Math 415 - Lecture 1

Introduction

Monday August 24 2015

- Textbook: Chapter 1.3, Chapter 2.2 (just the pages 78 and 79)
- Suggested Practice Exercise: in Chapter 1.3, Exercise 1,3, 5, 6, 11
- Khan Academy Video: Matrices: Reduced Row Echelon Form 1

1 Systems of Linear Equations

Definition. A **linear equation** is a equation of the form

$$a_1x_1 + \ldots + a_nx_n = b$$

where $a_1, ..., a_n, b$ are numbers and $x_1, ..., x_n$ are variables.

Example 1. Which of the following equations are linear equations (or can be rearranged to become linear equations)?

$$\begin{array}{l|ll}
4x_1 - 5x_2 + 2 = x_1 & 3x_1 - 5x_2 = -2 & \textit{Linear.} \\
x_2 = 2(\sqrt{6} - x_1) + x_3 & 2x_1 + x_2 - x_3 = 2\sqrt{6} & \textit{Linear.} \\
4x_1 - 6x_2 = x_1x_2 & 4x_1 - 6x_2 = \underline{x_1x_2} & \textit{Not linear.} \\
x_2 = 2\sqrt{x_1} - 7 & x_2 = \underline{2\sqrt{x_1}} - 7 & \textit{Not linear.}
\end{array}$$

This course will focus on linear equations.

Definition. A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same set of variables, say, $x_1, x_2, ..., x_n$.

Definition. A **solution** of a linear system is a list $(s_1, s_2, ..., s_n)$ of numbers that makes each equation in the system true when the values $s_1, s_2, ..., s_n$ are substituted for $x_1, x_2, ..., x_n$, respectively.

Definition. The **solution set** of a system of linear equations is the set of all possible solutions of a linear system.

Example 2. Two equations in two variables:

$$\begin{aligned} x_1 + x_2 &= 1 \\ -x_1 + x_2 &= 0. \end{aligned}$$

What is a solution for this system of linear equations?

Add them. $2x_2 = 1 \quad \Rightarrow \quad x_2 = 5$

Plug into first equation. $x_1 + .5 = 1 \quad \Rightarrow \quad x_1 = .5$

$(x_1, x_2) = (.5, .5)$ is the only solution.

Example 3. Does every system of linear equation have a solution?

$$\begin{aligned} x_1 - 2x_2 &= -3 \\ 2x_1 - 4x_2 &= 8. \end{aligned}$$

Multiply first equation by 2. $2x_1 - 4x_2 = -6$

Subtract from second equation. $0 = 14$

The equation $0 = 14$ is always false, so no solutions exist.

Example 4. How many solutions are there to the following system?

$$\begin{aligned} x_1 + x_2 &= 3 \\ -2x_1 - 2x_2 &= -6 \end{aligned}$$

Multiply first equation by 2. $2x_1 + 2x_2 = 6$

Add to second equation. $0 = 0$

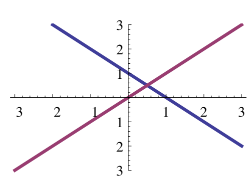
Any value of x_1 works. $x_2 = 3 - x_1$. Infinitely many solutions.

Theorem 1. *This is all there is: A linear system has either*

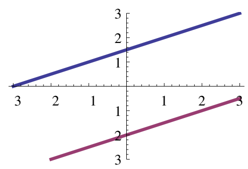
one unique solution *or* ***no solution*** *or* ***infinitely many solutions****.*

Can you draw the set of solutions of the above equations?

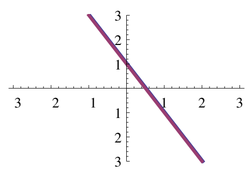
$$\begin{aligned} x_1 + x_2 &= 1 \\ -x_1 + x_2 &= 0. \end{aligned}$$



$$
\begin{aligned}
x_1 - 2x_2 &= -3 \\
2x_1 - 4x_2 &= 8.
\end{aligned}
$$



$$
\begin{aligned}
x_1 + x_2 &= 3 \\
-2x_1 - 2x_2 &= -6
\end{aligned}
$$



(The numbers in the graphs are not quite right.)

Take away: Whenever you have a linear system with n equations, then the set of solutions of this system is precisely the intersection of the sets of solutions of each of the n equations on its own.

1.1 Strategies for solving systems of linear equations

Definition. Two systems are **equivalent** if they have the same solution set.

The general strategy is to replace one system with an equivalent system that is easier to solve.

Example 5. Consider

$$
\begin{array}{rcrcr}
x_1 & - & 2x_2 & = & -1 \\
-x_1 & + & 3x_2 & = & 3
\end{array}
$$

$$
R2 \to R2 + R1 \quad
\begin{array}{rcrcr}
x_1 & - & 2x_2 & = & -1 \\
0 & + & x_2 & = & 2
\end{array}
$$

$x_2 = 2$, so $x_1 = 3$.

1.2 Matrix Notation

Matrix Notation

From a system of equations, we can get:

Coefficient Matrix

$$\begin{array}{rcrcr} x_1 & - & 2x_2 & = & -1 \\ -x_1 & + & 3x_2 & = & 3 \end{array} \qquad \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix}$$

Augmented matrix

$$\begin{array}{rcrcr} x_1 & - & 2x_2 & = & -1 \\ -x_1 & + & 3x_2 & = & 3 \end{array} \qquad \left[\begin{array}{cc|c} 1 & -2 & -1 \\ -1 & 3 & 3 \end{array}\right]$$

$R2 \to R2 + R1$

$$\left[\begin{array}{cc|c} 1 & -2 & -1 \\ 0 & 1 & 2 \end{array}\right]$$

$R1 \to R1 + 2R2$

$$\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \end{array}\right]$$

Solution: $x_1 = 3,\ x_2 = 2$

Definition. An **elementary row operation** is one of the following

(Replacement) Add a multiple of one row to another row,

(Interchange) Interchange two rows, or

(Scaling) Multiply all entries in a row by a nonzero constant.

Definition. Two matrices are **row equivalent**, if one matrix can be transformed into the other matrix by a sequence of elementary row operations.

Theorem 2. *If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.*

Example 6. Solve the following system (or show there is no solution):

$$\begin{aligned} x_1 - 2x_2 + x_3 &= \quad 0 \\ 2x_2 - 8x_3 &= \quad 8 \\ -4x_1 + 5x_2 + 9x_3 &= \ -9 \end{aligned}$$

$$
\begin{aligned}
x_1 - 2x_2 + x_3 &= 0 \\
2x_2 - 8x_3 &= 8 \\
-4x_1 + 5x_2 + 9x_3 &= -9
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 2 & -8 & 8 \\
-4 & 5 & 9 & -9
\end{array}\right]
$$

$$
\begin{aligned}
x_1 - 2x_2 + x_3 &= 0 \\
2x_2 - 8x_3 &= 8 \\
-3x_2 + 13x_3 &= -9
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 2 & -8 & 8 \\
0 & -3 & 13 & -9
\end{array}\right]
$$

$$
\begin{aligned}
x_1 - 2x_2 + x_3 &= 0 \\
x_2 - 4x_3 &= 4 \\
-3x_2 + 13x_3 &= -9
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & -4 & 4 \\
0 & -3 & 13 & -9
\end{array}\right]
$$

$$
\begin{aligned}
x_1 - 2x_2 + x_3 &= 0 \\
x_2 - 4x_3 &= 4 \\
x_3 &= 3
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & -4 & 4 \\
0 & 0 & 1 & 3
\end{array}\right]
$$

$$
\begin{aligned}
x_1 - 2x_2 &= -3 \\
x_2 &= 16 \\
x_3 &= 3
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & -2 & 0 & -3 \\
0 & 1 & 0 & 16 \\
0 & 0 & 1 & 3
\end{array}\right]
$$

$$
\begin{aligned}
x_1 &= 29 \\
{}_2 &= 16 \\
x_3 &= 3
\end{aligned}
\qquad
\left[\begin{array}{ccc|c}
1 & 0 & 0 & 29 \\
0 & 1 & 0 & 16 \\
0 & 0 & 1 & 3
\end{array}\right]
$$

