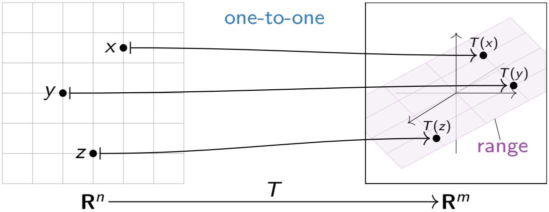
$$\begin{pmatrix} 1 & 0 & \star & 0 & \star \\ 0 & 1 & \star & 0 & \star \\ 0 & 0 & 0 & 1 & \star \end{pmatrix}$$

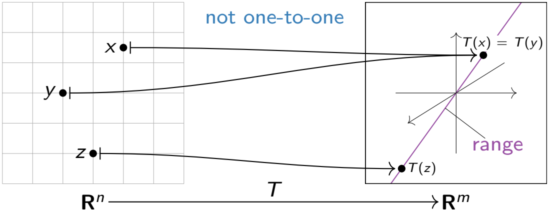
For instance, $\mathbf{R}^2$ is "too small" to map *onto* $\mathbf{R}^3$.

One-to-one Transformations

Definition

A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is **one-to-one** (or **into**, or **injective**) if different vectors in $\mathbf{R}^n$ map to different vectors in $\mathbf{R}^m$. In other words, each b in $\mathbf{R}^m$ is the image of *at most one* x in $\mathbf{R}^n$: different inputs have different outputs. Note that *not* one-to-one means different vectors in $\mathbf{R}^n$ have the same image.





Characterization of One-to-One Transformations

Theorem

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation with matrix A. Then the following are equivalent:

- T is one-to-one
- $T(x) = b$ has one or zero solutions for every b in $\mathbf{R}^m$
- $Ax = b$ has a unique solution or is inconsistent for every b in $\mathbf{R}^m$
- $Ax = 0$ has a unique solution
- The columns of A are linearly independent
- A has a pivot in every column.

Question

If $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is one-to-one, what can we say about the relative sizes of n and m?

Answer: T corresponds to an $m \times n$ matrix A. In order for A to have a pivot in every column, it must have *at least as many* rows as columns: $n \leq m$.

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

For instance, $\mathbf{R}^3$ is "too big" to map *into* $\mathbf{R}^2$.

Chapter 2

Matrix Algebra

Section 2.1

Matrix Operations

Motivation

Recall: we can turn any system of linear equations into a matrix equation

$$Ax = b.$$

This notation is suggestive. Can we solve the equation by "dividing by A"?

$$x \overset{??}{=} \frac{b}{A}$$

Answer: Sometimes, but you have to know what you're doing.

Today we'll study *matrix algebra*: adding and multiplying matrices.

More Notation for Matrices

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

We write a_{ij} for the entry in the ith row and the jth column. It is called the ***ij*th entry** of the matrix.

$$\begin{array}{cc} \begin{pmatrix} a_{11} & \cdots & \boxed{a_{1j}} & \cdots & a_{1n} \\ \vdots & & \boxed{\vdots} & & \vdots \\ \boxed{a_{i1}} & \boxed{\cdots} & \boxed{a_{ij}} & \boxed{\cdots} & \boxed{a_{in}} \\ \vdots & & \boxed{\vdots} & & \vdots \\ a_{m1} & \cdots & \boxed{a_{mj}} & \cdots & a_{mn} \end{pmatrix} & i\text{th row} \\ j\text{th column} & \end{array}$$

The entries $a_{11}, a_{22}, a_{33}, \ldots$ are the **diagonal entries**; they form the **main diagonal** of the matrix.

$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & \boxed{a_{22}} & a_{23} \end{pmatrix} \quad \begin{pmatrix} \boxed{a_{11}} & a_{12} \\ a_{21} & \boxed{a_{22}} \\ a_{31} & a_{32} \end{pmatrix}$$

A **diagonal matrix** is a *square* matrix whose only nonzero entries are on the main diagonal.

$$\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

The $n \times n$ **identity matrix** I_n is the diagonal matrix with all diagonal entries equal to 1. It is special because $I_n v = v$ for *all* v in $\mathbf{R}^n$.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

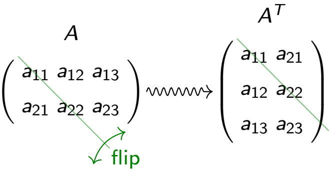
More Notation for Matrices

Continued

The **zero matrix** (of size $m \times n$) is the $m \times n$ matrix 0 with all zero entries.

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

The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose rows are the columns of A. In other words, the ij entry of A^T is a_{ji}.



$$\overset{A}{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}} \rightsquigarrow \overset{A^T}{\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}}$$

Addition and Scalar Multiplication

You add two matrices component by component, like with vectors.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{pmatrix}$$

Note you can only add two matrices *of the same size*.

You multiply a matrix by a scalar by multiplying each component, like with vectors.

$$c \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{pmatrix}.$$

These satisfy the expected rules, like with vectors:

Matrix Multiplication

Beware: matrix multiplication is more subtle than addition and scalar multiplication.

must be equal

Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix with columns $v_1, v_2 \ldots, v_p$:

$$B = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_p \\ | & | & & | \end{pmatrix}.$$

The **product** AB is the $m \times p$ matrix with columns $Av_1, Av_2, \ldots, Av_p$:

The equality is a definition

$$AB \overset{\text{def}}{=} \begin{pmatrix} | & | & & | \\ Av_1 & Av_2 & \cdots & Av_p \\ | & | & & | \end{pmatrix}.$$

In order for $Av_1, Av_2, \ldots, Av_p$ to make sense, the number of columns of A has to be the same as the number of rows of B.

Example

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 2 & -2 \\ 3 & -1 \end{pmatrix} =$$

Composition of Transformations

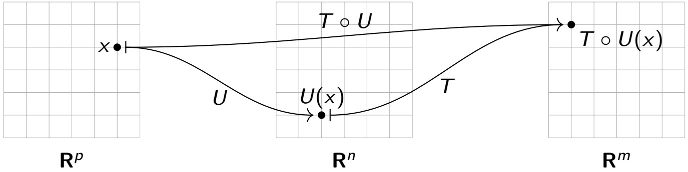
Why is this the correct definition of matrix multiplication?

Definition

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ and $U\colon \mathbf{R}^p \to \mathbf{R}^n$ be transformations. The **composition** is the transformation

$$T \circ U\colon \mathbf{R}^p \to \mathbf{R}^m \quad \text{defined by} \quad T \circ U(x) = T(U(x)).$$

This makes sense because $U(x)$ (the output of U) is in $\mathbf{R}^n$, which is the domain of T (the inputs of T).



Fact: If T and U are linear then so is $T \circ U$.

Guess: If A is the matrix for T, and B is the matrix for U, what is the matrix for $T \circ U$?

Composition of Linear Transformations

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ and $U\colon \mathbf{R}^p \to \mathbf{R}^n$ be *linear* transformations. Let A and B be their matrices:

$$A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix} \quad B = \begin{pmatrix} | & | & & | \\ U(e_1) & U(e_2) & \cdots & U(e_p) \\ | & | & & | \end{pmatrix}$$

Question

What is the matrix for $T \circ U$?

The matrix of the composition is the product of the matrices!

Composition of Linear Transformations

Example

Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be rotation by 45°, and let $U\colon \mathbf{R}^2 \to \mathbf{R}^2$ be projection onto the x-axis. Let's compute their standard matrices A and B:

$$\implies \quad A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Composition of Linear Transformations

Example, continued

So the matrix C for $T \circ U$ is

Check:

$$\implies C = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} \qquad \checkmark$$

Composition of Linear Transformations

Another example

Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -3 \\ 2 & -2 \\ 3 & -1 \end{pmatrix}.$$

Let $T(x) = Ax$ and $U(y) = By$, so

$$T\colon \mathbf{R}^3 \longrightarrow \mathbf{R}^2 \qquad U\colon \mathbf{R}^2 \longrightarrow \mathbf{R}^3 \qquad T \circ U\colon \mathbf{R}^2 \longrightarrow \mathbf{R}^2.$$

Let's find the matrix for $T \circ U$:

$$T \circ U(e_1) =$$

$$T \circ U(e_2) =$$

Before we computed $AB = \begin{pmatrix} 14 & -10 \\ 32 & -28 \end{pmatrix}$, so AB is the matrix of $T \circ U$.

Poll

The Row-Column Rule for Matrix Multiplication

Recall: A row vector of length n times a column vector of length n is a scalar:

$$\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + \cdots + a_n b_n.$$

Another way of multiplying a matrix by a vector is:

$$Ax = \begin{pmatrix} \text{—} \; r_1 \; \text{—} \\ \vdots \\ \text{—} \; r_m \; \text{—} \end{pmatrix} x = \begin{pmatrix} r_1 x \\ \vdots \\ r_m x \end{pmatrix}.$$

On the other hand, you multiply two matrices by

$$AB = A \begin{pmatrix} | & & | \\ c_1 & \cdots & c_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ Ac_1 & \cdots & Ac_p \\ | & & | \end{pmatrix}.$$

It follows that

$$AB = \begin{pmatrix} \text{—} \; r_1 \; \text{—} \\ \vdots \\ \text{—} \; r_m \; \text{—} \end{pmatrix} \begin{pmatrix} | & & | \\ c_1 & \cdots & c_p \\ | & & | \end{pmatrix} = \begin{pmatrix} r_1 c_1 & r_1 c_2 & \cdots & r_1 c_p \\ r_2 c_1 & r_2 c_2 & \cdots & r_2 c_p \\ \vdots & \vdots & & \vdots \\ r_m c_1 & r_m c_2 & \cdots & r_m c_p \end{pmatrix}$$

The Row-Column Rule for Matrix Multiplication

The ij entry of $C = AB$ is the ith row of A times the jth column of B:

$$c_{ij} = (AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

This is how everybody on the planet actually computes AB. Diagram ($AB = C$):

$$\begin{pmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ \boxed{a_{i1} \quad \cdots \quad a_{ik} \quad \cdots \quad a_{in}} & & & & \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mk} & \cdots & a_{mn} \end{pmatrix}_{i\text{th row}} \cdot \underset{j\text{th column}}{\begin{pmatrix} b_{11} & \cdots & \boxed{b_{1j}} & \cdots & b_{1p} \\ \vdots & & \vdots & & \vdots \\ b_{k1} & \cdots & b_{kj} & \cdots & b_{kp} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & \boxed{b_{nj}} & \cdots & b_{np} \end{pmatrix}} = \underset{ij\text{ entry}}{\begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1p} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & \boxed{c_{ij}} & \cdots & c_{ip} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mj} & \cdots & c_{mp} \end{pmatrix}}$$

Example

Properties of Matrix Multiplication

Mostly matrix multiplication works like you'd expect. Suppose A has size $m \times n$, and that the other matrices below have the right size to make multiplication work.

Most of these are easy to verify.

Associativity is $A(BC) = (AB)C$. It is a pain to verify using the row-column rule! Much easier: use associativity of linear transformations:

$$S \circ (T \circ U) = (S \circ T) \circ U.$$

This is a good example of an instance where having a conceptual viewpoint saves you a lot of work.

Recommended: Try to verify all of them on your own.

Properties of Matrix Multiplication

Caveats

Warnings!

- AB is usually not equal to BA.

  In fact, AB may be defined when BA is not.

- $AB = AC$ does not imply $B = C$, even if $A \neq 0$.

- $AB = 0$ does not imply $A = 0$ or $B = 0$.

Other Reading

Read about powers of a matrix and multiplication of transposes in §2.1.

Section 2.2

The Inverse of a Matrix

The Definition of Inverse

Recall: The multiplicative inverse (or reciprocal) of a nonzero number a is the number b such that $ab = 1$. We define the inverse of a matrix in almost the same way.

Definition

Let A be an $n \times n$ square matrix. We say A is **invertible** (or **nonsingular**) if there is a matrix B of the same size, such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

identity matrix

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

In this case, B is the **inverse** of A, and is written A^{-1}.

Example

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

The 2×2 case

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The **determinant** of A is the number

$$\boxed{\det(A) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.}$$

Facts:

1. If $\det(A) \neq 0$, then A is invertible and $\boxed{A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.}$
2. If $\det(A) = 0$, then A is not invertible.

Why 1?

Example

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} =$$

Solving Linear Systems via Inverses

Solving $Ax = b$ by "dividing by A"

Theorem

If A is invertible, then $Ax = b$ has exactly one solution for every b, namely:

$$x = A^{-1}b.$$

Why? Divide by A!

Example

Solve the system

$$\begin{aligned} 2x + 3y + 2z &= 1 \\ x \phantom{{}+3y} + 3z &= 1 \\ 2x + 2y + 3z &= 1 \end{aligned} \qquad \text{using} \qquad \begin{pmatrix} 2 & 3 & 2 \\ 1 & 0 & 3 \\ 2 & 2 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} -6 & -5 & 9 \\ 3 & 2 & -4 \\ 2 & 2 & -3 \end{pmatrix}.$$

Answer:

Some Facts

Say A and B are invertible $n \times n$ matrices.

1. A^{-1} is invertible and its inverse is $(A^{-1})^{-1} = A$.
2. AB is invertible and its inverse is $(AB)^{-1} = \xcancel{A^{-1}B^{-1}} \quad B^{-1}A^{-1}$.

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

   Why?

Computing A^{-1}

Let A be an $n \times n$ matrix. Here's how to compute A^{-1}.

1. Row reduce the augmented matrix $(\, A \mid I_n \,)$.
2. If the result has the form $(\, I_n \mid B \,)$, then A is invertible and $B = A^{-1}$.
3. Otherwise, A is not invertible.

Example

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

Computing A^{-1}

Example

Check:

Why Does This Work?

First answer: We can think of the algorithm as simultaneously solving the equations

$$Ax_1 = e_1 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$Ax_2 = e_2 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$Ax_3 = e_3 : \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right)$$

Now note $A^{-1}e_i = A^{-1}(Ax_i) = x_i$, and x_i is the ith column in the augmented part. Also $A^{-1}e_i$ is the ith column of A^{-1}.

Second answer: Elementary matrices.

Elementary Matrices

Definition

An **elementary matrix** is a square matrix E which differs from I_n by one row operation.

There are three kinds, corresponding to the three elementary row operations:

Fact: if E is the elementary matrix for a row operation, then EA differs from A by the same row operation.

Example:

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

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{pmatrix} \xrightarrow{R_2 = R_2 + 2R_1} \begin{pmatrix} 1 & 0 & 4 \\ 2 & 1 & 10 \\ 0 & -3 & -4 \end{pmatrix}$$

Elementary Matrices

Continued

Fact: if E is the elementary matrix for a row operation, then EA differs from A by the same row operation.

Consequence

Elementary matrices are invertible, and the inverse is the elementary matrix which un-does the row operation.

Why Does The Inversion Algorithm Work?

Second answer

Theorem

An $n \times n$ matrix A is invertible if and only if it is row equivalent to I_n. In this case, the sequence of row operations taking A to I_n also takes I_n to A^{-1}.

Why? Say the row operations taking A to I_n have elementary matrices $E_1, E_2, \ldots, E_k$.

This means if you do these same row operations to A and to I_n, you'll end up with I_n and A^{-1}. This is what you do when you row reduce the augmented matrix:

$$\left(\begin{array}{c|c} A & I_n \end{array}\right) \rightsquigarrow \left(\begin{array}{c|c} I_n & A^{-1} \end{array}\right)$$

Section 2.3

Characterization of Invertible Matrices

Invertible Transformations

Definition

A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^n$ is **invertible** if there exists another transformation $U\colon \mathbf{R}^n \to \mathbf{R}^n$ such that

$$T \circ U(x) = x \quad \text{and} \quad U \circ T(x) = x$$

for all x in $\mathbf{R}^n$. In this case we say U is the **inverse** of T, and we write $U = T^{-1}$.

In other words, $T(U(x)) = x$, so T "undoes" U, and likewise U "undoes" T.

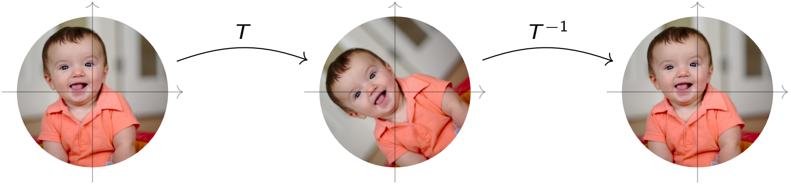
Fact

A transformation T is invertible if and only if it is both one-to-one and onto.

Invertible Transformations

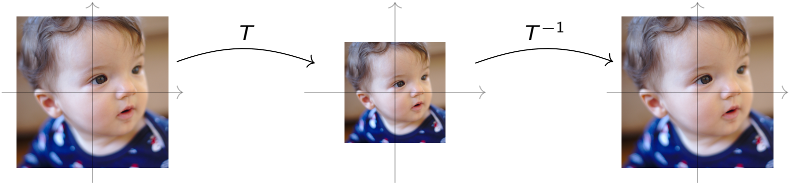
Examples

Let T = counterclockwise rotation in the plane by 45°. What is T^{-1}?



T^{-1} is *clockwise* rotation by 45°.

Let T = shrinking by a factor of $2/3$ in the plane. What is T^{-1}?



T^{-1} is *stretching* by $3/2$.

Let T = projection onto the x-axis. What is T^{-1}? It is not invertible: you can't undo it.

Invertible Linear Transformations

If $T\colon \mathbf{R}^n \to \mathbf{R}^n$ is an invertible *linear* transformation with matrix A, then what is the matrix for T^{-1}?

Fact

If T is an invertible linear transformation with matrix A, then T^{-1} is an invertible linear transformation with matrix A^{-1}.

Invertible Linear Transformations

Examples

Let $T =$ counterclockwise rotation in the plane by 45°. Its matrix is

Then $T^{-1} =$ counterclockwise rotation by -45°. Its matrix is

Check:

Let $T =$ shrinking by a factor of $2/3$ in the plane. Its matrix is

Then $T^{-1} =$ stretching by $3/2$. Its matrix is

Check:

The Invertible Matrix Theorem

A.K.A. The Really Big Theorem of Math 1553

The Invertible Matrix Theorem

Let A be an $n \times n$ matrix, and let $T\colon \mathbf{R}^n \to \mathbf{R}^n$ be the linear transformation $T(x) = Ax$. The following statements are equivalent.

1. A is invertible.
2. T is invertible.
3. A is row equivalent to I_n.
4. A has n pivots.
5. $Ax = 0$ has only the trivial solution.
6. The columns of A are linearly independent.
7. T is one-to-one.
8. $Ax = b$ is consistent for all b in $\mathbf{R}^n$.
9. The columns of A span $\mathbf{R}^n$.
10. T is onto.
11. A has a left inverse (there exists B such that $BA = I_n$).
12. A has a right inverse (there exists B such that $AB = I_n$).
13. A^T is invertible.

you really have to know these

The Invertible Matrix Theorem

Summary

There are two kinds of *square* matrices:

1. invertible (non-singular), and
2. non-invertible (singular).

For invertible matrices, all statements of the Invertible Matrix Theorem are true.

For non-invertible matrices, all statements of the Invertible Matrix Theorem are false.

Strong recommendation: If you want to understand invertible matrices, go through all of the conditions of the IMT and try to figure out on your own (or at least with help from the book) why they're all equivalent.

You know enough at this point to be able to reduce all of the statements to assertions about the pivots of a square matrix.

Section 2.8

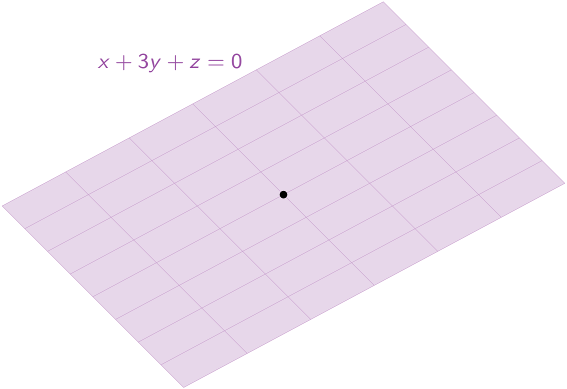
Subspaces of $\mathbf{R}^n$

Motivation

Today we will discuss **subspaces** of $\mathbf{R}^n$.

A subspace turns out to be the same as a span, except we don't know *which* vectors it's the span of.

This arises naturally when you have, say, a plane through the origin in $\mathbf{R}^3$ which is *not* defined (a priori) as a span, but you still want to say something about it.



Definition of Subspace

Definition

A **subspace** of $\mathbf{R}^n$ is a subset V of $\mathbf{R}^n$ satisfying:

1. The zero vector is in V. "not empty"
2. If u and v are in V, then $u + v$ is also in V. "closed under addition"
3. If u is in V and c is in $\mathbf{R}$, then cu is in V. "closed under × scalars"

What does this mean?

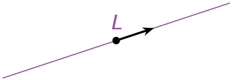
- If v is in V, then all scalar multiples of v are in V by (3). That is, the line through v is in V.
- If u, v are in V, then xu and yv are in V for scalars x, y by (3). So $xu + yv$ is in V by (2). So $\text{Span}\{u, v\}$ is contained in V.
- Likewise, if $v_1, v_2, \ldots, v_n$ are all in V, then $\text{Span}\{v_1, v_2, \ldots, v_n\}$ is contained in V.

A subspace V contains the span of any set of vectors in V.

Examples

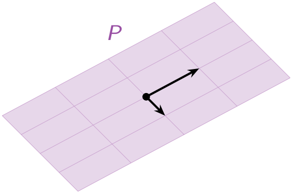
Example

A line L through the origin: this contains the span of any vector in L.



Example

A plane P through the origin: this contains the span of any vectors in P.



Example

All of $\mathbf{R}^n$: this contains 0, and is closed under addition and scalar multiplication.

Example

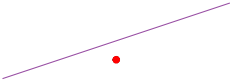
The subset $\{0\}$: this subspace contains only one vector.

Note these are all pictures of spans! (Line, plane, space, etc.)

Non-Examples

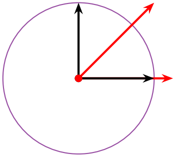
Non-Example

A line L (or any other set) that doesn't contain the origin is not a subspace. Fails: 1.



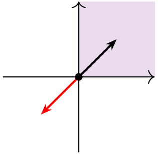
Non-Example

A circle C is not a subspace. Fails: 1,2,3. Think: a circle isn't a "linear space."



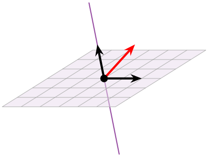
Non-Example

The first quadrant in $\mathbf{R}^2$ is not a subspace. Fails: 3 only.



Non-Example

A line union a plane in $\mathbf{R}^3$ is not a subspace. Fails: 2 only.



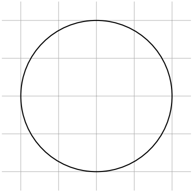
Subsets and Subspaces

They aren't the same thing

A **subset** of $\mathbf{R}^n$ is any collection of vectors whatsoever.

All of the non-examples are still subsets.

A **subspace** is a special kind of subset, which satisfies the three defining properties.



Subset: *yes*
Subspace: *no*

Spans are Subspaces

Theorem

Any $\text{Span}\{v_1, v_2, \ldots, v_n\}$ is a subspace.

!!!

Every subspace is a span, and every span is a subspace.

Definition

If $V = \text{Span}\{v_1, v_2, \ldots, v_n\}$, we say that V is the subspace **generated by** or **spanned by** the vectors $v_1, v_2, \ldots, v_n$.

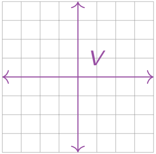
Poll

Question: What is the difference between $\{\}$ and $\{0\}$?

Subspaces

Verification

Let $V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbf{R}^2 \;\middle|\; ab = 0 \right\}$. Let's check if V is a subspace or not.



We conclude that V is *not* a subspace. A picture is above. (It doesn't look like a span.)

Column Space and Null Space

An $m \times n$ matrix A naturally gives rise to *two* subspaces.

Definition

- The **column space** of A is the subspace of $\mathbf{R}^m$ spanned by the columns of A. It is written $\operatorname{Col} A$.
- The **null space** of A is the set of all solutions of the homogeneous equation $Ax = 0$:
$$\operatorname{Nul} A = \{x \text{ in } \mathbf{R}^n \mid Ax = 0\}.$$
This is a subspace of $\mathbf{R}^n$.

The column space is defined as a span, so we know it is a subspace. It is the range (as opposed to the codomain) of the transformation $T(x) = Ax$.

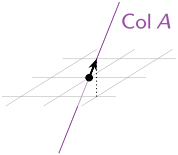
Check that the null space is a subspace:

Column Space and Null Space

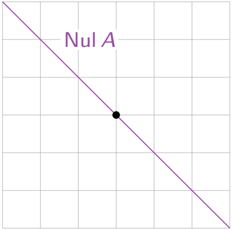
Example

Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Let's compute the column space:



Let's compute the null space:



The Null Space is a Span

The column space of a matrix A is defined to be a span (of the columns).

The null space is defined to be the solution set to $Ax = 0$. It is a subspace, so it is a span.

Question

How to find vectors which span the null space?

Answer: Parametric vector form! We know that the solution set to $Ax = 0$ has a parametric form that looks like

$$x_3 \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix}$$

if, say, x_3 and x_4 are the free variables. So

$$\operatorname{Nul} A = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

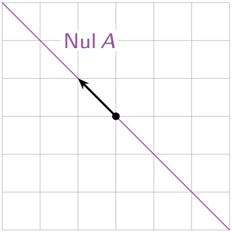
Refer back to the slides for §1.5 (Solution Sets).

Note: It is much easier to define the null space first as a subspace, then find spanning vectors *later*, if we need them. This is one reason subspaces are so useful.

The Null Space is a Span

Example, revisited

Find vector(s) that span the null space of $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$.



Subspaces

Summary

How do you check if a subset is a subspace?

- Is it a span? Can it be written as a span?
- Can it be written as the column space of a matrix?
- Can it be written as the null space of a matrix?
- Is it all of $\mathbf{R}^n$ or the zero subspace $\{0\}$?
- Can it be written as a type of subspace that we'll learn about later (eigenspaces, . . .)?

If so, then it's automatically a subspace.

If all else fails:

- Can you verify directly that it satisfies the three defining properties?

Basis of a Subspace

What is the *smallest number* of vectors that are needed to span a subspace?

Definition

Let V be a subspace of $\mathbf{R}^n$. A **basis** of V is a set of vectors $\{v_1, v_2, \ldots, v_m\}$ in V such that:

1. $V = \text{Span}\{v_1, v_2, \ldots, v_m\}$, and
2. $\{v_1, v_2, \ldots, v_m\}$ is linearly independent.

The number of vectors in a basis is the **dimension** of V, and is written $\dim V$.

Note the big red border here

Why is a basis the smallest number of vectors needed to span?

Recall: *linearly independent* means that every time you add another vector, the span gets bigger.

Hence, if we remove any vector, the span gets *smaller*: so any smaller set can't span V.

Important

A subspace has *many different* bases, but they all have the same number of vectors (see the exercises in §2.9).

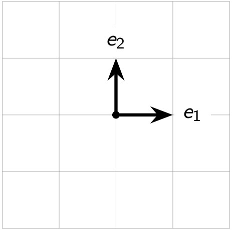
Bases of $\mathbf{R}^2$

Question

What is a basis for $\mathbf{R}^2$?

We need two vectors that *span* $\mathbf{R}^2$ and are *linearly independent*. $\{e_1, e_2\}$ is one basis.

1. They span: $\binom{a}{b} = ae_1 + be_2$.
2. They are linearly independent because they are not collinear.

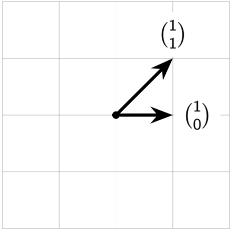


Question

What is another basis for $\mathbf{R}^2$?

Any two nonzero vectors that are not collinear. $\left\{\binom{1}{0}, \binom{1}{1}\right\}$ is also a basis.

1. They span: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has a pivot in every row.
2. They are linearly independent: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has a pivot in every column.



Bases of $\mathbf{R}^n$

The unit coordinate vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \ldots, \quad e_{n-1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

are a basis for $\mathbf{R}^n$.

The identity matrix has columns $e_1, e_2, \ldots, e_n$.

1. They span: I_n has a pivot in every row.
2. They are linearly independent: I_n has a pivot in every column.

In general: $\{v_1, v_2, \ldots, v_n\}$ is a basis for $\mathbf{R}^n$ if and only if the matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

has a pivot in every row and every column, i.e. if A is *invertible*.

Basis of a Subspace

Example

Example

Let

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ in } \mathbf{R}^3 \;\middle|\; x + 3y + z = 0 \right\} \qquad \mathcal{B} = \left\{ \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} \right\}.$$

Verify that $\mathcal{B}$ is a basis for V.

Basis for Nul A

Fact

The vectors in the parametric vector form of the general solution to $Ax = 0$ always form a basis for Nul A.

Example

Basis for Col A

Fact

The *pivot columns* of A always form a basis for Col A.

Warning: I mean the pivot columns of the *original* matrix A, not the row-reduced form. (Row reduction changes the column space.)

Example

Why? End of §2.8, or ask in office hours.

Section 2.9

Dimension and Rank

Coefficients of Basis Vectors

Recall: a **basis** of a subspace V is a set of vectors that *spans* V and is *linearly independent*.

Lemma ← like a theorem, but less important

If $\mathcal{B} = \{v_1, v_2, \dots, v_m\}$ is a basis for a subspace V, then any vector x in V can be written as a linear combination

$$x = c_1 v_1 + c_2 v_2 + \cdots + c_m v_m$$

for *unique* coefficients $c_1, c_2, \dots, c_m$.

Bases as Coordinate Systems

The unit coordinate vectors $e_1, e_2, \ldots, e_n$ form a basis for $\mathbf{R}^n$. Any vector is a unique linear combination of the e_i:

$$v = \begin{pmatrix} 3 \\ 5 \\ -2 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - 2\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3e_1 + 5e_2 - 2e_3.$$

Observe: the *coordinates* of v are exactly the *coefficients* of e_1, e_2, e_3.

We can go backwards: given any basis $\mathcal{B}$, we interpret the coefficients of a linear combination as "coordinates" with respect to $\mathcal{B}$.

Definition

Let $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ be a basis of a subspace V. Any vector x in V can be written uniquely as a linear combination $x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$. The coefficients $c_1, c_2, \ldots, c_m$ are the **coordinates of x with respect to $\mathcal{B}$**. The **$\mathcal{B}$-coordinate vector of x** is the vector

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \quad \text{in } \mathbf{R}^m.$$

Bases as Coordinate Systems

Example 1

Let $v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathcal{B} = \{v_1, v_2\}, \quad V = \text{Span}\{v_1, v_2\}.$

Verify that $\mathcal{B}$ is a basis:

Question: If $[x]_{\mathcal{B}} = \binom{5}{2}$, then what is x?

Question: Find the $\mathcal{B}$-coordinates of $x = \begin{pmatrix} 5 \\ 3 \\ 5 \end{pmatrix}$.

Bases as Coordinate Systems

Example 2

Let $v_1 = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} 2 \\ 8 \\ 6 \end{pmatrix}, \quad V = \text{Span}\{v_1, v_2, v_3\}.$

Question: Find a basis for V.

Question: Find the $\mathcal{B}$-coordinates of $x = \begin{pmatrix} 4 \\ 11 \\ 8 \end{pmatrix}$.

Bases as Coordinate Systems

Summary

If $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ is a basis for a subspace V and x is in V, then

$$\boxed{[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \quad \text{means} \quad x = c_1v_1 + c_2v_2 + \cdots + c_mv_m.}$$

Finding the $\mathcal{B}$-coordinates for x means solving the vector equation

$$x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$$

in the unknowns $c_1, c_2, \ldots, c_m$. This (usually) means row reducing the augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_m & x \\ | & | & & | & | \end{array}\right).$$

Question: What happens if you try to find the $\mathcal{B}$-coordinates of x *not* in V?

Bases as Coordinate Systems

Picture

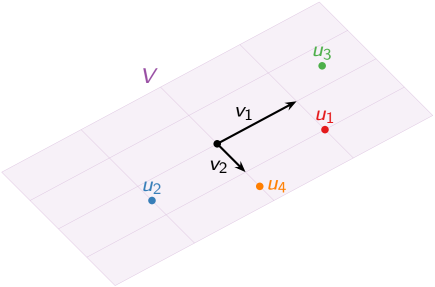
Let

$$v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

These form a basis $\mathcal{B}$ for the plane

$$V = \mathrm{Span}\{v_1, v_2\}$$

in $\mathbf{R}^3$.



Question: Estimate the $\mathcal{B}$-coordinates of these vectors:

$$[u_1]_{\mathcal{B}} = \qquad [u_2]_{\mathcal{B}} = \qquad [u_3]_{\mathcal{B}} = \qquad [u_4]_{\mathcal{B}} =$$

Remark

Many of you want to think of a plane in $\mathbf{R}^3$ as "being" $\mathbf{R}^2$. Choosing a basis $\mathcal{B}$ and using $\mathcal{B}$-coordinates is one way to make sense of that. But remember that the coordinates are the coefficients of a linear combination of the basis vectors.

The Rank Theorem

Recall:

- The **dimension** of a subspace V is the number of vectors in a basis for V.
- A basis for the column space of a matrix A is given by the pivot columns.
- A basis for the null space of A is given by the vectors attached to the free variables in the parametric vector form.

Definition

The **rank** of a matrix A, written $\operatorname{rank} A$, is the dimension of the column space $\operatorname{Col} A$.

Observe:

$$\begin{aligned}\operatorname{rank} A = \dim \operatorname{Col} A &= \text{the number of columns with pivots} \\ \dim \operatorname{Nul} A &= \text{the number of free variables} \\ &= \text{the number of columns without pivots.}\end{aligned}$$

Rank Theorem

If A is an $m \times n$ matrix, then

$$\operatorname{rank} A + \dim \operatorname{Nul} A = n = \text{the number of columns of } A.$$

The Rank Theorem

Example

$$A = \begin{pmatrix} 1 & 2 & 0 & -1 \\ -2 & -3 & 4 & 5 \\ 2 & 4 & 0 & -2 \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

basis of Col A

free variables

Poll

The Basis Theorem

Basis Theorem

Let V be a subspace of dimension m. Then:

- Any m linearly independent vectors in V form a basis for V.
- Any m vectors that span V form a basis for V.

Upshot

If you *already* know that $\dim V = m$, and you have m vectors $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ in V, then you only have to check *one* of

1. $\mathcal{B}$ is linearly independent, *or*
2. $\mathcal{B}$ spans V

in order for $\mathcal{B}$ to be a basis.

The Invertible Matrix Theorem

Addenda

The Invertible Matrix Theorem

Let A be an $n \times n$ matrix, and let $T\colon \mathbf{R}^n \to \mathbf{R}^n$ be the linear transformation $T(x) = Ax$. The following statements are equivalent.

1. A is invertible.
2. T is invertible.
3. A is row equivalent to I_n.
4. A has n pivots.
5. $Ax = 0$ has only the trivial solution.
6. The columns of A are linearly independent.
7. T is one-to-one.
8. $Ax = b$ is consistent for all b in $\mathbf{R}^n$.
9. The columns of A span $\mathbf{R}^n$.
10. T is onto.
11. A has a left inverse (there exists B such that $BA = I_n$).
12. A has a right inverse (there exists B such that $AB = I_n$).
13. A^T is invertible.
14. The columns of A form a basis for $\mathbf{R}^n$.
15. $\operatorname{Col} A = \mathbf{R}^n$.
16. $\dim \operatorname{Col} A = n$.
17. $\operatorname{rank} A = n$.
18. $\operatorname{Nul} A = \{0\}$.
19. $\dim \operatorname{Nul} A = 0$.

These are equivalent to the previous conditions by the Rank Theorem and the Basis Theorem.

Chapter 3

Determinants

Section 3.1

Introduction to Determinants

Orientation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$
  We've said most of what we'll say about this topic now.
- Solve the matrix equation $Ax = \lambda x$ (eigenvalue problem)
  We are now aiming at this.
- Almost solve the equation $Ax = b$
  This will happen later.

The next topic is *determinants*.

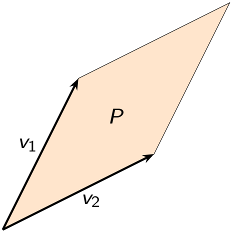
Dan Margalit has written some notes which, in my opinion, explain the topic in a much better way than Lay does. (Both cover the same material.)