Solution: $(29, 16, 3)$
Check: Is $(29, 16, 3)$ a solution of the ***original*** system?

$$
\begin{array}{rcrcrcr}
x_1 & - & 2x_2 & + & x_3 & = & 0 \\
 & & 2x_2 & - & 8x_3 & = & 8 \\
-4x_1 & + & 5x_2 & + & 9x_3 & = & -9
\end{array}
$$

$$
\begin{array}{rcrcrcrl}
29 & - & 32 & + & 3 & = & 0 & \checkmark \\
 & & 32 & - & 24 & = & 8 & \checkmark \\
-116 & + & 80 & + & 27 & = & -9 & \checkmark
\end{array}
$$

2 Two Fundamental Questions (Existence and Uniqueness)

Two Fundamental Questions (Existence and Uniqueness)
There are two fundamental question about linear equation:

(1) Is the system consistent? (I.e. does a solution **exist**?)

(2) If a solution exists, is it **unique**? (I.e. is there one only one solution?)

Example 7. Is this system consistent? If so, is the solution unique?

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ -4x_1 & + & 5x_2 & + & 9x_3 & = & -9 \end{array}$$

In the last example, this system was reduced to the triangular form:

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & x_2 & - & 4x_3 & = & 4 \\ & & & & x_3 & = & 3 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

This is sufficient to see that the system is consistent and unique. Why?

- The last row determines x_3 uniquely.
- Knowing x_3, the second row determines x_2 uniquely.
- Knowing x_2 and x_3, the first row determines x_1 uniquely.
- So, exactly one possible solution (x_1, x_2, x_3).

Example 8. Is this system consistent?

$$\begin{aligned} 3x_2 - 6x_3 &= 8 \\ x_1 - 2x_2 + 3x_3 &= -1 \\ 5x_1 - 7x_2 + 9x_3 &= 0 \end{aligned} \qquad \left[\begin{array}{rrr|r} 0 & 3 & -6 & 8 \\ 1 & -2 & 3 & -1 \\ 5 & -7 & 9 & 0 \end{array}\right]$$

Solution:

$$\xrightarrow[R1 \leftrightarrow R2]{} \left[\begin{array}{rrr|r} 1 & -2 & 3 & -1 \\ 0 & 3 & -6 & 8 \\ 5 & -7 & 9 & 0 \end{array}\right]$$

$$\xrightarrow[R3 \to R3 - 5R1]{} \left[\begin{array}{rrr|r} 1 & -2 & 3 & -1 \\ 0 & 3 & -6 & 8 \\ 0 & 3 & -6 & 5 \end{array}\right]$$

$$\xrightarrow[R3 \to R3 - R2]{} \left[\begin{array}{rrr|r} 1 & -2 & 3 & -1 \\ 0 & 3 & -6 & 8 \\ 0 & 0 & 0 & -3 \end{array}\right]$$

Equation notation of triangular form:

$$\begin{array}{rrrcr} x_1 & -2x_2 & +3x_3 & = & -1 \\ & 3x_2 & -6x_3 & = & 8 \\ & & 0 & = & -3 \end{array}$$

The original system is inconsistent!

Example 9. For what values of h will the following system be consistent?

$$\begin{array}{rcrcl} 3x_1 & - & 9x_2 & = & 4 \\ -2x_1 & + & 6x_2 & = & h \end{array}$$

Solution:

$$\left[\begin{array}{rr|c} 3 & -9 & 4 \\ -2 & 6 & h \end{array}\right] \xrightarrow[R1 \to \frac{1}{3}R1]{} \left[\begin{array}{rr|c} 1 & -3 & \frac{4}{3} \\ -2 & 6 & h \end{array}\right] \xrightarrow[R2 \to R2 + 2R1]{} \left[\begin{array}{rr|c} 1 & -3 & \frac{4}{3} \\ 0 & 0 & h + \frac{8}{3} \end{array}\right]$$

System is consistent if and only if $h = -\frac{8}{3}$.

Math 415 - Lecture 2

Echelon Forms, General Solution.

Wednesday August 26 2015

Textbook: Chapter 1.3, Chapter 2.2 (just the pages 78 and 79)

Suggested Practice Exercise: in Chapter 1.3, Exercise 17, 23, 24, in Chapter 2.2, Exercise 2 (just reduce A, B to echelon form), 8

Khan Academy Video: Matrices: Reduced Row Echelon Form 1

1 Row Reduction and Echelon Forms

Definition. A matrix is of **Echelon form** (or **row echelon form**) if

1. All nonzero rows are above any rows of all zeros.
2. The number of *leading zeroes* in each row increase going down.

   Or: Each *leading entry* (i.e. left most nonzero entry) of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zero.

A leading entry of an echelon form matrix is also called a ***PIVOT.***

Example 1. Are the following matrices in Echelon form?

(a) $\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ Echelon form ? 1. ✓ 2. ✓ 3. ✓

(b) $\begin{bmatrix} 0 & \blacksquare & * & * & * \\ \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ Not echelon form.
1. ✓ 2. Fails 3. Fails
Would be after $R1 \leftrightarrow R2$

(c) $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{bmatrix}$ Echelon form.
1. ✓ 2. ✓ 3. ✓

(d) $\begin{bmatrix} \blacksquare & 0 & 0 \\ * & \blacksquare & 0 \\ * & 0 & \blacksquare \\ * & 0 & 0 \end{bmatrix}$ Not echelon form.
1. ✓ 2. Fails 3. Fails

(e) $\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \end{bmatrix}$ Echelon form.
1. ✓ 2. ✓ 3. ✓
Leading column of 0s is OK.

Why Echelon Form?
The echelon form of an augmented matrix is good if you want to know if a system is consistent, and if so if there are infinitely many solutions. If you want to find the actual solutions (if any) you need to go further:

Definition. A matrix is of the **reduced echelon form** if in addition to conditions 1, 2, and 3 above it also satisfies

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

Example 2. Are the following matrices in reduced echelon form?

(a) $\begin{bmatrix} 0 & 1 & * & 0 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \end{bmatrix}$ Reduced row echelon form.
1. ✓ 2. ✓ 3. ✓ 4. ✓ 5. ✓

(b) $\begin{bmatrix} 1 & 0 & 5 & 0 & -7 \\ 0 & 2 & 4 & 0 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ No:
4. Fails 5. ✓

(c) $\begin{bmatrix} 1 & 0 & -2 & 3 & 2 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$ No:
4. ✓ 5. Fails

Theorem 1 (Uniqueness of The Reduced Echelon Form)**.** *Each matrix is row-equivalent to one and only one reduced echelon matrix.*

Question: Is the same statement true for Echelon from?
No:

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

Both are row-equivalent and in echelon form.

2 Pivots

Definition. A **pivot position** is the position of a leading entry in an echelon form of the matrix.

Definition. A **pivot** of a matrix is a (nonzero) number that appears in a pivot position.

In a Reduced Row Echelon Form matrix the pivots are 1. Pivots are used to create 0's.

Definition. A **pivot column** is a column that contains a pivot position.

Example 3. In this example, highlight the pivot positions and pivot columns.

$$\begin{bmatrix} 1 & 0 & 5 & 0 & 7 \\ 0 & 2 & 4 & 0 & 6 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{array}{ccccc} \downarrow & \downarrow & & \downarrow & \\ \end{array}$$

$$\begin{bmatrix} \boxed{1} & 0 & 5 & 0 & 7 \\ 0 & \boxed{2} & 4 & 0 & 6 \\ 0 & 0 & 0 & \boxed{5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Example 4. Row reduce to echelon form and locate the pivot columns for the following matrix.