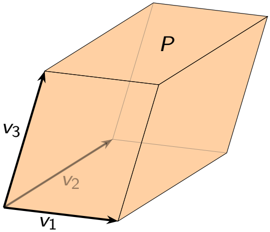
Prof. Margalit's notes are the primary reference for Chapter 3.

The Idea of Determinants

Let A be an $n \times n$ matrix. Determinants are only for square matrices.

The columns $v_1, v_2, \ldots, v_n$ give you n vectors in $\mathbf{R}^n$. These determine a **parallelepiped** P.





Observation: the volume of P is zero $\iff$ the columns are *linearly dependent* (P is "flat") $\iff$ the matrix A is not invertible.

The **determinant** of A will be a number $\det(A)$ whose absolute value is the volume of P. In particular, $\det(A) \neq 0 \iff A$ is invertible.

Determinants of 2 × 2 Matrices

Revisited

We already have a formula in the 2 × 2 case:

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

What does this have to do with volumes?

The area of the parallelogram is always $|ad - bc|$. If v_1 is not on the x-axis: it's a fun geometry problem!

Note: this shows $\det(A) \neq 0 \iff A$ is invertible in this case. (The volume is zero if and only if the columns are collinear.)

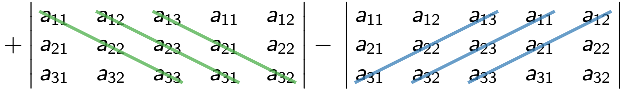
Question: What does the sign of the determinant mean?

Determinants of 3 × 3 Matrices

Here's the formula:

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ \quad - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{array}$$

How on earth do you remember this? Draw a bigger matrix, repeating the first two columns to the right:



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

Then add the products of the downward diagonals, and subtract the product of the upward diagonals. For example,

$$\det\begin{pmatrix} 5 & 1 & 0 \\ -1 & 3 & 2 \\ 4 & 0 & -1 \end{pmatrix} =$$

What does this have to do with volumes? Next time.

A Formula for the Determinant

When $n \geq 4$, the determinant isn't just a sum of products of diagonals. The formula is *recursive*: you compute a larger determinant in terms of smaller ones.

First some notation. Let A be an $n \times n$ matrix.

$$\begin{aligned} A_{ij} &= ij\text{th } \textbf{minor} \text{ of } A \\ &= (n-1) \times (n-1) \text{ matrix you get by deleting the } i\text{th row and } j\text{th column} \end{aligned}$$

$$\begin{aligned} C_{ij} &= (-1)^{i+j} \det A_{ij} \\ &= ij\text{th } \textbf{cofactor} \text{ of } A \end{aligned}$$

The signs of the cofactors follow a checkerboard pattern:

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix} \qquad \pm \text{ in the } ij \text{ entry is the sign of } C_{ij}$$

Definition

The **determinant** of an $n \times n$ matrix A is

$$\det(A) = \sum_{j=1}^{n} a_{1j} C_{1j} = a_{11} C_{11} + a_{12} C_{12} + \cdots + a_{1n} C_{1n}.$$

This formula is called **cofactor expansion along the first row**.

A Formula for the Determinant

1×1 Matrices

This is the beginning of the recursion.

$$\det(\, a_{11} \,) = a_{11}.$$

A Formula for the Determinant

2×2 Matrices

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

The minors are:

$A_{11} =$

$A_{12} =$

$A_{21} =$

$A_{22} =$

The cofactors are

$C_{11} =$

$C_{12} =$

$C_{21} =$

$C_{22} =$

The determinant is

$$\det A = a_{11}C_{11} + a_{12}C_{12} = a_{11}a_{22} - a_{12}a_{21}.$$

A Formula for the Determinant

3 × 3 Matrices

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

The top row minors and cofactors are:

$A_{11} =$ $\qquad$ $C_{11} =$

$A_{12} =$ $\qquad$ $C_{12} =$

$A_{13} =$ $\qquad$ $C_{13} =$

The determinant is the same formula as before (as it turns out):

$$\begin{aligned} \det A &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \\ &= a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \end{aligned}$$

A Formula for the Determinant

Example

$$\det \begin{pmatrix} 5 & 1 & 0 \\ -1 & 3 & 2 \\ 4 & 0 & -1 \end{pmatrix} =$$

$2n - 1$ More Formulas for the Determinant

Recall: the formula

$$\det(A) = \sum_{j=1}^{n} a_{1j} C_{1j} = a_{11} C_{11} + a_{12} C_{12} + \cdots + a_{1n} C_{1n}.$$

is called **cofactor expansion along the first row.** Actually, you can expand cofactors along any row or column you like!

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{for any fixed } i$$

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{for any fixed } j$$

Try this with a row or a column with a lot of zeros.

Cofactor Expansion

Example

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 5 & 9 & 1 \end{pmatrix}$$

It looks easiest to expand along the third column:

$$\det A =$$

Poll

The Determinant of an Upper-Triangular Matrix

The computation in the poll works for any matrix that is *upper-triangular* (all entries below the main diagonal are zero).

Theorem

The determinant of an upper-triangular matrix is the product of the diagonal entries:

$$\det\begin{pmatrix} \enclose{circle}{a_{11}} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & \enclose{circle}{a_{22}} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & \enclose{circle}{a_{33}} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \enclose{circle}{a_{nn}} \end{pmatrix} = a_{11}a_{22}a_{33}\cdots a_{nn}.$$

The same is true for lower-triangular matrices. (Repeatedly expand along the first row.)

A Formula for the Inverse

For fun—from §3.3

For 2 × 2 matrices we had a nice formula for the inverse:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\det A} \begin{pmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{pmatrix}.$$

Theorem

This last formula works for any $n \times n$ invertible matrix A:

(3, 1) entry

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} C_{11} & C_{21} & C_{31} & \cdots & C_{n1} \\ C_{12} & C_{22} & C_{32} & \cdots & C_{n2} \\ \boxed{C_{13}} & C_{23} & C_{33} & \cdots & C_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & C_{3n} & \cdots & C_{nn} \end{pmatrix} = \frac{1}{\det A} \left(C_{ij}\right)^{T}$$

Note that the cofactors are “transposed”: the (i, j) entry of the matrix is C_{ji}.

The proof uses Cramer’s rule. See Dan Margalit’s notes on the website for a nice explanation.

A Formula for the Inverse

Example

Compute A^{-1}, where $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

The minors are:

The cofactors are (don't forget to multiply by $(-1)^{i+j}$):

The determinant is (expanding along the first row):

$$\det A =$$

A Formula for the Inverse

Example, continued

Compute A^{-1}, where $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

The inverse is

$$A^{-1} =$$

Check:

Section 3.2

Properties of Determinants

Plan for Today

Last time, we gave a recursive formula for determinants in terms of cofactor expansions.

Plan for today:

- An abstract definition of the determinant in terms of its properties.
- Computing determinants using row operations.
- Determinants and products: $\det(AB) = \det(A)\det(B)$.
- Determinants and volumes.
- Determinants and linear transformations.

The determinant is one of the most amazing functions ever devised. Today is about beginning to understand why.

The Determinant is a Function

We can think of the determinant as a function of the entries of a matrix:

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{aligned} &a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}. \end{aligned}$$

The formula for the determinant of an $n \times n$ matrix has $n!$ terms. So the determinant of a 10×10 matrix has 3,628,800 terms!

When mathematicians encounter a function whose formula is too difficult to write down, we try to *characterize* it in terms of its properties.

The determinant function is characterized by how it is changed by row operations.

Defining the Determinant in Terms of its Properties

Definition

The **determinant** is a function

$$\det\colon \{\text{square matrices}\} \longrightarrow \mathbf{R}$$

with the following **defining properties**:

1. $\det(I_n) = 1$
2. If we do a row replacement on a matrix, the determinant does not change.
3. If we swap two rows of a matrix, the determinant scales by -1.
4. If we scale a row of a matrix by k, the determinant scales by k.

Why would we think of these properties? This is how volumes work!

1. The volume of the unit cube is 1.
2. Volumes don't change under a shear.
3. Volume of a mirror image is negative of the volume?
4. If you scale one coordinate by k, the volume is multiplied by k.

Properties of the Determinant

2×2 matrix

Properties of the Determinant

Elementary matrices

Since an elementary matrix differs from the identity matrix by one row operation, and since $\det(I_n) = 1$, it is easy to calulate the determinant of an elementary matrix:

$$\det \begin{pmatrix} 1 & 0 & 8 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\det \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} =$$

$$\det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 17 & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

Computing the Determinant by Row Reduction

We can use the properties of the determinant and row reduction to compute the determinant of any matrix! This means that det is completely characterized by its defining properties.

$$\det\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 5 & 7 & -4 \end{pmatrix} =$$

Computing the Determinant by Row Reduction

Saving some work

The determinant of an upper (or lower) triangular matrix is the product of the diagonal entries, so we can stop row reducing when we get to row echelon form.

$$\det\begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ 5 & 7 & -4\end{pmatrix} = \cdots = -\det\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & -9\end{pmatrix} = 9.$$

This is almost always the easiest way to compute the determinant of a large, complicated matrix, either by hand or by computer.

(Cofactor expansion is $O(n!) \sim O(n^n\sqrt{n})$, row reduction is $O(n^3)$.)

Poll

A Mathematical IOU

The characterization of the determinant function in terms of its properties is very useful. It gives us a fast way to compute determinants, and prove other properties (later). But. . .

The disadvantage of defining a function by its properties instead of a formula is: how do you know such a function exists? and if it exists, why is there only one function satisfying those properties?

In our case, we can compute the determinant of a matrix from its defining properties, so if it exists, it is unique. But how do we know that two different row reductions won't give two different answers for the determinant?

Here is a summary of the magical properties of the determinant. Prof. Margalit's notes (on the website) have very understandable proofs.

Magical Properties of the Determinant

you really have to know these

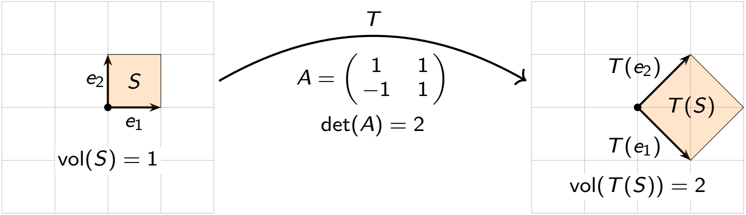
1. There is one and only one function $\det\colon \{\text{square matrices}\} \to \mathbf{R}$ satisfying the defining properties (1)–(4).
2. A is invertible if and only if $\det(A) \neq 0$.
3. If we row reduce A without row scaling, then
$$\det(A) = (-1)^{\#\text{swaps}}\bigl(\text{product of diagonal entries in REF}\bigr).$$
4. The determinant can be computed using any of the $2n$ cofactor expansions. (You get the same number every time!)
5. $\det(AB) = \det(A)\det(B)$ and $\det(A^{-1}) = \det(A)^{-1}$.
6. $\det(A) = \det(A^T)$.
7. $|\det(A)|$ is the volume of the parallelepiped defined by the columns of A.
8. If A is an $n \times n$ matrix with transformation $T(x) = Ax$, and S is a subset of $\mathbf{R}^n$, then the volume of $T(S)$ is $|\det(A)|$ times the volume of S. (Even for curvy shapes S.)
9. The determinant is multi-linear (we'll talk about this in a few slides).

Multiplicativity of the Determinant

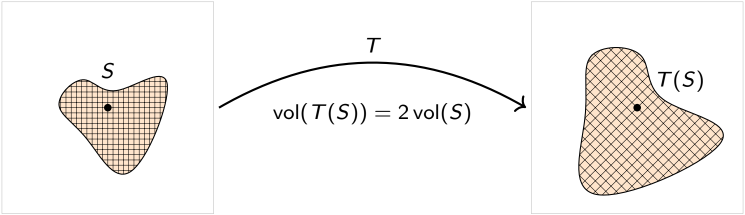
Why is Property 5 true? In Lay, there's a proof using elementary matrices. Here's a better one.

Determinants and Linear Transformations

Why is Property 8 true? For instance, if S is the unit cube, then $T(S)$ is the parallelepiped defined by the columns of A, since the columns of A are $T(e_1), T(e_2), \ldots, T(e_n)$. In this case, Property 8 is the same as Property 7.



For curvy shapes, you break S up into a bunch of tiny cubes. Each one is scaled by $|\det(A)|$; then you use *calculus* to reduce to the previous situation!



Multi-Linearity of the Determinant

We can also think of det as a function of the columns (or the rows) of an $n \times n$ matrix:

$$\det\colon \underbrace{\mathbf{R}^n \times \mathbf{R}^n \times \cdots \times \mathbf{R}^n}_{n \text{ times}} \longrightarrow \mathbf{R}$$

$$\det(v_1, v_2, \ldots, v_n) = \det\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Property 9 says that for any i and any vectors $v_1, v_2, \ldots, v_n$ and v_i' and any scalar c,

$$\begin{aligned} \det(v_1, \ldots, v_i + v_i', \ldots, v_n) &= \det(v_1, \ldots, v_i, \ldots, v_n) + \det(v_1, \ldots, v_i', \ldots, v_n) \\ \det(v_1, \ldots, cv_i, \ldots, v_n) &= c \det(v_1, \ldots, v_i, \ldots, v_n). \end{aligned}$$

In other words, scaling one column (or row) by c scales det by c (which we already knew), and if column i is a sum of two vectors v_i, v_i', then the determinant is the sum of two determinants, one with v_i in column i, and one with v_i' in column i. *This only works one column at a time.*

Proof: just expand cofactors along column i.

Chapter 5

Eigenvalues and Eigenvectors

Section 5.1

Eigenvectors and Eigenvalues

A Biology Question

Motivation

In a population of rabbits:

1. half of the newborn rabbits survive their first year;
2. of those, half survive their second year;
3. their maximum life span is three years;
4. rabbits have $0, 6, 8$ baby rabbits in their three years, respectively.

If you know the population one year, what is the population the next year?

$$\begin{aligned} f_n &= \text{first-year rabbits in year } n \\ s_n &= \text{second-year rabbits in year } n \\ t_n &= \text{third-year rabbits in year } n \end{aligned}$$

The rules say:

$$\begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} f_n \\ s_n \\ t_n \end{pmatrix} = \begin{pmatrix} f_{n+1} \\ s_{n+1} \\ t_{n+1} \end{pmatrix}.$$

Let $A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$ and $v_n = \begin{pmatrix} f_n \\ s_n \\ t_n \end{pmatrix}$. Then $\boxed{Av_n = v_{n+1}.}$ $\longleftarrow$ **difference equation**

A Biology Question

Continued

If you know v_0, what is v_{10}?

$$v_{10} = Av_9 = AAv_8 = \cdots = A^{10}v_0.$$

This makes it easy to compute examples by computer:

v_0	v_{10}	v_{11}
$\begin{pmatrix}3\\7\\9\end{pmatrix}$	$\begin{pmatrix}30189\\7761\\1844\end{pmatrix}$	$\begin{pmatrix}61316\\15095\\3881\end{pmatrix}$
$\begin{pmatrix}1\\2\\3\end{pmatrix}$	$\begin{pmatrix}9459\\2434\\577\end{pmatrix}$	$\begin{pmatrix}19222\\4729\\1217\end{pmatrix}$
$\begin{pmatrix}4\\7\\8\end{pmatrix}$	$\begin{pmatrix}28856\\7405\\1765\end{pmatrix}$	$\begin{pmatrix}58550\\14428\\3703\end{pmatrix}$

What do you notice about these numbers?

1. Eventually, each segment of the population doubles every year: $Av_n = v_{n+1} = 2v_n$.
2. The ratios get close to $(16 : 4 : 1)$:

$$v_n = (\text{scalar}) \cdot \begin{pmatrix}16\\4\\1\end{pmatrix}.$$

Translation: 2 is an eigenvalue, and $\begin{pmatrix}16\\4\\1\end{pmatrix}$ is an eigenvector!

Eigenvectors and Eigenvalues

Definition

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

Eigenvalues and eigenvectors are only for square matrices.

1. An **eigenvector** of A is a *nonzero* vector v in $\mathbf{R}^n$ such that $Av = \lambda v$, for some λ in $\mathbf{R}$. In other words, Av is a multiple of v.
2. An **eigenvalue** of A is a number λ in $\mathbf{R}$ such that the equation $Av = \lambda v$ has a *nontrivial* solution.

If $Av = \lambda v$ for $v \neq 0$, we say λ is the **eigenvalue for** v, and v is an **eigenvector for** λ.

Note: Eigenvectors are by definition nonzero. Eigenvalues may be equal to zero.

This is the most important definition in the course.

Verifying Eigenvectors

Example

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \qquad v = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix}$$

Multiply:

$$Av =$$

Hence v is an eigenvector of A, with eigenvalue $\lambda = 2$.

Example

$$A = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \qquad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Multiply:

$$Av =$$

Hence v is an eigenvector of A, with eigenvalue $\lambda = 4$.

Poll

Verifying Eigenvalues

Question: Is $\lambda = 3$ an eigenvalue of $A = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix}$?

In other words, does $Av = 3v$ have a nontrivial solution?

. . . does $Av - 3v = 0$ have a nontrivial solution?

. . . does $(A - 3I)v = 0$ have a nontrivial solution?

We know how to answer that! Row reduction!

$$A - 3I =$$

Eigenspaces

Definition

Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The λ-**eigenspace** of A is the set of all eigenvectors of A with eigenvalue λ, plus the zero vector:

$$\begin{aligned}\lambda\text{-eigenspace} &= \bigl\{v \text{ in } \mathbf{R}^n \mid Av = \lambda v\bigr\} \\ &= \bigl\{v \text{ in } \mathbf{R}^n \mid (A - \lambda I)v = 0\bigr\} \\ &= \mathrm{Nul}\bigl(A - \lambda I\bigr).\end{aligned}$$

Since the λ-eigenspace is a null space, it is a *subspace* of $\mathbf{R}^n$.

How do you find a basis for the λ-eigenspace? Parametric vector form!

Eigenspaces

Example

Find a basis for the 2-eigenspace of

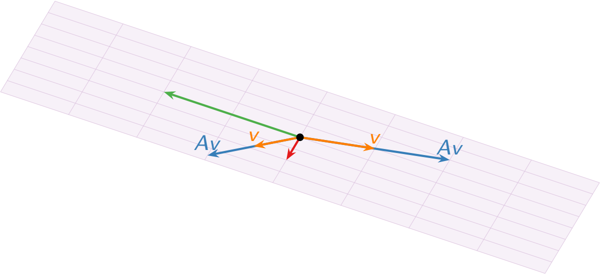
λ

$$A = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}.$$

Eigenspaces

Picture

A basis for the 2-eigenspace of $\begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}$ is $\left\{ \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\}$. What does this look like?



For any v in the 2-eigenspace, $Av = 2v$ by definition. So A acts by *scaling by* 2 on its 2-eigenspace. This is how eigenvalues and eigenvectors make matrices easier to understand.

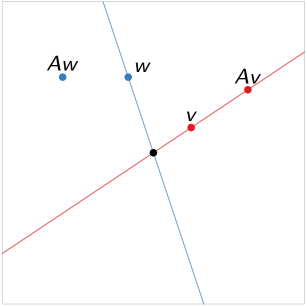
Eigenspaces

Geometry

Eigenvectors, geometrically

An eigenvector of a matrix A is a nonzero vector v such that:

- Av is a multiple of v, which means
- Av is collinear with v, which means
- Av and v are *on the same line*.



v is an eigenvector

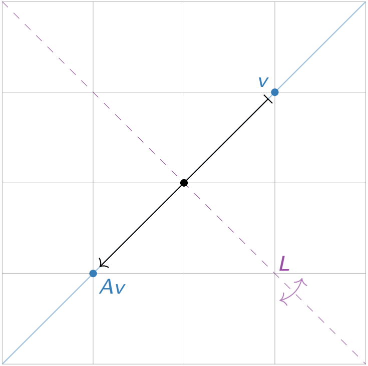
w is not an eigenvector

Eigenspaces

Geometry; example

Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

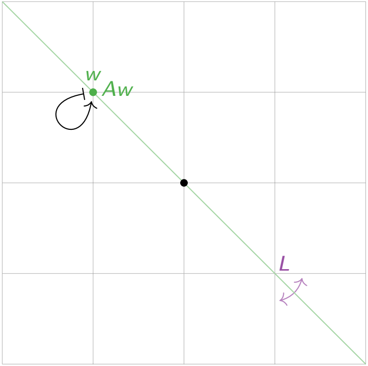
v is an eigenvector with eigenvalue -1.

Eigenspaces

Geometry; example

Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

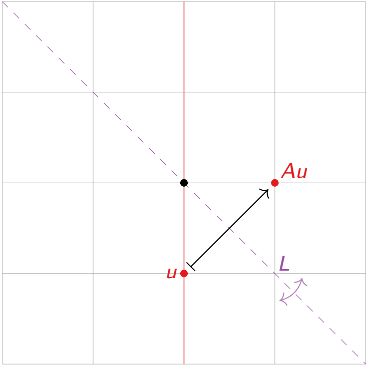
w is an eigenvector with eigenvalue 1.

Eigenspaces

Geometry; example

Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

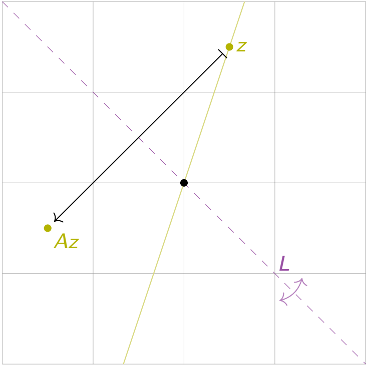
u is *not* an eigenvector.

Eigenspaces

Geometry; example

Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

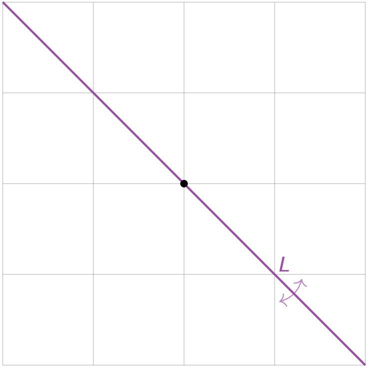
Neither is z.

Eigenspaces

Geometry; example

Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

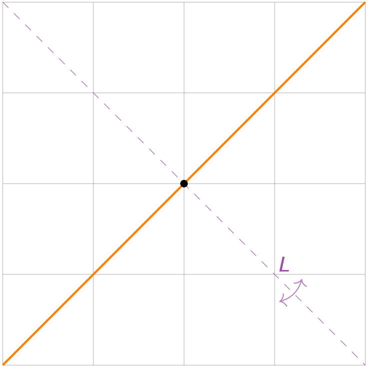
The 1-eigenspace is L
(all the vectors x where $Ax = x$).

Eigenspaces

Geometry; example

Let $T: \mathbf{R}^2 \to \mathbf{R}^2$ be reflection over the line L defined by $y = -x$, and let A be the matrix for T.

Question: What are the eigenvalues and eigenspaces of A? No computations!



Does anyone see any eigenvectors (vectors that don't move off their line)?

The (-1)-eigenspace is the line $y = x$ (all the vectors x where $Ax = -x$).

Eigenspaces

Summary

Let A be an $n \times n$ matrix and let λ be a number.

1. λ is an eigenvalue of A if and only if $(A - \lambda I)x = 0$ has a nontrivial solution, if and only if $\mathrm{Nul}(A - \lambda I) \neq \{0\}$.
2. In this case, finding a basis for the λ-eigenspace of A means finding a basis for $\mathrm{Nul}(A - \lambda I)$ as usual, i.e. by finding the parametric vector form for the general solution to $(A - \lambda I)x = 0$.
3. The eigenvectors with eigenvalue λ are the nonzero elements of $\mathrm{Nul}(A - \lambda I)$, i.e. the nontrivial solutions to $(A - \lambda I)x = 0$.

The Eigenvalues of a Triangular Matrix are the Diagonal Entries

We've seen that finding eigenvectors for a given eigenvalue is a row reduction problem.

Finding all of the eigenvalues of a matrix *is not a row reduction problem!* We'll see how to do it in general next time. For now:

Fact: The eigenvalues of a triangular matrix are the diagonal entries.

A Matrix is Invertible if and only if Zero is not an Eigenvalue

Fact: A is invertible if and only if 0 is not an eigenvalue of A.

Eigenvectors with Distinct Eigenvalues are Linearly Independent

Fact: If $v_1, v_2, \ldots, v_k$ are eigenvectors of A with *distinct* eigenvalues $\lambda_1, \ldots, \lambda_k$, then $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Why? If $k = 2$, this says v_2 can't lie on the line through v_1.

But the line through v_1 is contained in the λ_1-eigenspace, and v_2 does not have eigenvalue λ_1.

In general: see Lay, Theorem 2 in §5.1 (or work it out for yourself; it's not too hard).

Consequence: An $n \times n$ matrix has at most n distinct eigenvalues.

Let A be an $n \times n$ matrix. Suppose we want to solve $Av_n = v_{n+1}$ for all n. In other words, we want vectors $v_0, v_1, v_2, \ldots$, such that

$$Av_0 = v_1 \qquad Av_1 = v_2 \qquad Av_2 = v_3 \qquad \ldots$$

We saw before that $v_n = A^n v_0$. But it is inefficient to multiply by A each time.

If v_0 is an *eigenvector* with eigenvalue λ, then

$$v_1 = Av_0 = \lambda v_0 \qquad v_2 = Av_1 = \lambda v_1 = \lambda^2 v_0 \qquad v_3 = Av_2 = \lambda v_2 = \lambda^3 v_0.$$

In general, $v_n = \lambda^n v_0$. This is *much easier* to compute.

Example

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \qquad v_0 = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix} \qquad Av_0 = 2v_0.$$

So if you start with 16 baby rabbits, 4 first-year rabbits, and 1 second-year rabbit, then the population will exactly double every year. In year n, you will have $2^n \cdot 16$ baby rabbits, $2^n \cdot 4$ first-year rabbits, and 2^n second-year rabbits.

Section 5.2

The Characteristic Equation

The Invertible Matrix Theorem

Addenda

We have a couple of new ways of saying "A is invertible" now:

The Invertible Matrix Theorem

Let A be a square $n \times n$ matrix, and let $T\colon \mathbf{R}^n \to \mathbf{R}^n$ be the linear transformation $T(x) = Ax$. The following statements are equivalent.

1. A is invertible.
    2. T is invertible.
    3. A is row equivalent to I_n.
    4. A has n pivots.
    5. $Ax = 0$ has only the trivial solution.
    6. The columns of A are linearly independent.
    7. T is one-to-one.
    8. $Ax = b$ is consistent for all b in $\mathbf{R}^n$.
    9. The columns of A span $\mathbf{R}^n$.
    10. T is onto.
    11. A has a left inverse (there exists B such that $BA = I_n$).
    12. A has a right inverse (there exists B such that $AB = I_n$).
    13. A^T is invertible.
    14. The columns of A form a basis for $\mathbf{R}^n$.
    15. $\operatorname{Col} A = \mathbf{R}^n$.
    16. $\dim \operatorname{Col} A = n$.
    17. $\operatorname{rank} A = n$.
    18. $\operatorname{Nul} A = \{0\}$.
    19. $\dim \operatorname{Nul} A = 0$.
19. The determinant of A is *not* equal to zero.
20. The number 0 is *not* an eigenvalue of A.

The Characteristic Polynomial

Let A be a square matrix.

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff Ax = \lambda x \text{ has a nontrivial solution} \\ &\iff (A - \lambda I)x = 0 \text{ has a nontrivial solution} \\ &\iff A - \lambda I \text{ is not invertible} \\ &\iff \det(A - \lambda I) = 0. \end{aligned}$$

This gives us a way to compute the eigenvalues of A.

Definition

Let A be a square matrix. The **characteristic polynomial** of A is

$$f(\lambda) = \det(A - \lambda I).$$

The **characteristic equation** of A is the equation

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

Important

The eigenvalues of A are the roots of the characteristic polynomial $f(\lambda) = \det(A - \lambda I)$.

The Characteristic Polynomial

Example

Question: What are the eigenvalues of

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}?$$

The Characteristic Polynomial

Example

Question: What is the characteristic polynomial of

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}?$$

What do you notice about $f(\lambda)$?

- The constant term is $\det(A)$, which is zero if and only if $\lambda = 0$ is a root.
- The linear term $-(a+d)$ is the negative of the sum of the diagonal entries of A.

Definition

The **trace** of a square matrix A is $\mathrm{Tr}(A) =$ sum of the diagonal entries of A.

Shortcut

The characteristic polynomial of a 2×2 matrix A is

$$f(\lambda) = \lambda^2 - \mathrm{Tr}(A)\,\lambda + \det(A).$$

The Characteristic Polynomial

Example

Question: What are the eigenvalues of the rabbit population matrix

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}?$$

Algebraic Multiplicity

Definition

The **(algebraic) multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial.

This is not a very interesting notion *yet*. It will become interesting when we also define *geometric* multiplicity later.

Example

In the rabbit population matrix, $f(\lambda) = -(\lambda - 2)(\lambda + 1)^2$, so the algebraic multiplicity of the eigenvalue 2 is 1, and the algebraic multiplicity of the eigenvalue -1 is 2.

Example

In the matrix $\begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$, $f(\lambda) = (\lambda - (3 - 2\sqrt{2}))(\lambda - (3 + 2\sqrt{2}))$, so the algebraic multiplicity of $3 + 2\sqrt{2}$ is 1, and the algebraic multiplicity of $3 - 2\sqrt{2}$ is 1.

The Characteristic Polynomial

Poll

Fact: If A is an $n \times n$ matrix, the characteristic polynomial

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

turns out to be a polynomial of degree n, and its roots are the eigenvalues of A:

$$f(\lambda) = (-1)^n \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0.$$

Similarity

Definition

Two $n \times n$ matrices A and B are **similar** if there is an invertible $n \times n$ matrix C such that

$$A = CBC^{-1}.$$

What does this mean?

A acts on the standard coordinates of x in the same way that B acts on the $\mathcal{B}$-coordinates of x: $B[x]_{\mathcal{B}} = [Ax]_{\mathcal{B}}$.

Similarity

Example

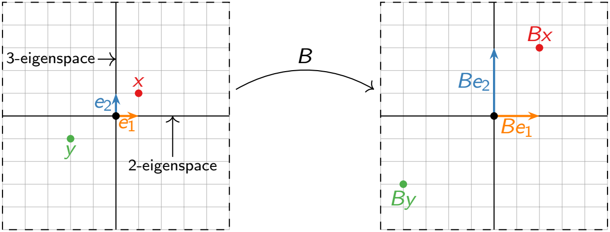
$$A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \quad C = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \implies \quad A = CBC^{-1}.$$

What does B do geometrically? It scales the x-direction by 2 and the y-direction by 3.

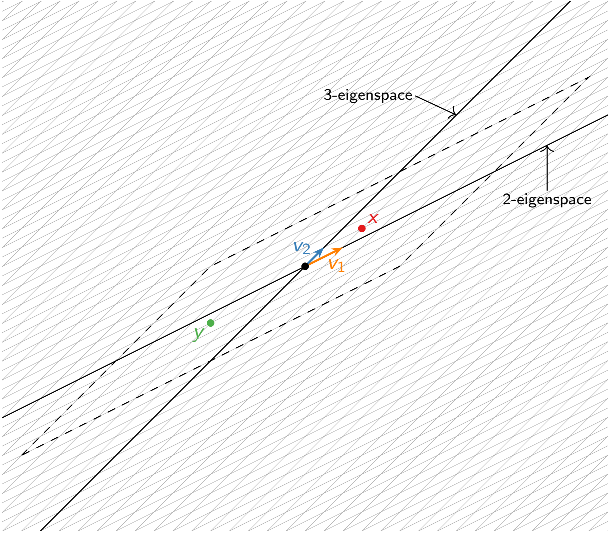
So A does to the standard coordinates what B does to the $\mathcal{B}$-coordinates, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

B acting on the usual coordinates



A does to the usual coordinates what B does to the $\mathcal{B}$-coordinates



$$\left.\begin{aligned} v_1 &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ v_2 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{aligned}\right\} \text{vectors in } \mathcal{B}$$

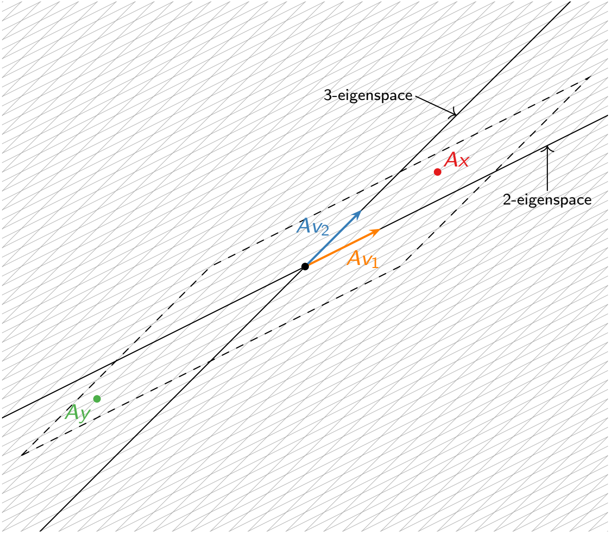
$$[x]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$x =$$

$$[y]_{\mathcal{B}} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

$$y =$$

A does to the usual coordinates what B does to the $\mathcal{B}$-coordinates



$Av_1 =$

$Av_2 =$

$B[x]_{\mathcal{B}} = \qquad = [Ax]_{\mathcal{B}}$

$Ax =$

$B[y]_{\mathcal{B}} = \qquad = [Ay]_{\mathcal{B}}$

$Ay =$

Check:

Similar Matrices Have the Same Characteristic Polynomial

Fact: If A and B are similar, then they have the same characteristic polynomial.

Why? Suppose $A = CBC^{-1}$.

Conseqence: similar matrices have the same eigenvalues!
(But different eigenvectors in general.)

Similarity

Caveats

Warning

1. Matrices with the same eigenvalues need not be similar. For instance,

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

   both only have the eigenvalue 2, but they are not similar.

2. Similarity has nothing to do with row equivalence. For instance,

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

   are row equivalent, but they have different eigenvalues.

Section 5.3

Diagonalization

Motivation

Difference equations

Many real-word linear algebra problems have the form:

$$v_1 = Av_0, \quad v_2 = Av_1 = A^2 v_0, \quad v_3 = Av_2 = A^3 v_0, \quad \ldots \quad v_n = Av_{n-1} = A^n v_0.$$

This is called a **difference equation**.

Our toy example about rabbit populations had this form.

The question is, what happens to v_n as $n \to \infty$?

- Taking powers of diagonal matrices is easy!
- Taking powers of *diagonalizable* matrices is still easy!
- Diagonalizing a matrix is an eigenvalue problem.

Powers of Diagonal Matrices

If D is diagonal, then D^n is also diagonal; its diagonal entries are the nth powers of the diagonal entries of D:

Powers of Matrices that are Similar to Diagonal Ones

What if A is not diagonal?

Example

Let $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$. Compute A^n.

In §5.2 lecture we saw that A is similar to a diagonal matrix:

$$A = PDP^{-1} \quad \text{where} \quad P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Then

$$A^2 =$$

$$A^3 =$$

$$\vdots$$

$$A^n =$$

Therefore

$$A^n =$$

Diagonalizable Matrices

Definition

An $n \times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix:

$$A = PDP^{-1} \quad \text{for } D \text{ diagonal.}$$

Important

$$\text{If } A = PDP^{-1} \text{ for } D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix} \text{ then}$$

$$A^k = PD^K P^{-1} = P \begin{pmatrix} d_{11}^k & 0 & \cdots & 0 \\ 0 & d_{22}^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^k \end{pmatrix} P^{-1}.$$

So diagonalizable matrices are easy to raise to any power.

Diagonalization

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In this case, $A = PDP^{-1}$ for

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \qquad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors, and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues (in the same order).

Corollary ← a theorem that follows easily from another theorem

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

The Corollary is true because eigenvectors with distinct eigenvalues are always linearly independent. We will see later that a diagonalizable matrix need not have n distinct eigenvalues though.

Diagonalization

Example

Problem: Diagonalize $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$.

Diagonalization

Another example

Problem: Diagonalize $A = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$.