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

Solution:

$$\xrightarrow[R4 \leftrightarrow R1]{} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

$$\xrightarrow[\substack{R2 \to R2+R1 \\ R3 \to R3+2R1}]{} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

$$\xrightarrow[\cdots]{} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

$$\xrightarrow{R3\leftrightarrow R4}\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{2} & 4 & -6 & -6 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note: There is no more than one pivot in any row. There is no more than one pivot in any column.

Example 5. Row reduce to echelon form and then to reduced echelon form:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Solution:

$$\xrightarrow{R1\leftrightarrow R3}\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

$$\xrightarrow{R2\to R2-R1}\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

$$\xrightarrow{R2\to R2-R1}\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

$$\xrightarrow{R3\to R3-\frac{3}{2}R2}\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 6 & 15 \\ 0 & \boxed{2} & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}$$

This is echelon form!

$$\xrightarrow[\substack{R1\to\frac{1}{3}R1 \\ R2\to\frac{1}{2}R2}]{}\begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 1 & -2 & 2 & 1 & -3 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow[\substack{R1\to R1-2R3 \\ R2\to R2-R1}]{}\begin{bmatrix} 1 & -3 & 4 & -3 & 0 & -3 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow{R1\to R1+3R2}\begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

This is reduced row echelon form (RREF)!

3 Solution of linear systems

Why do we care about pivots and pivot columns? Recall: each column of a coefficient matrix corresponds to one of the variables.

Definition. A **pivot variable** (or *basic variable*) is a variable that corresponds to a pivot column in the coefficient matrix of a system.

Definition. A **free variable** is variable that is *not* a pivot variable.

Example 6. Consider the following system of linear equations: $\left[\begin{array}{ccccc|c} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -8 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{array}\right]$

$$\begin{array}{ccccccl} x_1 & +6x_2 & & +3x_4 & & = 0 \\ & & x_3 & -8x_4 & & = 5 \\ & & & & x_5 & = 7 \end{array}$$

What are the pivot columns?

1st, 3rd, and 5th columns.

What are the pivot variables? x_1, x_3, and x_5.

What are the free variables?

x_2 and x_4.

Final Step in Solving a Consistent Linear System: After the augmented matrix is in **reduced** echelon form and the system is written down as a set of equations:

Solve each equation for the pivot variable in terms of the free variables (if any) in the equation.

Example 7 (A general solution).

$$\begin{array}{ccccccl} x_1 & +6x_2 & & +3x_4 & & = 0 \\ & & x_3 & -8x_4 & & = 5 \\ & & & & x_5 & = 7 \end{array} \qquad \begin{cases} x_1 & = -6x_2 - 3x_4 \\ x_2 & = \text{ free} \\ x_3 & = 8x_4 + 5 \\ x_4 & = \text{ free} \\ x_5 & = 7 \end{cases}$$

The **general solution** of the system provides a **parametric description of the solution set**.

- The free variables act as parameters.
- The above system has **infinitely many solutions.** Why?

Because you can pick any value of x_2 and x_4.

Warning: Use only the reduced echelon form to solve a system.

Example 8. Find the parametric description of the solution set of

$$\begin{array}{rrrrrr} & 3x_2 & -6x_3 & +6x_4 & +4x_5 & = -5 \\ 3x_1 & -7x_2 & +8x_3 & -5x_4 & +8x_5 & = 9 \\ 3x_1 & -9x_2 & +12x_3 & -9x_4 & +6x_5 & = 15 \end{array}$$

Its augmented matrix is

$$\left[\begin{array}{rrrrr|r} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{array}\right]$$

We determined earlier that it is reduced echelon form is

$$\left[\begin{array}{rrrrr|r} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right]$$

Equation form of the RREF matrix: $\begin{cases} x_1 \qquad\quad -2x_3 \;\; +3x_4 \qquad\qquad = -24 \\ \qquad x_2 \;\; -2x_3 \;\; +2x_4 \qquad\qquad = -7 \\ \qquad\qquad\qquad\qquad\qquad\quad x_5 \;\; = 4 \end{cases}$

Pivot variables: x_1, x_2, x_5

Free variables: x_3, x_4

General solution: $\begin{cases} x_1 &= 2x_3 - 3x_4 - 24 \\ x_2 &= 2x_3 - 2x_4 - 7 \\ x_3 &= \text{ free} \\ x_4 &= \text{ free} \\ x_5 &= 4 \end{cases}$

4 Existence And Uniqueness

We use the *reduced* echelon form to find the complete solution of a linear system.

The question whether a system has solution and whether it is unique, is much easier to answer than to find the complete solution.

- Echelon Form $\rightarrow$ Existence & Uniqueness.
- Reduced Echelon Form $\rightarrow$ Complete Solution.

Example 9. Let us go back to the following system

$$\begin{aligned} 3x_2 \qquad & - 6x_3 && + 6x_4 && + 4x_5 && = -5 \\ 3x_1 - 7x_2 \qquad & + 8x_3 && - 5x_4 && + 8x_5 && = 9 \\ 3x_1 - 9x_2 \qquad & + 12x_3 && - 9x_4 && + 6x_5 && = 15 \end{aligned}$$

Is the system *Consistent*? Is the solution *Unique*? Are there free variables? To answer these question we need just an echelon form. In an earlier example, we obtained the echelon form:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & | & 15 \\ 0 & 2 & -4 & 4 & 2 & | & -6 \\ 0 & 0 & 0 & 0 & 1 & | & 4 \end{bmatrix}$$

So for the echelon form matrix

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & | & 15 \\ 0 & 2 & -4 & 4 & 2 & | & -6 \\ 0 & 0 & 0 & 0 & 1 & | & 4 \end{bmatrix}$$

1. Is the system consistent? Yes/No? YES! Why? No row $\left[\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & 0 & b \end{array}\right]$!

2. What are the free variables? x_3, x_4.

3. How many solutions?

So we see that there are infinitely many solutions.

Theorem 2 (Existence and Uniqueness Theorem)**.** *A linear system is **consistent** if and only if an echelon form of the augmented matrix has **no** row of the form*

$$\left[\begin{array}{ccc|c} 0 & \dots & 0 & b \end{array}\right],$$

*where b is nonzero.**If** a linear system is consistent, then the solution contains either*

- *a unique solution (when there are no free variables) or*

- *infinitely many solutions (when there is at least one free variable).*

A consistent system can have 1 or ∞ many solutions. Look at the system with augmented matrix

$$\begin{bmatrix} 3 & 4 & | & -3 \\ 2 & 5 & | & 5 \\ -2 & -3 & | & 1 \end{bmatrix}$$

How many pivot variables can this matrix have? Do you expect the system to be consistent? Well, there are at most 2 pivots, so the last row of an echelon form should be $\left[\begin{array}{cc|c} 0 & 0 & b \end{array}\right]$. We cannot predict the value of b without doing some work. We need an echelon form.

The (reduced) echelon form of

$$\begin{bmatrix} 3 & 4 & | & -3 \\ 2 & 5 & | & 5 \\ -2 & -3 & | & 1 \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} 1 & 0 & | & -5 \\ 0 & 1 & | & 3 \\ 0 & 0 & | & 0 \end{bmatrix}$$

So what is b? Is the system consistent? So how many pivots? How many free variables? How many solutions?

Look now at the system with augmented matrix

$$\begin{bmatrix} 3 & 4 & | & -3 \\ 3 & 4 & | & -3 \\ 6 & 8 & | & -6 \end{bmatrix}$$

How many free variables can this matrix have? What is the Echelon form? $\begin{bmatrix} 3 & 4 & | & -3 \\ 0 & 0 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$. Is the system consistent? How many free variables? How many solutions?

Math 415 - Lecture 3

Existence and Uniqueness, linear combinations

Friday August 28 2015

Textbook: Chapter 1.2

Suggested Practice Exercise: Read section 1.2, do problem 1.3:9 (drawing optional)

Khan Academy Video: Linear Combinations and Span

1 Review

Existence and Uniqueness Theorem

Theorem 1 (Existence and Uniqueness Theorem)**.** *A linear system is **consistent** if and only if an echelon form of the augmented matrix has **no** row of the form*

$$\left[\begin{array}{ccc|c} 0 & ... & 0 & b \end{array}\right]$$

*where b is nonzero.**If** a linear system is consistent, then the solution contains either*

- *a unique solution (when there are no free variables) or*
- *infinitely many solutions (when there is at least one free variable).*

Example 1. A consistent system can have 1 or ∞ many solutions. Look at the system with augmented matrix

$$\left[\begin{array}{cc|c} 3 & 4 & -3 \\ 2 & 5 & 5 \\ -2 & -3 & 1 \end{array}\right]$$

How many pivot variables can this matrix have? Do you expect the system to be consistent? Well, there are at most 2 pivots, so the last row of an echelon form should be $\left[\begin{array}{cc|c} 0 & 0 & b \end{array}\right]$. We cannot predict the value of b without doing some work. We need an echelon form.

The (reduced) echelon form of

$$\left[\begin{array}{cc|c} 3 & 4 & -3 \\ 2 & 5 & 5 \\ -2 & -3 & 1 \end{array}\right] \quad \text{is} \quad \left[\begin{array}{cc|c} 1 & 0 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array}\right]$$

- So what is b? Is the system consistent? $b = 0$ so consistent.
- So how many pivots? 2 pivots.
- How many free variables? No free variables.
- How many solutions? Exactly one!

Look now at the system with augmented matrix

$$\left[\begin{array}{cc|c} 3 & 4 & -3 \\ 3 & 4 & -3 \\ 6 & 8 & -6 \end{array}\right]$$

- How many free variables can this matrix have? One or two.Need to calculate.
- What is the Echelon form? $\begin{bmatrix} 3 & 4 & | & -3 \\ 0 & 0 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$.
- Is the system consistent? Yes!
- How many free variables? Exactly one free variable!
- How many solutions? ∞ly many!

1.1 Recap

Recap: Using Row Reduction to Solve Linear Systems

Use the following algorithm:

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent. If not, stop; otherwise go to the next step.
3. Continue row reduction to obtain the reduced echelon form.
4. Write the system of equations corresponding to the matrix obtained in step 3.
5. State the solution by expressing each pivot variable in terms of the free variables and declare the free variables.

1.2 Questions

Questions to check understanding

- On an exam, you are asked to find all solutions to a system of linear equations. You find exactly two solutions. Should you be worried? **YES!**
- True or false?
  - There is no more than one pivot in any row. **TRUE!**
  - There is no more than one pivot in any column. **TRUE!**
  - There cannot be more free variables than pivot variables. **FALSE!**
  - Why? Look at the equation

    $$x_1 + x_2 + x_3 = 0.$$

    How many pivot variables? Free variables?

2 Geometry of Linear Equations

Definition. A **vector** in $\mathbb{R}^n$ is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

i.e., a column with n numbers $x_1, x_2, \ldots, x_n$ in it.

Definition. The **Sum** of $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ is $\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$.

Let c be a real number. Then we define the **Scalar Multiple** $c\mathbf{u}$ by

$$c\mathbf{u} = \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix}$$

Example 2. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Then $\mathbf{u} + \mathbf{v}$ is $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $2\mathbf{u}$ and $-\frac{3}{2}\mathbf{u}$ are $\begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -3/2 \\ -\frac{9}{2} \end{bmatrix}$.

2.1 Linear Combinations

Definition. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given **scalars** $c_1, c_2, \dots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ using scalars (or weights) $c_1, c_2, \dots, c_p$.

Example 3. Linear combinations don't all look the same. The following are linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$:

- $3\mathbf{v}_1 + 2\mathbf{v}_2$,
- $\frac{1}{3}\mathbf{v}_1$,
- $\mathbf{v}_1 - 2\mathbf{v}_2$,
- $\mathbf{0}$.

Example 4. Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$. Express each of the following as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{a} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$$

Solution. Try first $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{a}$ or $c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Staring at this you see that $c_1 = c_2$ and hence $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{a}$ or $\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Try the others for your selves.

Example 5. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 4 \\ 2 \\ 14 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -1 \\ 8 \\ -5 \end{bmatrix}$.

Determine if $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

Solution. Vector $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ if can we find scalars (weights) x_1, x_2, x_3 such that $x_1 \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 2 \\ 14 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix} = \begin{bmatrix} -1 \\ 8 \\ -5 \end{bmatrix}$. It is easy to check that $x_1 = 1, x_2 = -2, x_3 = 2$ works, so $\mathbf{b}$ is indeed a linear combination. How to find these numbers?

How to find these numbers?: $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ and $\mathbf{b}$ are columns of the augmented matrix

$$\begin{array}{cccc} & & & \\ \left[\begin{array}{ccc|c} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 3 & 14 & 10 & -5 \end{array}\right] & & & \\ \uparrow \quad \uparrow \quad \uparrow \quad \uparrow & & & \\ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{b} & & & \end{array}$$

Solution to

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$$

is found by solving the linear system whose augmented matrix is

$$\left[\begin{array}{ccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{b} \end{array}\right].$$

2.2 Linear combinations and linear systems

Motto
Solving linear systems is the same as finding linear combinations!

Theorem 2. *A vector equation*

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

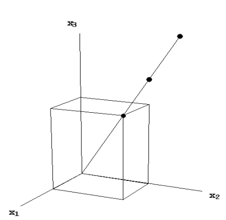
has the same solution set as the linear system whose augmented matrix is

$$\left[\begin{array}{cccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{array}\right]$$

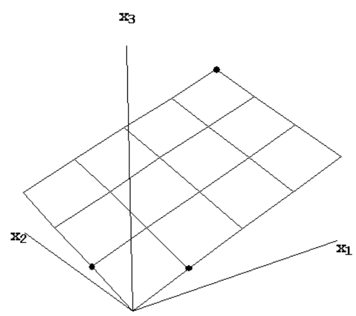
In particular, $\mathbf{b}$ *can be generated by a linear combination of* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ *if and only if there is a solution to the linear system corresponding to the augmented matrix.*

2.3 Span

Example 6. Let $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$. The origin $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ together with $\mathbf{v}$, $2\mathbf{v}$ and $1.5\mathbf{v}$ all lie on the same line.



Span$\{\mathbf{v}\}$ is the set of all vectors of the form $c\mathbf{v}$. Here, **Span**$\{\mathbf{v}\}$ = a line through the origin.



Example 7. Label $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}+\mathbf{v}$ and $3\mathbf{u}+4\mathbf{v}$ on the graph below. $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}+\mathbf{v}$ and $3\mathbf{u}+4\mathbf{v}$ all lie in the same plane. $\mathbf{Span}\{\mathbf{u},\mathbf{v}\}$ is the set of all vectors of the form $x_1\mathbf{u}+x_2\mathbf{v}$. Here, $\mathbf{Span}\{\mathbf{u},\mathbf{v}\}$ = a plane through the origin.

Definition. Suppose $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_p$ are in $\mathbb{R}^n$; then the $\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_p\}$ is defined as the set of all linear combinations of $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_p$.

Stated another way: $\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_p\}$ is the collection of all vectors that can be written as

$$x_1\mathbf{v}_1+x_2\mathbf{v}_2+\cdots+x_p\mathbf{v}_p$$

where $x_1,x_2,\ldots,x_p$ are scalars.