Diagonalization

Another example, continued

Problem: Diagonalize $A = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$.

Note: In this case, there are three linearly independent eigenvectors, but only two distinct eigenvalues.

Diagonalization

A non-diagonalizable matrix

Problem: Show that $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.

Conclusion: A has only one linearly independent eigenvector, so by the "only if" part of the diagonalization theorem, A is not diagonalizable.

Poll

Diagonalization

Procedure

How to diagonalize a matrix A:

1. Find the eigenvalues of A using the characteristic polynomial.
2. For each eigenvalue λ of A, compute a basis $\mathcal{B}_\lambda$ for the λ-eigenspace.
3. If there are fewer than n total vectors in the union of all of the eigenspace bases $\mathcal{B}_\lambda$, then the matrix is not diagonalizable.
4. Otherwise, the n vectors $v_1, v_2, \ldots, v_n$ in your eigenspace bases are linearly independent, and $A = PDP^{-1}$ for

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where λ_i is the eigenvalue for v_i.

Diagonalization

Proof

Why is the Diagonalization Theorem true?

Non-Distinct Eigenvalues

Definition

Let λ be an eigenvalue of a square matrix A. The **geometric multiplicity** of λ is the dimension of the λ-eigenspace.

Theorem

Let λ be an eigenvalue of a square matrix A. Then

$$1 \leq (\text{the geometric multiplicity of } \lambda) \leq (\text{the algebraic multiplicity of } \lambda).$$

The proof is beyond the scope of this course.

Corollary

Let λ be an eigenvalue of a square matrix A. If the algebraic multiplicity of λ is 1, then the geometric multiplicity is also 1.

The Diagonalization Theorem (Alternate Form)

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

1. A is diagonalizable.
2. The sum of the geometric multiplicities of the eigenvalues of A equals n.
3. The sum of the algebraic multiplicities of the eigenvalues of A equals n, and *the geometric multiplicity equals the algebraic multiplicity* of each eigenvalue.

Non-Distinct Eigenvalues

Examples

Example

If A has n distinct eigenvalues, then the algebraic multiplicity of each equals 1, hence so does the geometric multiplicity, and therefore A is diagonalizable.

For example, $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$ has eigenvalues 2 and 3, so it is diagonalizable.

Example

The matrix $A = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$ has characteristic polynomial

$$f(\lambda) = -(\lambda - 1)^2(\lambda - 2).$$

The algebraic multiplicities of 1 and 2 are 2 and 1, respectively. They sum to 3.

We showed before that the geometric multiplicity of 1 is 2 (the 1-eigenspace has dimension 2). The eigenvalue 2 automatically has geometric multiplicity 1.

Hence the geometric multiplicities add up to 3, so A is diagonalizable.

Non-Distinct Eigenvalues

Another example

Example

The matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has characteristic polynomial $f(\lambda) = (\lambda - 1)^2$.

It has one eigenvalue 1 of algebraic multiplicity 2.

We showed before that the geometric multiplicity of 1 is 1 (the 1-eigenspace has dimension 1).

Since the geometric multiplicity is smaller than the algebraic multiplicity, the matrix is *not* diagonalizable.

Applications to Difference Equations

Let $D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$.

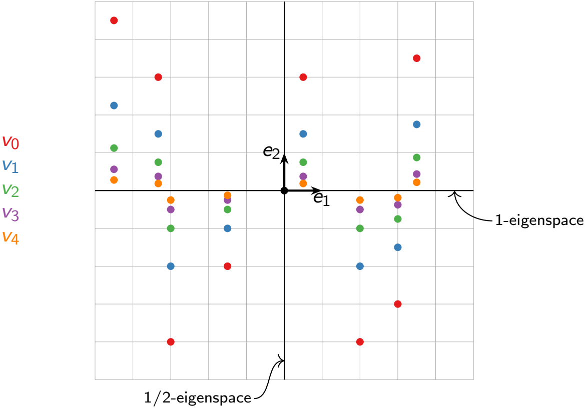
Fix a vector v_0, and let $v_1 = Dv_0$, $v_2 = Dv_1$, etc., so $v_n = D^n v_0$.

Question: What happens to the v_i's for different choices of v_0?

Applications to Difference Equations

Picture

$$D\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b/2 \end{pmatrix}$$



So all vectors get “sucked into the x-axis,” which is the 1-eigenspace.

Applications to Difference Equations

More complicated example

Let $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$.

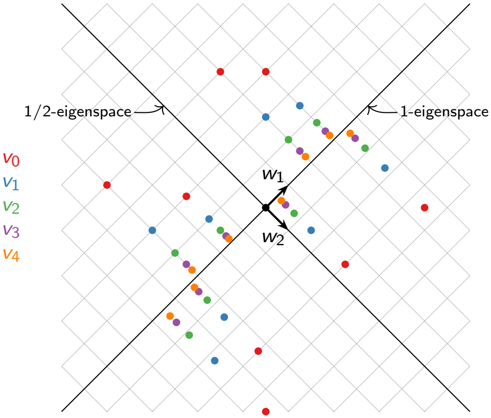
Fix a vector v_0, and let $v_1 = Av_0$, $v_2 = Av_1$, etc., so $v_n = A^n v_0$.

Question: What happens to the v_i's for different choices of v_0?

Applications to Difference Equations

Picture of the more complicated example

Recall: $A^n = PD^nP^{-1}$ acts on the usual coordinates of v_0 in the same way that D^n acts on the $\mathcal{B}$-coordinates, where $\mathcal{B} = \{w_1, w_2\}$.



So all vectors get "sucked into the 1-eigenspace."

Applications to Difference Equations

Remark

The matrix $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$ is called a **stochastic matrix**.

We will study such matrices in detail next time.

Application

Stochastic Matrices and PageRank

Stochastic Matrices

Definition

A square matrix A is **stochastic** if all of its entries are nonnegative, and the sum of the entries of each column is 1.

We say A is **positive** if all of its entries are positive.

These arise very commonly in modeling of probabalistic phenomena (Markov chains).

You'll be responsible for knowing basic facts about stochastic matrices and the Perron–Frobenius theorem, but we will not cover them in depth. These slides are the primary reference; see also §4.9 in Lay.

The specifics of the PageRank algorithm are just for fun.

Stochastic Matrices

Example

Red Box has kiosks all over where you can rent movies. You can return them to any other kiosk. Let A be the matrix whose ij entry is the probability that a customer renting a movie from location j returns it to location i. For example, if there are three locations, maybe

$$A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}.$$

30% probability a movie rented from location 3 gets returned to location 2

The columns sum to 1 because there is a 100% chance that the movie will get returned to *some* location. This is a positive stochastic matrix.

Note that, if $v = (x, y, z)$ represents the number of movies at the three locations, then (assuming the number of movies is large), Red Box will have approximately

$$Av = A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .3x + .4y + .5z \\ \boxed{.3x + .4y + .3z} \\ .4x + .2y + .2z \end{pmatrix}$$

"The number of movies returned to location 2 will be (on average):
30% of the movies from location 1;
40% of the movies from location 2;
30% of the movies from location 3"

movies in its three locations the next day. The *total number* of movies doesn't change because the columns sum to 1.

Stochastic Matrices and Difference Equations

If x_n, y_n, z_n are the numbers of movies in locations $1, 2, 3$, respectively, on day n, and $v_n = (x_n, y_n, z_n)$, then:

$$v_n = Av_{n-1} = A^2 v_{n-2} = \cdots = A^n v_0.$$

Recall: This is an example of a **difference equation**.

Red Box probably cares about what v_n is as n gets large: it tells them where the movies will end up *eventually*. This seems to involve computing A^n for large n, but as we will see, they actually only have to compute one eigenvector.

In general: A difference equation $v_{n+1} = Av_n$ is used to model a state change controlled by a matrix:

- v_n is the "state at time n",
- v_{n+1} is the "state at time $n + 1$", and
- $v_{n+1} = Av_n$ means that A is the "change of state matrix."

Eigenvalues of Stochastic Matrices

Fact: 1 is an eigenvalue of a stochastic matrix.

Why? If A is stochastic, then 1 is an eigenvalue of A^T:

$$\begin{pmatrix} .3 & .3 & .4 \\ .4 & .4 & .2 \\ .5 & .3 & .2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Lemma

A and A^T have the same eigenvalues.

Proof: $\det(A - \lambda I) = \det\big((A - \lambda I)^T\big) = \det(A^T - \lambda I)$, so they have the same characteristic polynomial.

Note: This doesn't give a new procedure for finding an eigenvector with eigenvalue 1; it only shows one exists.

Eigenvalues of Stochastic Matrices

Continued

Fact: if λ is an eigenvalue of a stochastic matrix, then $|\lambda| \leq 1$. Hence 1 is the *largest* eigenvalue (in absolute value).

Better fact: if $\lambda \neq 1$ is an eigenvalue of a *positive* stochastic matrix, then $|\lambda| < 1$.

Diagonalizable Stochastic Matrices

Example from §5.3

Let $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$. This is a positive stochastic matrix.

We saw last time that A is diagonalizable (and 1 is the largest eigenvalue):

$$A = PDP^{-1} \quad \text{for} \quad P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Let $w_1 = \binom{1}{1}$ and $w_2 = \binom{1}{-1}$ be the columns of P, and let $\mathcal{B} = \{w_1, w_2\}$.

Recall: A^n acts on the usual coordinates of a vector in the same way that D acts on the $\mathcal{B}$-coordinates: $[A^n x]_{\mathcal{B}} = D^n[x]_{\mathcal{B}}$.

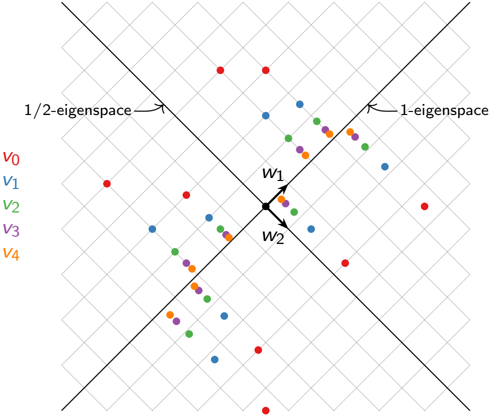
When n is large, the second term disappears, so $A^n x$ approaches $c_1 w_1$, which is an *eigenvector with eigenvalue* 1 (assuming $c_1 \neq 0$).

So all vectors get "sucked into the 1-eigenspace," which is spanned by $w_1 = \binom{1}{1}$.

Diagonalizable Stochastic Matrices

Example, continued

Recall: $A^n = PD^nP^{-1}$ acts on the usual coordinates of v_0 in the same way that D^n acts on the $\mathcal{B}$-coordinates, where $\mathcal{B} = \{w_1, w_2\}$.



All vectors get "sucked into the 1-eigenspace."

Diagonalizable Stochastic Matrices

The Red Box matrix $A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$ has characteristic polynomial

$$f(\lambda) = -\lambda^3 + 0.12\lambda - 0.02 = -(\lambda - 1)(\lambda + 0.2)(\lambda - 0.1).$$

So 1 is indeed the largest eigenvalue. Since A has 3 distinct eigenvalues, it is diagonalizable:

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & -.2 \end{pmatrix} P^{-1} = PDP^{-1}.$$

Hence it is easy to compute the powers of A:

$$A^n = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & (.1)^n & 0 \\ 0 & 0 & (-.2)^n \end{pmatrix} P^{-1} = PD^n P^{-1}.$$

Let w_1, w_2, w_3 be the columns of P, i.e. the eigenvectors of P with respective eigenvalues $1, .1, -.2$. Let $\mathcal{B} = \{w_1, w_2, w_3\}$.

Recall: A^n acts on the usual coordinates of a vector in the same way that D acts on the $\mathcal{B}$-coordinates: $[A^n x]_{\mathcal{B}} = D^n [x]_{\mathcal{B}}$.

Diagonalizable Stochastic Matrices

Continued

Recall: A^n acts on the usual coordinates of a vector in the same way that D acts on the $\mathcal{B}$-coordinates: $[A^n x]_{\mathcal{B}} = D^n [x]_{\mathcal{B}}$.

As n becomes large, this approaches $c_1 w_1$, which is an *eigenvector with eigenvalue* 1 (assuming $c_1 \neq 0$).

So all vectors get "sucked into the 1-eigenspace," which (I computed) is spanned by

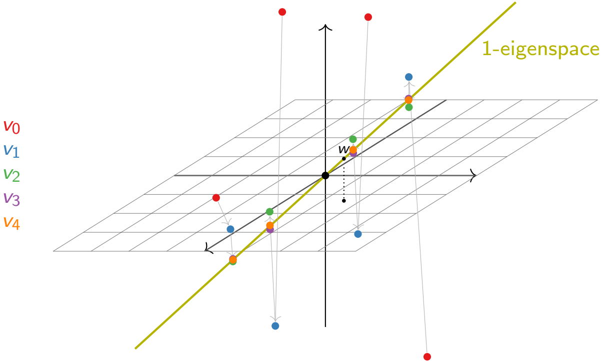
$$w = w_1 = \frac{1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}.$$

(We'll see in a moment why I chose that eigenvector.)

Diagonalizable Stochastic Matrices

Picture

Start with a vector v_0 (the number of movies on the first day), let $v_1 = Av_0$ (the number of movies on the second day), let $v_2 = Av_1$, etc.



We see that v_n approaches an eigenvector with eigenvalue 1 as n gets large: all vectors get "sucked into the 1-eigenspace."

Diagonalizable Stochastic Matrices

Interpretation

If A is the Red Box matrix, and v_n is the vector representing the number of movies in the three locations on day n, then

$$v_{n+1} = Av_n.$$

For any starting distribution v_0 of videos in red boxes, after enough days, the distribution v ($= v_n$ for n large) is an eigenvector with eigenvalue 1:

$$Av = v.$$

In other words, eventually each kiosk has the same number of movies, every day.

Moreover, we know exactly what v is: it is the multiple of $w \sim (0.39, 0.33, 0.28)$ that represents the same number of videos as in v_0. (Remember the total number of videos never changes.)

Presumably, Red Box really does have to do this kind of analysis to determine how many videos to put in each box.

Perron–Frobenius Theorem

Definition

A *steady state* for a stochastic matrix A is an eigenvector w with eigenvalue 1, such that all entries are *positive* and sum to 1.

Perron–Frobenius Theorem

If A is a positive stochastic matrix, then it admits a unique steady state vector w, which spans the 1-eigenspace.

Moreover, for any vector v_0 with entries summing to some number c, the iterates $v_1 = Av_0$, $v_2 = Av_1$, . . . , $v_n = Av_{n-1}$, . . . , approach cw as n gets large.

Translation: The Perron–Frobenius Theorem says the following:

- The 1-eigenspace of a positive stochastic matrix A is a line.
- To compute the steady state, find any 1-eigenvector (as usual), then divide by the sum of the entries; the resulting vector w has entries that sum to 1, and are *automatically* positive.
- Think of w as a vector of steady state *percentages*: if the movies are distributed according to these percentages today, then they'll be in the same distribution tomorrow.
- The sum c of the entries of v_0 is the total number of movies; eventually, the movies arrange themselves according to the steady state percentage, i.e., $v_n \to cw$.

Steady State

Red Box example

Consider the Red Box matrix $A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$.

This says that eventually, 39% of the movies will be in location 1, 33% will be in location 2, and 28% will be in location 3, every day.

So if you start with 100 total movies, eventually you'll have
$100w = (39, 33, 28)$ movies in the three locations, every day.

The Perron–Frobenius Theorem says that our analysis of the Red Box matrix works for *any* positive stochastic matrix—whether or not it is diagonalizable!

Google's PageRank

Internet searching in the 90's was a pain. Yahoo or AltaVista would scan pages for your search text, and just list the results with the most occurrences of those words.

Not surprisingly, the more unsavory websites soon learned that by putting the words "Alanis Morissette" a million times in their pages, they could show up first every time an angsty teenager tried to find *Jagged Little Pill* on Napster.

Larry Page and Sergey Brin invented a way to rank pages by *importance*. They founded Google based on their algorithm.

Here's how it works. (roughly)

Reference:

`http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html`

The Importance Rule

Each webpage has an associated importance, or **rank**. This is a positive number.

The Importance Rule

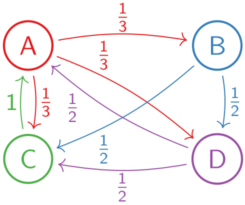
If page P links to n other pages $Q_1, Q_2, \ldots, Q_n$, then each Q_i should inherit $\frac{1}{n}$ of P's importance.

- So if a very important page links to your webpage, your webpage is considered important.
- And if a ton of unimportant pages link to your webpage, then it's still important.
- But if only one crappy site links to yours, your page isn't important.

Random surfer interpretation: a "random surfer" just sits at his computer all day, randomly clicking on links. The pages he spends the most time on should be the most important. This turns out to be equivalent to the rank.

The Importance Matrix

Consider the following Internet with only four pages. Links are indicated by arrows.



Page A has 3 links, so it passes $\frac{1}{3}$ of its importance to pages B, C, D.
Page B has 2 links, so it passes $\frac{1}{2}$ of its importance to pages C, D.
Page C has one link, so it passes all of its importance to page A.
Page D has 2 links, so it passes $\frac{1}{2}$ of its importance to pages A, C.

In terms of matrices, if $v = (a, b, c, d)$ is the vector containing the ranks a, b, c, d of the pages A, B, C, D, then

importance matrix: ij entry is importance page j passes to page i

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} \overset{\text{Importance Rule}}{=} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

The 25 Billion Dollar Eigenvector

Observations:

- The importance matrix is a stochastic matrix! The columns each contain $1/n$ ($n =$ number of links), n times.
- The rank vector is an eigenvector with eigenvalue 1!

Random surfer interpretation: If a random surfer has probability (a, b, c, d) to be on page A, B, C, D, respectively, then after clicking on a random link, the probability he'll be on each page is

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix}.$$

The rank vector is a *steady state* for the importance matrix: it's the probability vector (a, b, c, d) such that, after clicking on a random link, the random surfer will have the *same probability* of being on each page.

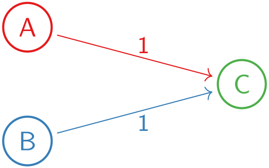
So, the important (high-ranked) pages are those where a random surfer will end up most often.

Problems with the Importance Matrix

Dangling pages

Observation: the importance matrix is *not* positive: it's only nonnegative. So we can't apply the Perron–Frobenius theorem. Does this cause problems? Yes!