Example 8. Let $\mathbf{v}_1=\begin{bmatrix}2\\1\end{bmatrix}$ and $\mathbf{v}_2=\begin{bmatrix}4\\2\end{bmatrix}$.

(a) Find a vector in $\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}$.

(b) Describe $\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}$ geometrically.

Solution. (a) For instance $2\mathbf{v}_1=\mathbf{v}_2$.

(b) *Span*$\{\mathbf{v}_1,\mathbf{v}_2\}$ is the collection of all vectors in the direction of $\mathbf{v}_1$ (or $\mathbf{v}_2$!). It is a line through the origin.

So the **Span** of two vectors is a plane if and only if they don't point in the same direction.

Example 9. Let $\mathbf{v}_1=\begin{bmatrix}4\\2\\2\end{bmatrix}$ and $\mathbf{v}_2=\begin{bmatrix}6\\3\\2\end{bmatrix}$. Is $\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}$ a line or a plane?

Is v_1 a multiple of v_2? Do they point in the same direction?

Example 10. Let $A=\begin{bmatrix}1&2\\3&1\\0&5\end{bmatrix}$ and $\mathbf{b}=\begin{bmatrix}8\\3\\17\end{bmatrix}$. Is $\mathbf{b}$ in the plane spanned by the columns of A?

Solution.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 8 \\ 3 \\ 17 \end{bmatrix}$$

Do x_1 and x_2 exist such that

$$x_1 \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 3 \\ 17 \end{bmatrix}?$$

pause Try and find the answer at home.

Math 415 - Lecture 6

Elementary Matrices, LU Decomposition

Friday September 4th 2015

Textbook: Chapter 1.4, 1.5

Suggested Practice Exercise: Chapter 1.4 Exercise 22, 27, Chapter 1.5: 4, 5, 11, 23, 29

Khan Academy Video: Matrix multiplication (part I), Matrix multiplication (part II), Defined and undefined matrix operations

Review of matrix multiplication

- **Matrix multiplication is linear combination**: $A\mathbf{x}$ is a linear combination of the columns of A with weights given by the entries of $\mathbf{x}$.

  Example 1.

$$Ax = \begin{bmatrix} 1 & 2 & 3 \\ 4 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + (-1) \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

- **Linear Combination is Linear System**

  Example 2.

$$x_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix} \leftrightarrow \begin{aligned} x_1 + 2x_2 + 3x_3 &= -2 \\ 4x_1 + (-1)x_2 + 0x_3 &= 4 \end{aligned}$$

- $A\mathbf{x} = \mathbf{b}$ is the matrix form of the linear system with augmented matrix $\begin{bmatrix} A & | & \mathbf{b} \end{bmatrix}$.

  Example 3.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix} \leftrightarrow \begin{aligned} x_1 + 2x_2 + 3x_3 &= -2 \\ 4x_1 + (-1)x_2 + 0x_3 &= 4 \end{aligned}$$

$$\leftrightarrow \begin{bmatrix} 1 & 2 & 3 & | & -2 \\ 4 & -1 & 0 & | & 4 \end{bmatrix}$$

- Each column of AB is a linear combination of the columns of A with weights given by the corresponding column of B: $AB = A\begin{bmatrix}\mathbf{b_1} & \dots & \mathbf{b_p}\end{bmatrix} = \begin{bmatrix}A\mathbf{b_1} & \dots & A\mathbf{b_p}\end{bmatrix}$

  Example 4. If $A = \begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$, $B = \begin{bmatrix}3 & 4\\ 0 & 2\end{bmatrix}$, then

$$AB = \left[A\begin{bmatrix}3\\0\end{bmatrix} \quad A\begin{bmatrix}4\\2\end{bmatrix}\right] = \left[3\begin{bmatrix}1\\2\end{bmatrix} + 0\begin{bmatrix}2\\1\end{bmatrix} \quad 4\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\1\end{bmatrix}\right]$$
$$= \begin{bmatrix}3 & 8\\ 6 & 10\end{bmatrix}$$

- Row-column rule: The ij-th entry of AB is $a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$.

  Example 5. If $A = \begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$, $B = \begin{bmatrix}3 & 4\\ 0 & 2\end{bmatrix}$, then the 22 entry of AB is

$$AB_{22} = \begin{bmatrix} & \\ 2 & 1\end{bmatrix}\begin{bmatrix} & 4\\ & 2\end{bmatrix} = 2 \times 4 + 1 \times 2 = 10$$

- Matrix multiplication is not commutative: usually, $AB \neq BA$.

Powers of A

Powers of A

We write: $A^k = A \cdots A$, k-times.

For which matrices A does this make sense? If A is $m \times n$ what can m, n be?

Example 6.

$$\begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix}^3 = \begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix}$$
$$= \begin{bmatrix}1 & 0\\ 9 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\ 3 & 2\end{bmatrix}$$
$$= \begin{bmatrix}1 & 0\\ 21 & 8\end{bmatrix}$$

Calculating high powers of large matrix is hard. We will later learn a clever way to do this efficiently.

Transpose

Definition. If A is $m \times n$, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A. In terms of matrix elements $(A^T)_{ij} = A_{ji}$.

Example 7. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 8 \\ 7 & 6 & 5 & 4 & 3 \end{bmatrix}$.

Then $A^T = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 7 & 6 \\ 3 & 8 & 5 \\ 4 & 9 & 4 \\ 5 & 8 & 3 \end{bmatrix}$

Example 8. Let $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix}$. Compute AB, $(AB)^T$, $A^T B^T$ and $B^T A^T$.

Solution.

$$\begin{aligned}
AB &= \begin{bmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 10 \end{bmatrix} \\
(AB)^T &= \begin{bmatrix} 1 & 1 \\ 4 & 10 \end{bmatrix} \\
A^T B^T &= \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 3 & 10 \\ 2 & 0 & -4 \\ 2 & 1 & 4 \end{bmatrix} \\
B^T A^T &= \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 4 & 10 \end{bmatrix}
\end{aligned}$$

Conclusion
The transpose of a product is the product of transposes **IN OPPOSITE ORDER**:

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

Definition. A is **symmetric** if $A = A^T$.

Example 9. Which of these is symmetric?

$$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}^T =$$

$$\begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix}^T =$$

Theorem 1. *Let A and B denote matrices whose sizes are appropriate for the following sums and products.*

(a) $\left(A^T\right)^T = A$,

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

(c) *For any scalar* r, $(rA)^T = rA^T$

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

Example 10. Prove that $(ABC)^T = C^T B^T A^T$.

Solution. By part d of the Theorem,$(ABC)^T = (A(BC))^T = (BC)^T A^T = C^T B^T A^T$.

Elementary matrices

Definition. The $n \times n$ **identity matrix** I_n has all entries 0, except on the main diagonal where the entries are 1. For example

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

Definition. An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix. A **permutation matrix** is one that is obtained by performing row exchanges on an identity matrix.

Example 11. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$.

E_1, E_2, and E_3 are elementary matrices. Why? Are there any permutation matrices?

Solution. Observe the following products and describe how these products can be obtained by elementary row operations on A.

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ 2d & 2e & 2f \\ g & h & i \end{bmatrix}$$

$$E_2 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ g & h & i \\ d & e & f \end{bmatrix}$$

$$E_3 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g+3a & h+3b & i+3c \end{bmatrix}$$

Theorem 2. *If an elementary row operation is performed on an* $m \times n$*- matrix* A*, the resulting matrix can be written as* EA*, where the* $m \times m$*-matrix* E *is created by performing the same row operations on* I_m.

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

We write this as

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

More on inverses soon.