Consider the following Internet:



The importance matrix is $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$. This has characteristic polynomial

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 1 & 1 & -\lambda \end{pmatrix} = -\lambda^3.$$

So 1 is not an eigenvalue at all: there is no rank vector! (It is not stochastic.)

Problems with the Importance Matrix

Disconnected internet

Consider the following Internet:

The importance matrix is $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$. This has linearly independent eigenvectors $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, both with eigenvalue 1. So there is more than one rank vector!

The Google Matrix

Here is Page and Brin's solution. Fix p in $(0, 1)$, called the **damping factor**. (A typical value is $p = 0.15$.) The **Google Matrix** is

$$M = (1 - p) \cdot A + p \cdot B \quad \text{where} \quad B = \frac{1}{N} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

N is the total number of pages, and A is the importance matrix.

In the random surfer interpretation, this matrix M says: with probability p, our surfer will surf to a completely random page; otherwise, he'll click a random link.

Lemma

The Google matrix is a positive stochastic matrix.

The PageRank vector is the steady state for the Google Matrix.

This exists and has positive entries by the Perron–Frobenius theorem. The hard part is calculating it: the Google matrix has 1 gazillion rows.

Section 5.5

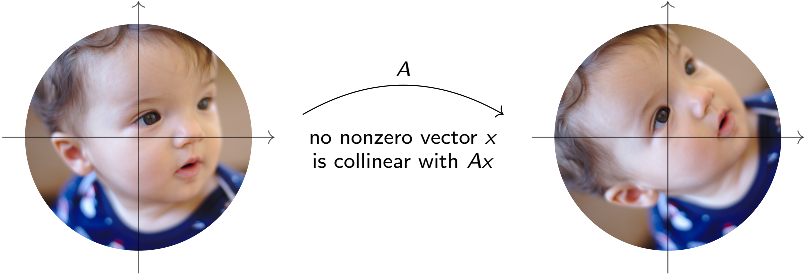
Complex Eigenvalues

A Matrix with No Eigenvectors

In recitation you discussed the linear transformation for rotation by $\pi/4$ in the plane. The matrix is:

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

This matrix has no eigenvectors, as you can see geometrically:



no nonzero vector x
is collinear with Ax

or algebraically:

$$f(\lambda) = \lambda^2 - \mathrm{Tr}(A)\,\lambda + \det(A) = \lambda^2 - \sqrt{2}\,\lambda + 1 \implies \lambda = \frac{\sqrt{2} \pm \sqrt{-2}}{2}.$$

Complex Numbers

It makes us sad that -1 has no square root. If it did, then $\sqrt{-2} = \sqrt{2} \cdot \sqrt{-1}$.

Mathematician's solution: we're just not using enough numbers! We're going to declare by *fiat* that there exists a square root of -1.

Definition

The number i is defined such that $i^2 = -1$.

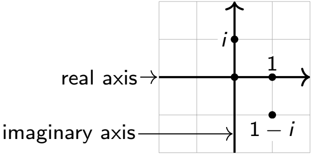
Once we have i, we have to allow numbers like $a + bi$ for real numbers a, b.

Definition

A *complex number* is a number of the form $a + bi$ for a, b in $\mathbf{R}$. The set of all complex numbers is denoted $\mathbf{C}$.

Note $\mathbf{R}$ is contained in $\mathbf{C}$: they're the numbers $a + 0i$.

We can identify $\mathbf{C}$ with $\mathbf{R}^2$ by $a + bi \longleftrightarrow \binom{a}{b}$. So when we draw a picture of $\mathbf{C}$, we draw the plane:



Why This Is Not A Weird Thing To Do

An anachronistic historical aside

In the beginning, people only used counting numbers for, well, counting things: $1, 2, 3, 4, 5, \ldots$. Then someone (Persian mathematician Muḥammad ibn Mūsā al-Khwārizmī, 825) had the ridiculous idea that there should be a number 0 that represents an absence of quantity. This blew everyone's mind.

Then it occurred to someone (Chinese mathematician Liu Hui, c. 3rd century) that there should be *negative* numbers to represent a deficit in quantity. That seemed reasonable, until people realized that $10 + (-3)$ would have to equal 7. This is when people started saying, "bah, math is just too hard for me."

At this point it was inconvenient that you couldn't divide 2 by 3. Thus someone (Indian mathematician Aryabhatta, c. 5th century) invented fractions (rational numbers) to represent fractional quantities. These proved very popular. The Pythagoreans developed a whole belief system around the notion that any quantity worth considering could be broken down into whole numbers in this way.

Then the Pythagoreans (c. 6th century BCE) discovered that the hypotenuse of an isosceles right triangle with side length 1 (i.e. $\sqrt{2}$) is not a fraction. This caused a serious existential crisis and led to at least one death by drowning. The real number $\sqrt{2}$ was thus invented to solve the equation $x^2 - 2 = 0$.

So what's so strange about inventing a number i to solve the equation $x^2 + 1 = 0$?

Operations on Complex Numbers

Addition:

Multiplication:

Complex conjugation: $\overline{a+bi} = a - bi$ is the **complex conjugate** of $a + bi$.
Check: $\overline{z+w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z} \cdot \overline{w}$.

Absolute value: $|a+bi| = \sqrt{a^2+b^2}$. This is a *real* number.
Note: $(a+bi)(\overline{a+bi}) = (a+bi)(a-bi) = a^2 - (bi)^2 = a^2 + b^2$. So $|z| = \sqrt{z\overline{z}}$.
Check: $|zw| = |z| \cdot |w|$.

Division by a nonzero real number: $\dfrac{a+bi}{c} = \dfrac{a}{c} + \dfrac{b}{c}i$.

Division by a nonzero complex number: $\dfrac{z}{w} = \dfrac{z\overline{w}}{w\overline{w}} = \dfrac{z\overline{w}}{|w|^2}$.

Example:

$$\frac{1+i}{1-i} =$$

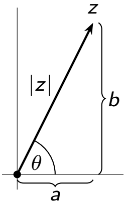
Real and imaginary part: $\mathrm{Re}(a+bi) = a \qquad \mathrm{Im}(a+bi) = b$.

Polar Coordinates for Complex Numbers

Any complex number $z = a + bi$ has the polar coordinates

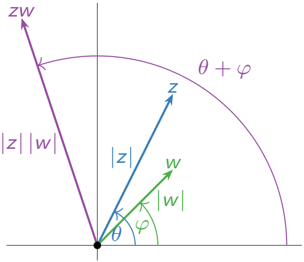
$$z = |z|(\cos\theta + i\sin\theta).$$

The angle θ is called the **argument** of z, and is denoted $\theta = \arg(z)$. Note $\arg(\overline{z}) = -\arg(z)$.



When you multiply complex numbers, you multiply the absolute values and add the arguments:

$$|zw| = |z|\,|w| \qquad \arg(zw) = \arg(z) + \arg(w).$$



The Fundamental Theorem of Algebra

The whole point of using complex numbers is to solve polynomial equations. It turns out that they are enough to find all solutions of all polynomial equations:

Fundamental Theorem of Algebra

Every polynomial of degree n has exactly n complex roots, counted with multiplicity.

Equivalently, if $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ is a polynomial of degree n, then

$$f(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_n)$$

for (not necessarily distinct) complex numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Important

If f is a polynomial with *real* coefficients, and if λ is a root of f, then so is $\overline{\lambda}$:

$$\begin{aligned} 0 = \overline{f(\lambda)} &= \overline{\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0} \\ &= \overline{\lambda}^n + a_{n-1}\overline{\lambda}^{n-1} + \cdots + a_1\overline{\lambda} + a_0 = f\big(\overline{\lambda}\big). \end{aligned}$$

Therefore complex roots of real polynomials come in *conjugate pairs*.

The Fundamental Theorem of Algebra

Examples

Degree 2: The quadratic formula gives you the (real or complex) roots of any degree-2 polynomial:

$$f(x) = x^2 + bx + c \implies x = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

For instance, if $f(\lambda) = \lambda^2 - \sqrt{2}\lambda + 1$ then

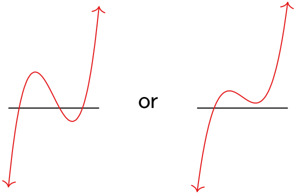
$$\lambda =$$

Note the roots are complex conjugates if b, c are real.

The Fundamental Theorem of Algebra

Examples

Degree 3: A real cubic polynomial has either three real roots, or one real root and a conjugate pair of complex roots. The graph looks like:



respectively.

Example: let $f(\lambda) = 5\lambda^3 - 18\lambda^2 + 21\lambda - 10$.

Poll

The characteristic polynomial of

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

is $f(\lambda) = \lambda^2 - \sqrt{2}\lambda + 1$. This has two complex roots $(1 \pm i)/\sqrt{2}$.

A Matrix *with* an Eigenvector

Every matrix is guaranteed to have *complex* eigenvalues and eigenvectors. Using rotation by $\pi/4$ from before:

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{has eigenvalues} \quad \lambda = \frac{1 \pm i}{\sqrt{2}}.$$

Let's compute an eigenvector for $\lambda = (1+i)/\sqrt{2}$:

A similar computation shows that an eigenvector for $\lambda = (1-i)/\sqrt{2}$ is $\begin{pmatrix} -i \\ 1 \end{pmatrix}$.

So is $i\begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ i \end{pmatrix}$ (you can scale by *complex* numbers).

A Trick for Computing Eigenvectors of 2×2 Matrices

Very useful for complex eigenvalues

Let A be a 2×2 matrix, and let λ be an eigenvalue of A.

Then $A - \lambda I$ is not invertible, so the second row is *automatically* a multiple of the first. (Think about it for a while: otherwise the rref is $\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$.)

Hence the second row disappears in the rref, so *we don't care what it is!*

If $A - \lambda I = \begin{pmatrix} a & b \\ \star & \star \end{pmatrix}$, then $(A - \lambda I)\begin{pmatrix} b \\ -a \end{pmatrix} = 0$, so $\begin{pmatrix} b \\ -a \end{pmatrix}$ is an eigenvector.

So is $\begin{pmatrix} -b \\ a \end{pmatrix}$.

Example:

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \lambda = \frac{1-i}{\sqrt{2}}.$$

Conjugate Eigenvectors

For $A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$,

$$\text{the eigenvalue } \frac{1+i}{\sqrt{2}} \text{ has eigenvector } \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

$$\text{the eigenvalue } \frac{1-i}{\sqrt{2}} \text{ has eigenvector } \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Do you notice a pattern?

Fact

Let A be a real square matrix. If λ is an eigenvalue with eigenvector v, then $\overline{\lambda}$ is an eigenvalue with eigenvector $\overline{v}$.

Why?

$$Av = \lambda \implies A\overline{v} = \overline{Av} = \overline{\lambda v} = \overline{\lambda}\overline{v}.$$

Both eigenvalues and eigenvectors of real square matrices occur in conjugate pairs.

A 3×3 Example

Find the eigenvalues and eigenvectors of

$$A = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The characteristic polynomial is

We computed the roots of this polynomial (times 5) before:

$$\lambda = \quad 2, \quad \frac{4+3i}{5}, \quad \frac{4-3i}{5}.$$

We eyeball an eigenvector with eigenvalue 2 as $(0, 0, 1)$.

A 3×3 Example

Continued

$$A = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

To find the other eigenvectors, we row reduce:

Geometric Interpretation of Complex Eigenvectors

2×2 case

Theorem

Let A be a 2×2 matrix with complex (non-real) eigenvalue λ, and let v be an eigenvector. Then

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

where

$$P = \begin{pmatrix} | & | \\ \operatorname{Re} v & \operatorname{Im} v \\ | & | \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}.$$

The matrix C is a composition of rotation by $-\arg(\lambda)$ and scaling by $|\lambda|$:

$$C = \begin{pmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{pmatrix} \begin{pmatrix} \cos(-\arg(\lambda)) & -\sin(-\arg(\lambda)) \\ \sin(-\arg(\lambda)) & \cos(-\arg(\lambda)) \end{pmatrix}.$$

A 2×2 matrix with complex eigenvalue λ is similar to (rotation by the argument of $\overline{\lambda}$) composed with (scaling by $|\lambda|$). This is multiplication by $\overline{\lambda}$ in $\mathbf{C} \sim \mathbf{R}^2$.

Geometric Interpretation of Complex Eigenvalues

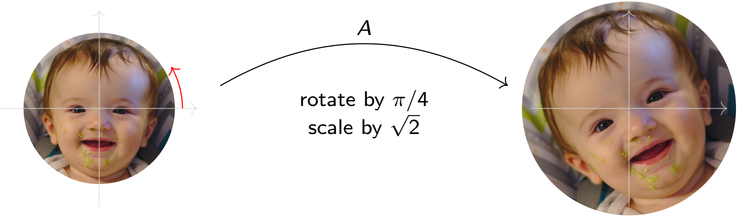
2 × 2 example

What does $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ do geometrically?

Geometric Interpretation of Complex Eigenvalues

2×2 example, continued

$$A = C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \lambda = 1 - i$$



A

rotate by $\pi/4$
scale by $\sqrt{2}$

Geometric Interpretation of Complex Eigenvalues

Another 2 × 2 example

What does $A = \begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix}$ do geometrically?

Geometric Interpretation of Complex Eigenvalues

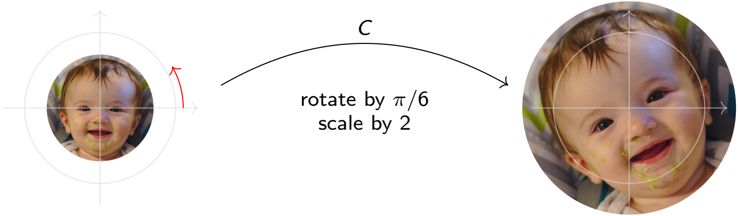
Another 2 × 2 example, continued

$$A = \begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix} \qquad C = \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \qquad \lambda = \sqrt{3} - i$$

Geometric Interpretation of Complex Eigenvalues

Another 2 × 2 example: picture

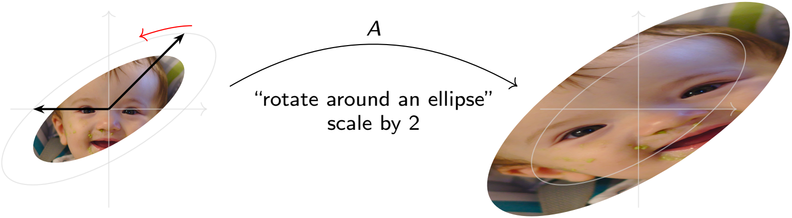
What does $A = \begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix}$ do geometrically?



C

rotate by $\pi/6$
scale by 2

$A = PCP^{-1}$ does the same thing, but with respect to the basis $\mathcal{B} = \left\{ \binom{1}{1}, \binom{-1}{0} \right\}$ of columns of P:



A

"rotate around an ellipse"
scale by 2

Classification of 2×2 Matrices with a Complex Eigenvalue

Triptych

Let A be a real matrix with a complex eigenvalue λ. One way to understand the geometry of A is to consider the difference equation $v_{n+1} = Av_n$, i.e. the sequence of vectors $v, Av, A^2v, \ldots$

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix}$$

$$\lambda = \frac{\sqrt{3}-i}{\sqrt{2}}$$

$|\lambda| > 1$

$$A = \frac{1}{2}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix}$$

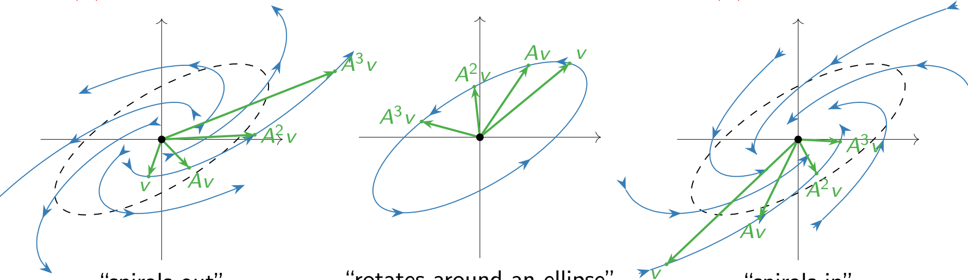
$$\lambda = \frac{\sqrt{3}-i}{2}$$

$|\lambda| = 1$

$$A = \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{3}+1 & -2 \\ 1 & \sqrt{3}-1 \end{pmatrix}$$

$$\lambda = \frac{\sqrt{3}-i}{2\sqrt{2}}$$

$|\lambda| < 1$



“spirals out”

“rotates around an ellipse”

“spirals in”

Complex Versus Two Real Eigenvalues

An analogy

Theorem

Let A be a 2×2 matrix with complex eigenvalue $\lambda = a + bi$ (where $b \neq 0$), and let v be an eigenvector. Then

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

where

$$P = \begin{pmatrix} | & | \\ \operatorname{Re} v & \operatorname{Im} v \\ | & | \end{pmatrix} \quad \text{and} \quad C = (\text{rotation}) \cdot (\text{scaling}).$$

This is very analogous to diagonalization. In the 2×2 case:

Theorem

Let A be a 2×2 matrix with linearly independent eigenvectors v_1, v_2 and associated eigenvalues λ_1, λ_2. Then

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

where

$$P = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

scale x-axis by λ_1
scale y-axis by λ_2

Picture with 2 Real Eigenvalues

We can draw analogous pictures for a matrix with 2 real eigenvalues.

Example: Let $A = \frac{1}{4}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$.

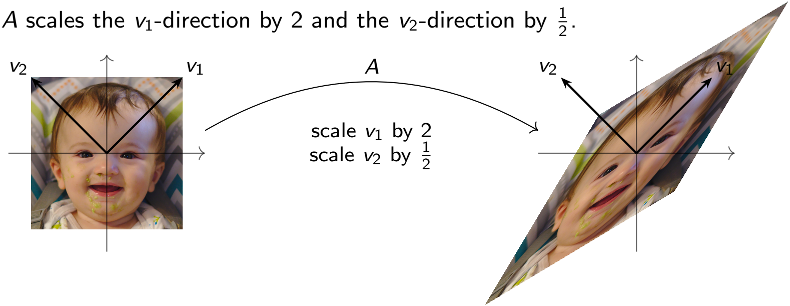
This has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = \frac{1}{2}$, with eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Therefore, $A = PDP^{-1}$ with

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

So A scales the v_1-direction by 2 and the v_2-direction by $\frac{1}{2}$.

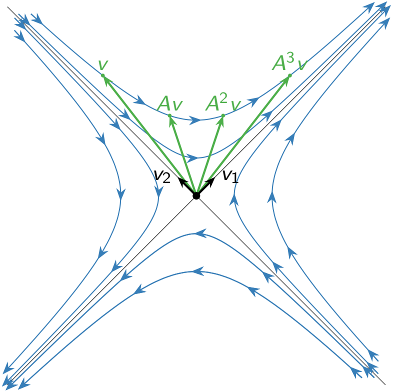


scale v_1 by 2
scale v_2 by $\frac{1}{2}$

Picture with 2 Real Eigenvalues

We can also draw a picture from the perspective a difference equation: in other words, we draw $v, Av, A^2v, \ldots$

$$A = \frac{1}{4}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \qquad \begin{array}{cc} \lambda_1 = 2 & \lambda_2 = \frac{1}{2} \\ |\lambda_1| > 1 & |\lambda_1| < 1 \end{array}$$



Exercise: Draw analogous pictures when $|\lambda_1|, |\lambda_2|$ are any combination of $< 1, = 1, > 1$.

The Higher-Dimensional Case

Theorem

Let A be a real $n \times n$ matrix. Suppose that for each (real or complex) eigenvalue, the dimension of the eigenspace equals the algebraic multiplicity. Then $A = PCP^{-1}$, where P and C are as follows:

1. C is **block diagonal**, where the blocks are 1×1 blocks containing the real eigenvalues (with their multiplicities), or 2×2 blocks containing the matrices $\begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}$ for each non-real eigenvalue λ (with multiplicity).
2. The columns of P form bases for the eigenspaces for the real eigenvectors, or come in pairs $(\operatorname{Re} v \;\; \operatorname{Im} v)$ for the non-real eigenvectors.

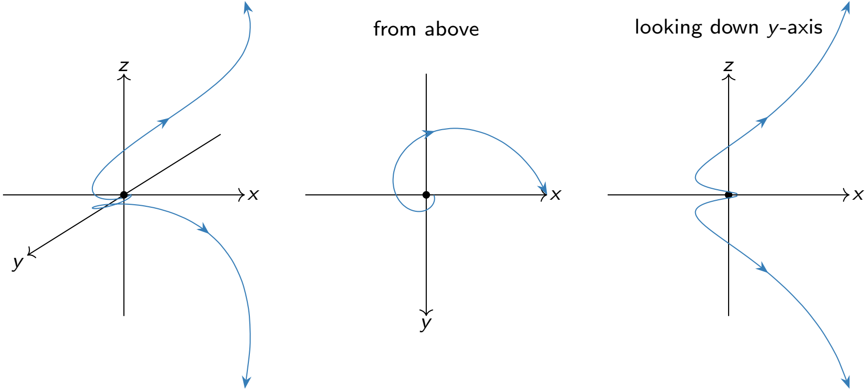
For instance, if A is a 3×3 matrix with one real eigenvalue λ_1 with eigenvector v_1, and one conjugate pair of complex eigenvalues $\lambda_2, \overline{\lambda}_2$ with eigenvectors $v_2, \overline{v}_2$, then

$$P = \begin{pmatrix} | & | & | \\ v_1 & \operatorname{Re} v_2 & \operatorname{Im} v_2 \\ | & | & | \end{pmatrix} \qquad C = \begin{pmatrix} \boxed{\lambda_1} & \begin{matrix} 0 & 0 \end{matrix} \\ \begin{matrix} 0 \\ 0 \end{matrix} & \boxed{\begin{matrix} \operatorname{Re}\lambda_2 & \operatorname{Im}\lambda_2 \\ -\operatorname{Im}\lambda_2 & \operatorname{Re}\lambda_2 \end{matrix}} \end{pmatrix}$$

The Higher-Dimensional Case

Example

Let $A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. This acts on the xy-plane by rotation by $\pi/4$ and scaling by $\sqrt{2}$. This acts on the z-axis by scaling by 2. Pictures:



Remember, in general $A = PCP^{-1}$ is only *similar* to such a matrix C: so the x, y, z axes have to be replaced by the columns of P.

Chapter 6

Orthogonality and Least Squares

Section 6.1

Inner Product, Length, and Orthogonality

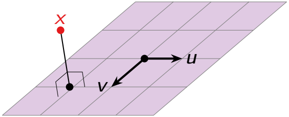
Orientation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$
- Solve the matrix equation $Ax = \lambda x$
- Almost solve the equation $Ax = b$

We are now aiming at the last topic.

Idea: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution. What do you do? The real value is probably the *closest* point to x on $\text{Span}\{u, v\}$. Which point is that?

The Dot Product

We need a notion of *angle* between two vectors, and in particular, a notion of *orthogonality* (i.e. when two vectors are perpendicular). This is the purpose of the dot product.

Definition

The **dot product** of two vectors x, y in $\mathbf{R}^n$ is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \overset{\text{def}}{=} x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Thinking of x, y as column vectors, this is the same as $x^T y$.

Example

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} =$$

Properties of the Dot Product

Many usual arithmetic rules hold, as long as you remember you can only dot two vectors together, and that the result is a scalar.

- $x \cdot y = y \cdot x$
- $(x + y) \cdot z = x \cdot z + y \cdot z$
- $(cx) \cdot y = c(x \cdot y)$

Dotting a vector with itself is special:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Hence:

- $x \cdot x \geq 0$
- $x \cdot x = 0$ if and only if $x = 0$.

Important: $x \cdot y = 0$ does *not* imply $x = 0$ or $y = 0$. For example, $\binom{1}{0} \cdot \binom{0}{1} = 0$.

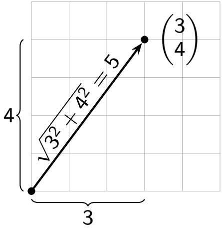
The Dot Product and Length

Definition

The **length** or **norm** of a vector x in $\mathbf{R}^n$ is

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Why is this a good definition? The Pythagorean theorem!



$$\left\|\begin{pmatrix}3\\4\end{pmatrix}\right\| = \sqrt{3^2+4^2} = 5$$

Fact

If x is a vector and c is a scalar, then $\|cx\| = |c| \cdot \|x\|$.

$$\left\|\begin{pmatrix}6\\8\end{pmatrix}\right\| = \left\|2\begin{pmatrix}3\\4\end{pmatrix}\right\| =$$

The Dot Product and Distance

Definition

The **distance** between two points x, y in $\mathbf{R}^n$ is

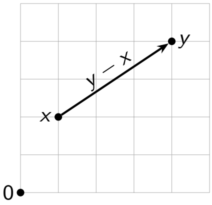
$$\mathrm{dist}(x, y) = \|y - x\|.$$

This is just the length of the vector from x to y.

Example

Let $x = (1, 2)$ and $y = (4, 4)$. Then

$$\mathrm{dist}(x, y) =$$



Unit Vectors

Definition

A **unit vector** is a vector v with length $\|v\| = 1$.

Example

The unit coordinate vectors are unit vectors:

$$\|e_1\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + 0^2} = 1$$

Definition

Let x be a nonzero vector in $\mathbf{R}^n$. The **unit vector in the direction of** x is the vector $\dfrac{x}{\|x\|}$.

This is in fact a unit vector:

$$\left\| \frac{x}{\underset{\text{scalar}}{\boxed{\|x\|}}} \right\| = \frac{1}{\|x\|}\|x\| = 1.$$

Unit Vectors

Example

Example

What is the unit vector in the direction of $x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$?

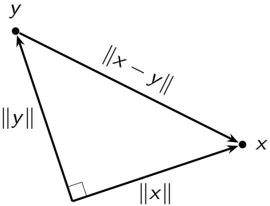
Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.

Notation: $x \perp y$ means $x \cdot y = 0$.

Why is this a good definition? The Pythagorean theorem / law of cosines!



Fact: $x \perp y \iff \|x - y\|^2 = \|x\|^2 + \|y\|^2$

Orthogonality

Example

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

Orthogonality

Example

Problem: Find *all* vectors orthogonal to both $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Orthogonality

General procedure

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \ldots, v_m$ in $\mathbf{R}^n$.

This is the same as finding all vectors x such that

$$0 = v_1^T x = v_2^T x = \cdots = v_m^T x.$$

Putting the *row* vectors $v_1^T, v_2^T, \ldots, v_m^T$ into a matrix, this is the same as finding all x such that

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix} x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix} = 0.$$

Important

The set of all vectors orthogonal to some vectors $v_1, v_2, \ldots, v_m$ in $\mathbf{R}^n$ is the *null space* of the $m \times n$ matrix

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}.$$

In particular, this set is a subspace!

Orthogonal Complements

Definition

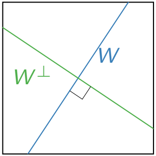
Let W be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is

$$W^{\perp} = \{ v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W \} \qquad \text{read "}W\text{ perp".}$$

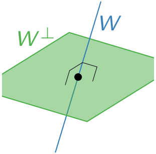
$W^{\perp}$ is orthogonal complement
A^{T} is transpose

Pictures:

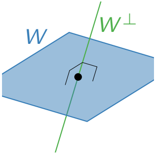
The orthogonal complement of a line in $\mathbf{R}^2$ is the perpendicular line.



The orthogonal complement of a line in $\mathbf{R}^3$ is the perpendicular plane.



The orthogonal complement of a plane in $\mathbf{R}^3$ is the perpendicular line.



Poll

Orthogonal Complements

Basic properties

Let W be a subspace of $\mathbf{R}^n$.

Facts:

1. $W^\perp$ is also a subspace of $\mathbf{R}^n$
2. $(W^\perp)^\perp = W$
3. $\dim W + \dim W^\perp = n$
4. If $W = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$, then

$$\begin{aligned} W^\perp &= \text{all vectors orthogonal to each } v_1, v_2, \ldots, v_m \\ &= \bigl\{x \text{ in } \mathbf{R}^n \mid x \cdot v_i = 0 \text{ for all } i = 1, 2, \ldots, m\bigr\} \\ &= \mathrm{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}. \end{aligned}$$

Orthogonal Complements

Computation

Problem: if $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$, compute $W^{\perp}$.

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^{\perp} = \text{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}$$

Orthogonal Complements

Row space, column space, null space

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A. Equivalently, it is the column span of A^T:

$$\text{Row}\, A = \text{Col}\, A^T.$$

It is a subspace of $\mathbf{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\, A.$$

Hence we have shown:

Fact: $(\text{Row}\, A)^\perp = \text{Nul}\, A$.

Replacing A by A^T, and remembering $\text{Row}\, A^T = \text{Col}\, A$:

Fact: $(\text{Col}\, A)^\perp = \text{Nul}\, A^T$.

Using property 2 and taking the orthogonal complements of both sides, we get:

Fact: $(\text{Nul}\, A)^\perp = \text{Row}\, A$ and $\text{Col}\, A = (\text{Nul}\, A^T)^\perp$.

Orthogonal Complements

Reference sheet

Orthogonal Complements of Most of the Subspaces We've Seen

For any vectors $v_1, v_2, \ldots, v_m$:

$$\operatorname{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \operatorname{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}$$

For any matrix A:

$$\operatorname{Row} A = \operatorname{Col} A^T$$

and

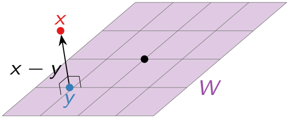
$$(\operatorname{Row} A)^\perp = \operatorname{Nul} A \qquad \operatorname{Row} A = (\operatorname{Nul} A)^\perp$$
$$(\operatorname{Col} A)^\perp = \operatorname{Nul} A^T \qquad \operatorname{Col} A = (\operatorname{Nul} A^T)^\perp$$

Section 6.2

Orthogonal Sets

Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.



Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

How do you know that y is the closest point? The vector from y to x is orthogonal to W: it is in the *orthogonal complement* $W^{\perp}$.

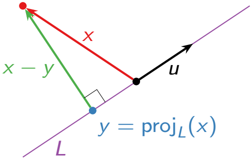
Orthogonal Projection onto a Line

Theorem

Let $L = \mathrm{Span}\{u\}$ be a line in $\mathbf{R}^n$, and let x be in $\mathbf{R}^n$. The closest point to x on L is the point

$$\mathrm{proj}_L(x) = \frac{x \cdot u}{u \cdot u}\, u.$$

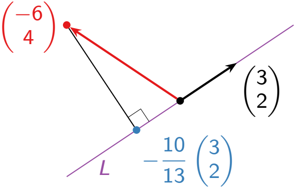
This point is called the **orthogonal projection of** x **onto** L**.**



Orthogonal Projection onto a Line

Example

Compute the orthogonal projection of $x = \begin{pmatrix} -6 \\ 4 \end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.



Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Example: $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ is an orthogonal set. Check:

Lemma

An orthogonal set of vectors is linearly independent.

Orthogonal Bases

An orthogonal set $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ forms a basis for $W = \operatorname{Span}\mathcal{B}$.

An advantage of orthogonal bases is it's *very easy* to compute the $\mathcal{B}$-coordinates of a vector in W.

Theorem

Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span}\mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

In other words, the $\mathcal{B}$-coordinates of x are $\left(\frac{x \cdot u_1}{u_1 \cdot u_1}, \frac{x \cdot u_2}{u_2 \cdot u_2}, \ldots, \frac{x \cdot u_m}{u_m \cdot u_m}\right)$.

Orthogonal Bases

Geometric reason

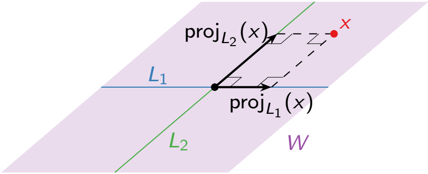
Theorem

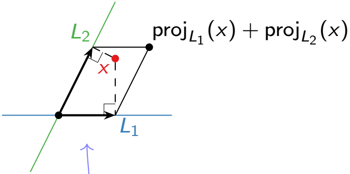
Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span}\mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \overset{\text{proj}_{L_2}(u_2)}{\boxed{\frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2}} + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

If L_i is the line spanned by u_i, then this says

$$x = \text{proj}_{L_1}(x) + \text{proj}_{L_2}(x) + \cdots + \text{proj}_{L_m}(x).$$





Warning: This only works for an *orthogonal* basis.

Orthogonal Bases

Example

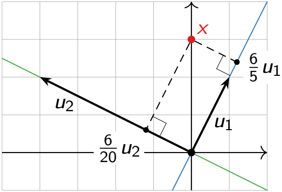
Problem: Find the $\mathcal{B}$-coordinates of $x = \binom{0}{3}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix} \right\}.$$

Old way:

$$\left(\begin{array}{cc|c} 1 & -4 & 0 \\ 2 & 2 & 3 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 6/5 \\ 0 & 1 & 6/20 \end{array}\right) \implies [x]_{\mathcal{B}} = \begin{pmatrix} 6/5 \\ 6/20 \end{pmatrix}.$$

New way: note $\mathcal{B}$ is an *orthogonal* basis.



Orthogonal Bases

Example

Problem: Find the $\mathcal{B}$-coordinates of $x = (6, 1, -8)$ where

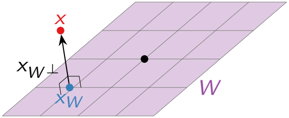
$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

Section 6.3

Orthogonal Projections

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^{\perp}$:



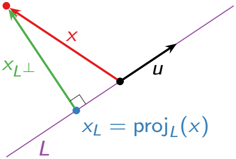
Reformulation: Every vector x can be decompsed uniquely as

$$x = x_W + x_{W^\perp}$$

where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^{\perp}$.

Example: Let $u = \binom{3}{2}$ and let $L = \text{Span}\{u\}$. Let $x = \binom{-6}{4}$. Then the closest point to x in L is $\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u$, so

$$x_L = \text{proj}_L(x) = -\frac{10}{13}\begin{pmatrix}3\\2\end{pmatrix} \qquad x_{L^\perp} = x - \text{proj}_L(x) = \begin{pmatrix}-6\\4\end{pmatrix} + \frac{10}{13}\begin{pmatrix}3\\2\end{pmatrix}.$$



Orthogonal Projections

Definition

Let W be a subspace of $\mathbf{R}^n$, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\text{proj}_W(x) \overset{\text{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Question: What is the difference between this and the formula for $[x]_{\mathcal{B}}$ from before?

Theorem

Let W be a subspace of $\mathbf{R}^n$, and let x be a vector in $\mathbf{R}^n$. Then $\text{proj}_W(x)$ is the closest point to x in W. Therefore

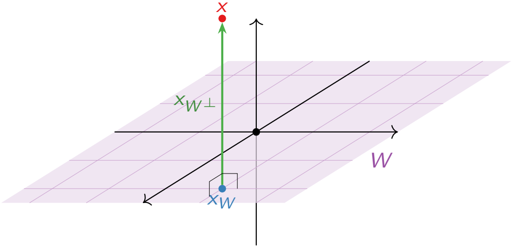
$$\boxed{x_W = \text{proj}_W(x) \qquad x_{W^\perp} = x - \text{proj}_W(x).}$$

Orthogonal Projections

Easy example

What is the projection of $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ onto the *xy*-plane?

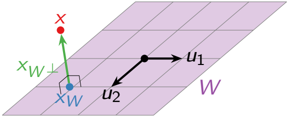
So this is the same projection as before.



Orthogonal Projections

More complicated example

What is the projection of $x = \begin{pmatrix} -1.1 \\ 1.4 \\ 1.45 \end{pmatrix}$ onto $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1.1 \\ -.2 \end{pmatrix} \right\}$?



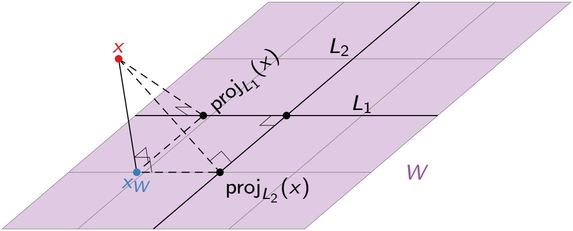
Orthogonal Projections

Picture

Let W be a subspace of $\mathbf{R}^n$, and let $\{u_1, u_2, \ldots, u_m\}$ be an orthogonal basis for W. Let $L_i = \mathrm{Span}\{u_i\}$. Then

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \sum_{i=1}^{m} \mathrm{proj}_{L_i}(x).$$

So the orthogonal projection is formed by adding orthogonal projections onto perpendicular lines.