Remark. Elementary matrices are *invertible* because row operations are *reversible*. To determine the inverse of an elementary matrix E, determine the elementary row operation needed to transform E back into I and apply this operation to I to find the inverse.

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

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

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Example 12. Compute the following products using the row or column interpretation of matrix multiplication. (Don't just use the row-column rule.)

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

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

Math 415 - Lecture 7

LU-decomposition

Wednesday September 9th 2015

Textbook: Chapter 1.5

Suggested Practice Exercise: Chapter 1.5 Exercise 4, 5, 11, 23, 29

1 Review - Elementary matrices

- Multiply row 3 by 7:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 7g & 7h & 7i \end{bmatrix}$$

- Switch rows 2 and 3:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ g & h & i \\ d & e & f \end{bmatrix}$$

- $R3 \to 3R1 + R3$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 3a+g & 3b+h & 3c+i \end{bmatrix}$$

- Taking the inverse of an elementary matrix:

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

2 Triangular matrices

Definition. An $n \times n$ matrix A is called upper triangular if it is of the form

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

An $n \times n$ matrix B is called lower triangular if it is of the form

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

Definition. A matrix A has LU factorization if there is a lower triangular matrix L and a upper triangular matrix U such that

$$A = LU.$$

(In practice, L will have all 1's on the main diagonal.)

When is this possible?

Theorem 1. *Let A be a $n \times n$-matrix. If A can be transformed into echelon form without the use of row exchanges, then A has LU factorization.*

Example 1. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$. Can we transform it into Echelon form without row exchanges?

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boxed{2} & 1 & 1 \\ \boxed{4} & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}, \ell_{21} = 2$$

$$E_2 (E_1 A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boxed{2} & 1 & 1 \\ 0 & -8 & -2 \\ \boxed{-2} & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix}, \ell_{31} = -1$$

$$E_3 (E_2 E_1 A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & \boxed{-8} & -2 \\ 0 & \boxed{8} & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}, \ell_{32} = -1$$

We got an upper triangular matrix!

$$E_3E_2E_1A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U$$

(Always works - if an $n \times n$ matrix is in echelon form, then it is upper triangular.)
We need to reverse these operations:

$$E_3E_2E_1A = U \implies A = E_1^{-1}E_2^{-1}E_3^{-1}U$$

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

$$E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} = L$$

This is a lower triangular matrix!

Motto
Product of lower triangulars is lower triangular.

So the LU decomposition is

$$A = LU = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Summary:

- If A can be brought in echelon form without row exchanges we have $A = LU$,
- U - Echelon form of A
- $L = E_1^{-1}E_2^{-1}E_3^{-1}$ where E_1, E_2, E_3 were elementary matrices that put A into Echelon form. (No row exchanges!)
- $L = I+$ strictly lower triangular, and ℓ_{ij} is the factor between pivot and the entry you want to make zero in the elimination process: see the boxed numbers.

3 Row exchanges

Recall that a permutation matrix P is a square matrix obtained from the identity matrix by reordering the rows.

Theorem 2. *Let A be a $n \times n$-matrix that can be brought to echelon form. Then there is permutation matrix P such that PA has LU factorization.*

Reason: If A can be brought to echelon form with the help of row exchanges, we can do those exchange first. So there is a permutation matrix P such that PA can be brought to echelon form without row exchanges. $\square$

Example 2. Let $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \end{bmatrix}$. Find PA that has a LU factorization.

Let's rearrange the rows:

- Move the 2nd row to the 1st row
- Move the 3rd row to the 2nd row
- Move the 1st row to the 3rd row

Do these moves to the identity matrix to get P:

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

Doing these moves to A gives the same matrix as PA:

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

We can bring PA to echelon form without row exchanges (check this!) so $PA = LU$:

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

4 Applications

Theorem 3. *Let A be an $n \times n$-matrix such that $A = LU$, where L is a lower triangular matrix and U is a upper triangular matrix. Then x will be solution of*

$$Ax = b$$

if and only if x is a solution of

$$Ux = c,$$

where c satisfies $Lc = b$.

Point:$Ux = c$ and $Lc = b$ are *triangular systems*, easy to solve by substitution.

Proof. If $Lc = b$ and $Ux = c$, then

$$Ax = (LU)x = L(Ux) = Lc = b$$

On the other hand, suppose $Ax = b$. We take c to be the vector Ux. Then Lc is equal to $L(Ux) = Ax = b$, so in total we have both $Ux = c$ and $Lc = b$. $\square$

Example 3. Solve

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

We found already a LU factorization for this matrix A. So you first have to solve $Lc = b$ for c:

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}.$$

Use **forward substitution:**

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 2 & 1 & 0 & -2 \\ -1 & -1 & 1 & 9 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -12 \\ 0 & -1 & 1 & 14 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

So

$$c = \begin{bmatrix} 5 \\ -12 \\ 2 \end{bmatrix}$$

Then solve $Ux = c$ for x:

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -12 \\ 2 \end{bmatrix}.$$

This uses **backwards substitution.**

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

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

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

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

So

$$x = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

5 Practice problems

- Is $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ upper triangular? Lower triangular?

  Yes, it is both upper and lower triangular.

- Is $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ upper triangular? Lower triangular?

  No, it is neither upper nor lower triangular.

- True or false? A permutation matrix is one that is obtained by performing column exchanges on an identity matrix.

  Technically yes, but this isn't how we defined it. We defined it as row exchanges on an identity matrix.

- Why do we care about LU decomposition if we already have Gaussian elimination?

  It's faster, especially if we have to feed in lots of different values of **b**.

Example 4. Solve

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \\ -3 \end{bmatrix}.$$

using the factorization we already have.

Math 415 - Lecture 8

Inverses.

Wednesday September 11th 2015

Textbook: Chapter 1.6

Suggested Practice Exercise: Chapter 1.6 Exercise 1, 2, 4, 6, 10, 11, 18, 35, 36, 37, 38, 40, 49, 50

Khan Academy Video: Inverse Matrix (part I), Inverse Matrix (part II)

1 Review

- Elementary matrices perform row operations:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ -2a+d & -2b+e & -2c+f \\ g & h & i \end{bmatrix}$$

- Gaussian elimination on A gives a decomposition $A = LU$:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

  U is the echelon form, L records the reverse of the row operations we did.

- LU decomposition lets us solve $A\mathbf{x} = \mathbf{b}$ quickly for many different $\mathbf{b}$.

1.1 Today's goal

- We know how to reverse a single row operation:

$$\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Inverting a more complicated matrix is harder:

$$\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & b & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ab & -b & 1 \end{bmatrix}$$

Goal today: how to find an "inverse" to any (square!) matrix. *Today A will be an $n \times n$ matrix*

2 The inverse of a matrix

The inverse of a real number a is denoted by a^{-1}. For example, $7^{-1} = 1/7$ and

$$7 \cdot 7^{-1} = 7^{-1} \cdot 7 = 1.$$

Remember that the identity matrix I_n is the $n \times n$-matrix

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

Definition. An $n \times n$ matrix A is said to be invertible if there is an $n \times n$ matrix C satisfying

$$CA = AC = I_n$$

where I_n is the $n \times n$ identity matrix. We call C the inverse of A .

Example 1. We already know that an elementary matrix is invertible:

$$\begin{bmatrix} 1 & 0 & 0 \\ 8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In fact:

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

(Check this at home!) So the definition works!

Theorem 1. *Let A be an invertible matrix, then its inverse C is unique.*

Proof. Assume B and C are both inverses of A. Then

$$B = BI_n = BAC = I_nC = C$$

$\square$

- We will write A^{-1} for the inverse of A. Multiplying by A^{-1} is like "dividing by A."
- Do not write $\frac{A}{B}$. Why?

  It is unclear whether this means AB^{-1} or $B^{-1}A$, and these two matrices are *different.*
- Fact: if $AB = I$ then $A^{-1} = B$ and so $BA = I$. (Not so easy to show at this stage.)

Remark. Not all $n \times n$ matrices are invertible. For example, the 2×2 matrix

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

is not invertible. Try to find an inverse!

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix} \neq I_2$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a \\ 0 & c \end{bmatrix} \neq I_2$$

Definition. A matrix which is *not* invertible is sometimes called a singular matrix. An invertible matrix is also called nonsingular matrix.

Theorem 2. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *If* $ad - bc \neq 0$, *then* A *is invertible and*

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

If $ad - bc = 0$, *then* A *is not invertible.*

Proof. Calculate

$$\begin{aligned} \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \frac{1}{ad - bc} \begin{bmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

$\square$

Quick question: when is the 1×1 matrix $[\ a\ ]$ invertible?
When $a \neq 0$. Its inverse is $[\ \frac{1}{a}\ ]$.

Theorem 3. *If A is an invertible $n \times n$ matrix, then for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.*

Proof. The vector $A^{-1}\mathbf{b}$ is a solution, because

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

Suppose there is another solution $\mathbf{w}$, then

$$\begin{aligned} A\mathbf{w} &= \mathbf{b} \\ A^{-1}A\mathbf{w} &= A^{-1}\mathbf{b} \\ I_n\mathbf{w} &= A^{-1}\mathbf{b} \\ \mathbf{w} &= A^{-1}\mathbf{b} \end{aligned}$$

□

Example 2. Use the inverse of $A = \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}$ to solve

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

Solution. Matrix form of the linear system:

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

$$A^{-1} = \frac{1}{14-15} \begin{bmatrix} -2 & -3 \\ -5 & -7 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}$$

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 17 \end{bmatrix}$$

Check this works!

3 Calculational rules

Theorem 4. *Suppose A and B are invertible. Then the following results hold:*

(a) A^{-1} *is invertible and* $\left(A^{-1}\right)^{-1} = A$ *(i.e. A is the inverse of A^{-1}).*

(b) AB *is invertible and* $(AB)^{-1} = B^{-1}A^{-1}$

(c) A^T *is invertible and* $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$

Note: the inverse of the product is the product of inverses *in opposite order*. Think about putting on socks and shoes. How do you undo those two operations?

Proof. (a) A^{-1} is invertible and $\left(A^{-1}\right)^{-1} = A$

$$AA^{-1} = I = A^{-1}A \quad \checkmark$$

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

$$(B^{-1}A^{-1})(AB) = B^{-1}IB = B^{-1}B = I \quad \checkmark$$
$$(AB)(B^{-1}A^{-1}) = AIA^{-1} = AA^{-1} = I \quad \checkmark$$

(c) A^T is invertible and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I \quad \checkmark$$
$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I \quad \checkmark$$

$\square$

4 An algorithm for computing the inverse matrix

Idea:

- To solve $Ax = b$ we row reduce $[\ A \mid b\ ]$.
- To solve $AX = I_n$ we row reduce $[\ A \mid I\ ]$.

Theorem 5. *An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n will also transform I_n to A^{-1}.*

So here is the algorithm:

- Place A and I side-by-side to form an augmented matrix $[\ A \mid I\ ]$. This is an $n \times 2n$ matrix (*B*ig *A*ugmented *M*atrix), instead of $n \times (n+1)$, for the usual augmented matrix.
- Then perform row operations on this matrix (which will produce identical operations on A and I).
- So by Theorem 5:

$$[\ A \mid I\ ] \text{ will row reduce to } \left[\ I \mid A^{-1}\ \right]$$

or A is not invertible.

Example 3. Find the inverse of $A = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, if it exists.

Solution:

$$[A \quad I] = \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \sim \cdots \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

So

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

Example 4 (Let's do the previous example step by step.).

$$[\ A \mid I\ ] = \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R1 \to \frac{1}{2}R1}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R2 \to R2+3R1}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R2 \leftrightarrow R3}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

Check at home that $AA^{-1} = I_3$.

Remark. Why does this algorithm work?

- At each step, we get

$$[\ A \mid I\ ] \rightsquigarrow [\ E_1 A \mid E_1\ ] \rightsquigarrow [\ E_2 E_1 A \mid E_2 E_1\ ] \rightsquigarrow \ldots$$

- So each step is of the form

$$[\ FA \mid F\ ], \quad F = E_r \ldots E_3 E_2 E_1$$

- If we succeed in row reducing A to I, the final step is

$$[\ FA \mid F\ ] = [\ I \mid F\ ]$$

- So $FA = I$, which means that $A^{-1} = F$.

Practice Problems. Find the inverse of A:

- $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix}$. Hint: What is $A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$?
- $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 6 & 7 & 8 & 0 \\ 9 & 0 & 1 & 0 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

Math 415 - Lecture 9

Vector spaces and subspaces

Monday September 14th 2015

Textbook: Chapter 2.1.

Suggested practice exercises: Chapter 2.1: 1, 2, 10, 11, 17, 18.

Khan Academy video: Linear Subspaces

We know how to find the inverse of a 2×2 matrix. What about 3×3,....,$n\times n$? We use:

Theorem 1. *An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n will also transform I_n to A^{-1}.*

Note that this tells us a lot about $Ax = b$ if A is invertible.

- $Ax = b$ has how many pivots in A?
- How many free variables?
- Can $Ax = b$ be inconsistent?

Here is the algorithm to find the inverse of a matrix A, called the *Gauss-Jordan Method*

- Place A and I side-by-side to form an augmented matrix $[\ A \mid I\ ]$. This is an $n \times 2n$ matrix (*B*ig *A*ugmented *M*atrix), instead of $n \times (n+1)$, for the usual augmented matrix.
- Then perform row operations on this matrix (which will produce identical operations on A and I).
- So by the Theorem:

$$[\ A \mid I\ ] \text{ will row reduce to } \left[\ I \mid A^{-1}\ \right]$$

  or A is not invertible.

Example 1. Find the inverse of $A = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, if it exists.

Solution:

$$\left[A \mid I\ \right] = \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \sim \cdots \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

So

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

Example 2 (Let's do the previous example step by step.).

$$[\ A \mid I\ ] = \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R1 \to \frac{1}{2}R1}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R2 \to R2+3R1}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R2 \leftrightarrow R3}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

Check at home that $AA^{-1} = I_3$.

Remark. Why does this algorithm work?

- At each step, we get

$$[\ A \mid I\ ] \rightsquigarrow [\ E_1 A \mid E_1\ ] \rightsquigarrow [\ E_2 E_1 A \mid E_2 E_1\ ] \rightsquigarrow \ldots$$

- So each step is of the form

$$[\ FA \mid F\ ], \quad F = E_r \ldots E_3 E_2 E_1$$

- If we succeed in row reducing A to I, the final step is

$$[\ FA \mid F\ ] = [\ I \mid F\ ]$$

- So $FA = I$, which means that $A^{-1} = F$.

Example 3. Use the Gauss Jordan method to compute the inverse of

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

Solution.

$$[\, A \mid I \,] = \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{R3 \to R1+R3}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 & 0 & 1 \end{array}\right]$$

$$\underset{R3 \to R2+R3}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{array}\right]$$

Failure: the reduced row echelon form of A will not be I, so A has no inverse!

Practice Problems. Find the inverse of A:

- $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ -2 & 1 & 1 \end{bmatrix}$. Hint: What is $A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$?
- $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 6 & 7 & 8 & 0 \\ 9 & 0 & 1 & 0 \end{bmatrix}$.
- $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

1 Vector Spaces and Subspaces

- The most important property of column vectors in $\mathbb{R}^n$ is that you can take *linear combinations* of them.
- There are many mathematical objects $X, Y, \ldots$ for which a linear combination $cX + dY$ make sense, and have the usual properties of linear combination in $\mathbb{R}^n$

- We are going to define a *vector space* in general as a collection of objects for which linear combinations make sense. The objects of such a set are called vectors.