Orthogonal Projections

Properties

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of $\mathbf{R}^n$, and let x be a vector in $\mathbf{R}^n$. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathbf{R}^n \longrightarrow \mathbf{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of $\mathbf{R}^n$.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) = x$.
3. For every x in $W^\perp$, we have $\mathrm{proj}_W(x) = 0$.
4. The range of proj_W is W.

Poll

Orthogonal Projections

Matrices

What is the matrix for $\operatorname{proj}_W \colon \mathbf{R}^3 \to \mathbf{R}^3$, where

$$W = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}?$$

Orthogonal Projections

Matrix facts

Let W be an m-dimensional subspace of $\mathbf{R}^n$, let $\text{proj}_W\colon \mathbf{R}^n \to W$ be the projection, and let A be the matrix for proj_L.

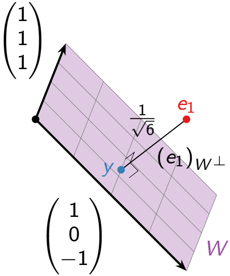
Fact 1: A is diagonalizable with eigenvalues 0 and 1; it is similar to the diagonal matrix with m ones and $n - m$ zeros on the diagonal.

Fact 2: $A^2 = A$.

Orthogonal Projections

Minimum distance

What is the distance from e_1 to $W = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$?



Section 6.4

The Gram–Schmidt Process

Motivation

All of the procedures we learned in §§6.2–6.3 require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

- Finding the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i$$

- Finding the orthogonal projection of a vector x onto the span W of $u_1, u_2, \ldots, u_m$:

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Problem: What if your basis isn't orthogonal?

Solution: The Gram–Schmidt process: take any basis and make it orthogonal.

The Gram–Schmidt Process

Let $\{v_1, v_2, \dots, v_m\}$ be a basis for a subspace W of $\mathbf{R}^n$. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$\vdots$

m. $u_m = v_m - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \dots, u_{m-1}\}}(v_m) \; = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i}\, u_i$

Then $\{u_1, u_2, \dots, u_m\}$ is an *orthogonal* basis for the same subspace W.

Remark

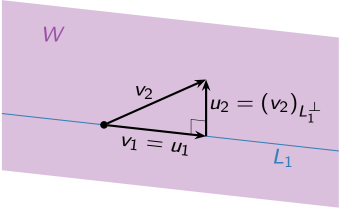
In fact, for every i between 1 and n, the set $\{u_1, u_2, \dots, u_i\}$ is an orthogonal basis for $\mathrm{Span}\{v_1, v_2, \dots, v_i\}$.

The Gram–Schmidt Process

Two vectors

Find an orthogonal basis $\{u_1, u_2\}$ for $W = \mathrm{Span}\{v_1, v_2\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$



Important: $\mathrm{Span}\{u_1, u_2\} = \mathrm{Span}\{v_1, v_2\} = W$: this is an *orthogonal* basis for the *same* subspace.

The Gram–Schmidt Process

Three vectors

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \text{Span}\{v_1, v_2, v_3\} = \mathbf{R}^3$, where

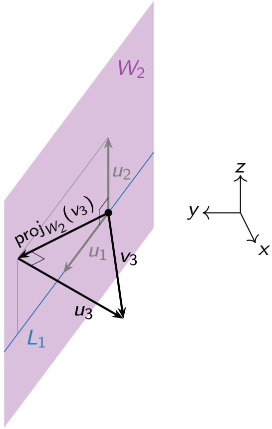
$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

Important: $\text{Span}\{u_1, u_2, u_3\} = \text{Span}\{v_1, v_2, v_3\} = W$: this is an *orthogonal* basis for the *same* subspace.

The Gram–Schmidt Process

Three vectors, continued

$$v_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ v_2 = \begin{pmatrix}1\\1\\1\end{pmatrix},\ v_3 = \begin{pmatrix}3\\1\\1\end{pmatrix} \overset{\text{G–S}}{\rightsquigarrow} u_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ u_3 = \begin{pmatrix}1\\-1\\0\end{pmatrix}$$



The Gram–Schmidt Process

Three vectors in $\mathbf{R}^4$

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \text{Span}\{v_1, v_2, v_3\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_2 = \begin{pmatrix} -1 \\ 4 \\ 4 \\ -1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 4 \\ -2 \\ -2 \\ 0 \end{pmatrix}.$$

Poll

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

The columns of A are a basis for $W = \operatorname{Col} A$. The columns of Q come from Gram–Schmidt as applied to the columns of A, after normalizing to unit vectors. The columns of R come from the steps in Gram–Schmidt.

Here is the procedure for producing a QR factorization.

QR Factorization

Example

Find the QR factorization of $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(The columns of A are the vectors v_1, v_2, v_3 from a previous example.)

Step 1: Run Gram–Schmidt and solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3.

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_1 = u_1$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = v_2 - 1\, u_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad v_2 = u_1 + u_2$$

$$\begin{aligned} u_3 &= v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2 \\ &= v_3 - 2\, u_1 - 1\, u_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \end{aligned} \qquad v_3 = 2u_1 + u_2 + u_3$$

QR Factorization

Example, continued

$$v_1 = \textcircled{1}u_1 \qquad v_2 = \textcircled{1}u_1 + \textcircled{1}u_2 \qquad v_3 = \textcircled{2}u_1 + \textcircled{1}u_2 + \textcircled{1}u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

Do this by putting the above equations in matrix form:

$$A \longrightarrow \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \textcircled{1} & \textcircled{1} & \textcircled{2} \\ 0 & \textcircled{1} & \textcircled{1} \\ 0 & 0 & \textcircled{1} \end{pmatrix}$$

$\widehat{Q}$ $\widehat{R}$

$$\text{first column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1u_1 = v_1$$

$$\text{second column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 1u_1 + 1u_2 = v_2$$

$$\text{third column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 2u_1 + 1u_2 + 1u_3 = v_3$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0/1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0/1 & -1/\sqrt{2} \\ 0/\sqrt{2} & 1/1 & 0/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \cdot \sqrt{2} & 1 \cdot \sqrt{2} & 2 \cdot \sqrt{2} \\ 0 \cdot 1 & 1 \cdot 1 & 1 \cdot 1 \\ 0 \cdot \sqrt{2} & 0 \cdot \sqrt{2} & 1 \cdot \sqrt{2} \end{pmatrix}.$$

Note that the entries in the ith column of Q multiply by the entries in the ith row of R, so this doesn't change the product.

The final QR decomposition is:

$$A = QR \qquad Q = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \qquad R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & 1 & 1 \\ 0 & 0 & \sqrt{2} \end{pmatrix}$$

QR Factorization

Another example

Find the QR factorization of $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & -2 \\ 1 & -1 & 0 \end{pmatrix}$.

(The columns are vectors from a previous example.)

Step 1: Run Gram–Schmidt and solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3:

QR Factorization

Another example, continued

$$v_1 = 1\,u_1 \qquad v_2 = \frac{3}{2}\,u_1 + 1\,u_2 \qquad v_3 = 0\,u_1 - \frac{4}{5}\,u_2 + 1\,u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

QR Factorization

Another example, continued

$$A = \widehat{Q}\widehat{R} \qquad \widehat{Q} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix} \qquad \widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Normalize the columns of $\widehat{Q}$ and the rows of $\widehat{R}$ to get Q and R:

The final QR decomposition is

$$A = QR \qquad Q = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix} \qquad R = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}.$$

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.
But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

It follows that $\det(Q) = \pm 1$.

(Since $\det(R) > 0$, in fact $\det(Q)$ has the same sign as $\det(A)$.)

Therefore,

$$\det(A) = \det(Q)\det(R) = \pm\det(R).$$

But R is upper-triangular, so it's easy to compute its determinant!

In fact, if $v_1, v_2, \ldots, v_n$ are the columns of A, and $u_1, u_2, \ldots, u_n$ are the vectors you obtain by applying Gram–Schmidt, then the (i, i) entry of R is $\|u_i\|$, so

$$\det(A) = \pm\|u_1\|\,\|u_2\|\cdots\|u_n\|.$$

So you can use Gram–Schmidt to compute determinants (up to sign)!

QR Factorization

Application: computing eigenvalues

Let A be an $n \times n$ matrix with real eigenvalues. Here is an algorithm:

$$\begin{aligned} A &= Q_1R_1 && QR \text{ factorization} \\ A_1 &= R_1Q_1 && \text{swap the } Q \text{ and } R \\ &= Q_2R_2 && \text{find its } QR \text{ factorization} \\ A_2 &= R_2Q_2 && \text{swap the } Q \text{ and } R \\ &= Q_3R_3 && \text{find its } QR \text{ factorization} \\ & && \text{et cetera} \end{aligned}$$

Theorem

The matrices A_k converge to an upper triangular matrix, and the diagonal entries converge (quickly!) to the eigenvalues of A.

This gives a computationally efficient way (called the *QR* algorithm) to find the eigenvalues of a matrix.

Section 6.5

Least Squares Problems

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is $\widehat{b} = \operatorname{proj}_{\operatorname{Col} A}(b)$.

Then $A\widehat{x} = \widehat{b}$ is a consistent equation.

A solution $\widehat{x}$ to $A\widehat{x} = \widehat{b}$ is a **least squares solution.**

Least Squares Solutions

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

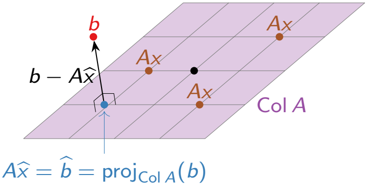
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$\|b - A\widehat{x}\| \leq \|b - Ax\|$$

for all x in $\mathbf{R}^n$.

Note that $b - A\widehat{x}$ is in $(\text{Col}\, A)^{\perp}$.



In other words, a least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Equivalently, a least squares solution to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$A\widehat{x} = \widehat{b} = \text{proj}_{\text{Col}\, A}(b).$$

This is because $\widehat{b}$ is the closest vector to b such that $A\widehat{x} = \widehat{b}$ is consistent.

Least Squares Solutions

Computation

Theorem

The least squares solutions to $Ax = b$ are the solutions to

$$(A^T A)\widehat{x} = A^T b.$$

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!
Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Why is this true?

Alternative when A has orthogonal columns $v_1, v_2, \ldots, v_n$:

$$\widehat{b} = \operatorname{proj}_{\mathrm{Col}\,A}(b) = \sum_{i=1}^{n} \frac{b \cdot v_i}{v_i \cdot v_i}\, v_i$$

The right hand side equals $A\widehat{x}$, where $\widehat{x} = \left(\dfrac{b \cdot v_1}{v_1 \cdot v_1}, \ \dfrac{b \cdot v_2}{v_2 \cdot v_2}, \ \cdots, \ \dfrac{b \cdot v_n}{v_n \cdot v_n} \right)$.

Least Squares Solutions

Example

Find the least squares solutions to $Ax = b$ where:

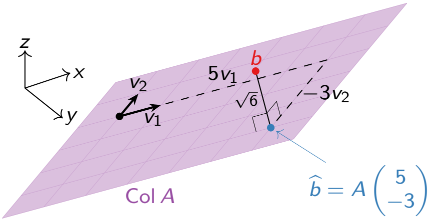
$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

So the only least squares solution is $\widehat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

Least Squares Solutions

Example, continued

How close did we get?



Let

$$v_1 = \begin{pmatrix}1\\1\\1\end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix}0\\1\\2\end{pmatrix}$$

be the columns of A, and let $\mathcal{B} = \{v_1, v_2\}$.

Note $\widehat{x} = \binom{5}{-3}$ is just the $\mathcal{B}$-coordinates of $\widehat{b}$, in $\operatorname{Col} A = \operatorname{Span}\{v_1, v_2\}$.

Least Squares Solutions

Second example

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

So the only least squares solution is $\widehat{x} = \begin{pmatrix} 1/3 \\ -1/3 \end{pmatrix}$.

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

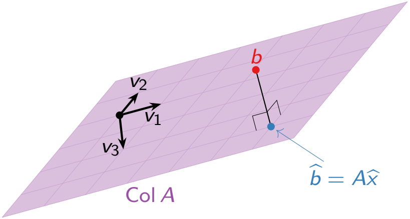
Theorem

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

1. $Ax = b$ has a *unique* least squares solution for all b in $\mathbf{R}^n$.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Why? If the columns of A are linearly *de*pendent, then $A\widehat{x} = \widehat{b}$ has many solutions:

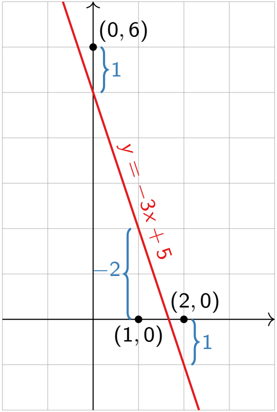


Note: $A^T A$ is always a square matrix, but it need not be invertible.

Application

Data modeling: best fit line

Find the best fit line through $(0, 6)$, $(1, 0)$, and $(2, 0)$.



$$A\begin{pmatrix}5\\-3\end{pmatrix}-\begin{pmatrix}6\\0\\0\end{pmatrix}=\begin{pmatrix}1\\-2\\1\end{pmatrix}$$

Poll

Application

Best fit ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

So we want to solve:

$$\begin{aligned}
(0)^2 + \quad A(2)^2 + \quad B(0)(2) + \quad C(0) + \quad D(2) + E &= 0 \\
(2)^2 + \quad A(1)^2 + \quad B(2)(1) + \quad C(2) + \quad D(1) + E &= 0 \\
(1)^2 + A(-1)^2 + \quad B(1)(-1) + \quad C(1) + D(-1) + E &= 0 \\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0 \\
(-3)^2 + \quad A(1)^2 + \quad B(-3)(1) + C(-3) + \quad D(1) + E &= 0
\end{aligned}$$

In matrix form:

$$\begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

Best fit ellipse:

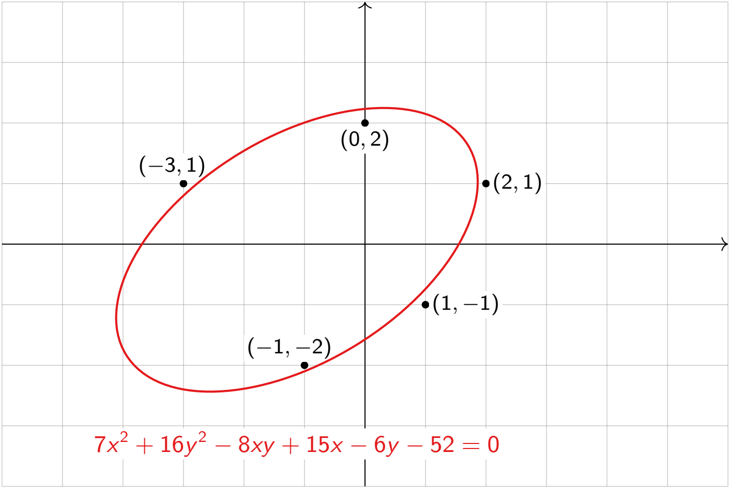
$$x^2 + \frac{16}{7}y^2 - \frac{8}{7}xy + \frac{15}{7}x - \frac{6}{7}y - \frac{52}{7} = 0$$

or

$$7x^2 + 16y^2 - 8xy + 15x - 6y - 52 = 0.$$

Application

Best fit ellipse, picture



Remark: Gauss invented the method of least squares to do exactly this: he predicted the (elliptical) orbit of the asteroid Ceres as it passed behind the sun in 1801.

Application

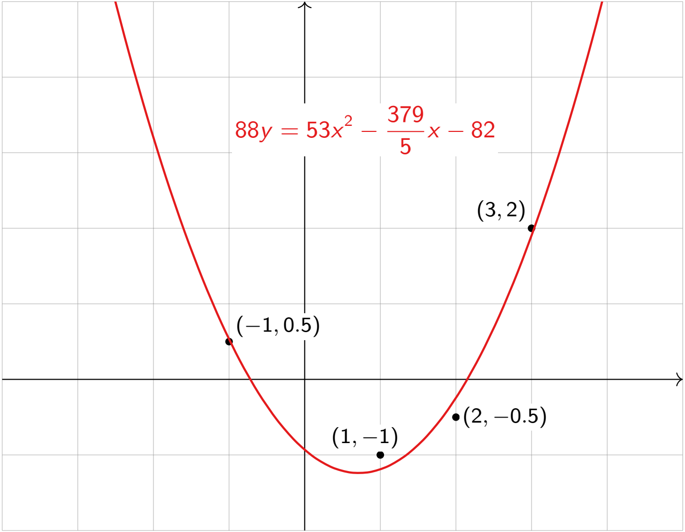
Best fit parabola

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

Answer: $$88y = 53x^2 - \frac{379}{5}x - 82$$

Application

Best fit parabola, picture



Application

Best fit linear function

What least squares problem $Ax = b$ finds the best linear function $f(x, y)$ fitting the following data?

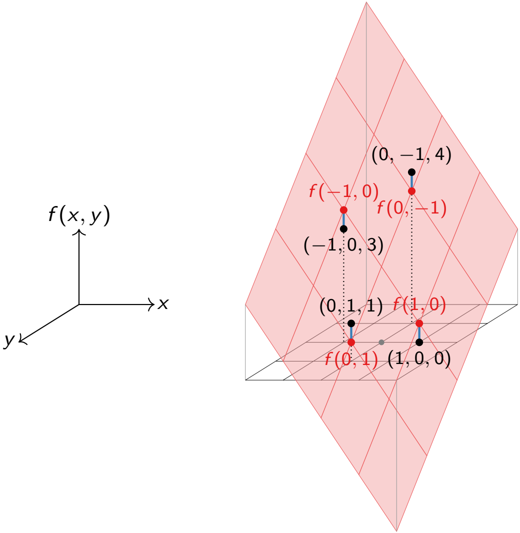
x	y	$f(x, y)$
1	0	0
0	1	1
-1	0	3
0	-1	4

Answer:

$$f(x, y) = -\frac{3}{2}x - \frac{3}{2}y + 2$$

Application

Best fit linear function, picture



Graph of

$$f(x,y) = -\frac{3}{2}x - \frac{3}{2}y + 2$$