Definition. A **vector space** is a non-empty set V of objects, called *vectors*, for which linear combinations make sense. More precisely: on V there are defined two operations, called *addition* and *multiplication* by scalars (real numbers), subject to the ten axioms below. The axioms must hold for all u, v, and w in V and for all scalars c and d.

1. $\mathbf{u}+\mathbf{v}$ is in V. (V is "closed under addition".)
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.
4. There is a vector (called the zero vector) $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V satisfying $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.

Definition Continued

6. $c\mathbf{u}$ is in V. (V is "closed under scalar multiplication".)
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.
9. $(cd)\mathbf{u}=c(d\mathbf{u})$.
10. $1\mathbf{u}=\mathbf{u}$.

2 Vector Space Examples

Example 4. Let $M_{2x2}=\left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} : a,b,c,d\in\mathbb{R}\right\}$. This is a vector space.

We need to say what the two operations are. Addition:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}+\begin{bmatrix} e & f \\ g & h \end{bmatrix}=\begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}.$$

Scalar Multiplication:

$$e\cdot\begin{bmatrix} a & b \\ c & d \end{bmatrix}=\begin{bmatrix} ea & eb \\ ec & ed \end{bmatrix}.$$

Next we need to say what the zero vector is. Question: What is the matrix $\mathbf{0}$ such that $\mathbf{0}+A=A$ for any (2×2) matrix A? Answer: We see that the $\mathbf{0}$ vector is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Then we need to check all the 10 axioms. They follow from the corresponding properties of ordinary numbers.

Remarks

- We can take instead of matrices of size 2×2 matrices of any shape: you can check that the set $M_{m\times n}$ of $m \times n$ matrices is also a vector space, in the same way as we indicated above.
- Confusing: in the vector space $M_{2\times 2}$ the vectors are in fact 2×2 matrices!
- In the definition of the vector space $M_{2\times 2}$ the multiplication of matrices plays no role; matrix multiplication will show up when we study the connections *between* vector spaces.
- a "vector" $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ behaves very much like a column vector $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$. A fancy person would say that the vector spaces $M_{2\times 2}$ and $\mathbb{R}^4$ are *isomorphic*.

Example 5. Let $n \geq 0$ be an integer and let

$$\mathbf{P}_n = \text{the set of all polynomials of degree at most } n.$$

This is a vector space.

Members of $\mathbf{P}_n$ have the form

$$\mathbf{p}(t) = a_0 + a_1 t + \cdots + a_n t^n$$

where $a_0, a_1, \ldots, a_n$ are real numbers and t is a variable. We will just verify 3 out of the 10 axioms here.

Vector Space Examples

Let $\mathbf{p}(t) = a_0 + a_1 t + \cdots + a_n t^n$ and $\mathbf{q}(t) = b_0 + b_1 t + \cdots + b_n t^n$ and let c be a scalar. The polynomial $\mathbf{p} + \mathbf{q}$ is defined as follows:

$$(\mathbf{p} + \mathbf{q})(t) = \mathbf{p}(t) + \mathbf{q}(t).$$

Therefore,

$$\begin{aligned} (\mathbf{p} + \mathbf{q})(t) &= \mathbf{p}(t) + \mathbf{q}(t) \\ &= (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n. \end{aligned}$$

which is also a polynomial of degree at most n. So $\mathbf{p} + \mathbf{q}$ is in $\mathbf{P}_n$ (i.e. $\mathbf{P}_n$ is closed under addition). This verifies Axiom 1.

Vector Space Examples

Next we need to find a zero vector. **Question:** What this is the polynomial $\mathbf{0}(t)$ such that $\mathbf{0}(t) + p(t) = p(t)$? **Answer:** Take $\mathbf{0}(t) = 0 + 0t + \cdots + 0t^n$ (zero vector in $\mathbf{P}_n$) Then

$$\begin{aligned} (\mathbf{p} + \mathbf{0})(t) &= (a_0 + a_1 t + \cdots + a_n t^n) + (0 + 0t + \cdots + 0t^n) \\ &= (a_0 + 0) + (a_1 + 0)t + \cdots + (a_n + 0)t^n \\ &= a_0 + a_1 t + \cdots + a_n t^n \end{aligned}$$

and so $\mathbf{p} + \mathbf{0} = \mathbf{p}$. This verifies Axiom 4. Next we define scalar multiplication. Remember $\mathbf{p}(t) = a_0 + a_1 t + \cdots + a_n t^n$. We define

$$(c\mathbf{p})(t) = c\mathbf{p}(t) = (ca_0) + (ca_1)t + \cdots + (ca_n)t^n$$

which is in $\mathbf{P}_n$. so that Axiom 6 holds. The other 7 axioms also hold, so $\mathbf{P}_n$ is a vector space.

3 Subspaces

New vector spaces may be formed from subsets of other vector spaces. These are called **subspaces**.

Definition. A *subspace* of a vector space V is a subset H of V that satisfies 3 properties:

- The zero vector (of V) belongs to H.
- If $\mathbf{u}, \mathbf{v}$ both belong to H also the sum $\mathbf{u} + \mathbf{v}$ belongs to H. (H is *closed* under vector addition).
- If $\mathbf{u}$ is in H and c is any scalar also $c\mathbf{u}$ belongs to H. (H is closed under scalar multiplication.)

Note that if the subset H satisfies these three properties, then H itself is a vector space.

Example 6. $Z = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$ is a subspace of $\mathbb{R}^2$. Why?

Check:

- $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in Z.
- $\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0+0 \\ 0+0 \end{bmatrix}$ is in Z.
- $c \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c0 \\ c0 \end{bmatrix}$ is in Z.

Z is called the zero subspace of $\mathbb{R}^2$. Every vectorspace has a zero subspace consisting just of the zero vector.

Example 7. $H = span \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is a subspace of $\mathbb{R}^2$. Why?

Check:

- $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in H.

- $\begin{bmatrix} a \\ a \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ a+b \end{bmatrix}$ is in H.
- $c \begin{bmatrix} a \\ a \end{bmatrix} = \begin{bmatrix} ca \\ ca \end{bmatrix}$ is in H.

Example 8. Let $H = \left\{ \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Show that H is a subspace of $\mathbb{R}^3$.

Verify properties $1, 2$, and 3 of the definition of a subspace.

- The zero vector of $\mathbb{R}^3$ is in H.

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in H, \quad (a = b = 0)$$

Subspaces

- Adding two vectors in H always produces another vector whose second entry is 0 and therefore the sum of two sectors in H is also in H. (H is closed under addition.)

$$\begin{bmatrix} a \\ 0 \\ b \end{bmatrix} + \begin{bmatrix} c \\ 0 \\ d \end{bmatrix} = \begin{bmatrix} a+c \\ 0 \\ b+d \end{bmatrix}.$$

- Multiplying a vector in H by a scalar produces another vector in H. (H is closed under scalar multiplication.)

$$c \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} = \begin{bmatrix} ca \\ 0 \\ cb \end{bmatrix}.$$

Since those three properties hold, H is a subspace of $\mathbb{R}^3$.

Remark. Vectors $(a, 0, b)$ look and act like the points (a, b) in $\mathbb{R}^2$. But they are **not** the same!

Example 9. Is $H = \left\{ \begin{bmatrix} x \\ x+1 \end{bmatrix} : x \in \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$? (i.e. does H satisfy the properties of a subspace?)

H does not contain the zero vector (property 1).

$$\begin{bmatrix} x \\ x+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

cannot be true for any value of x. Therefore, H is **not** a subspace!

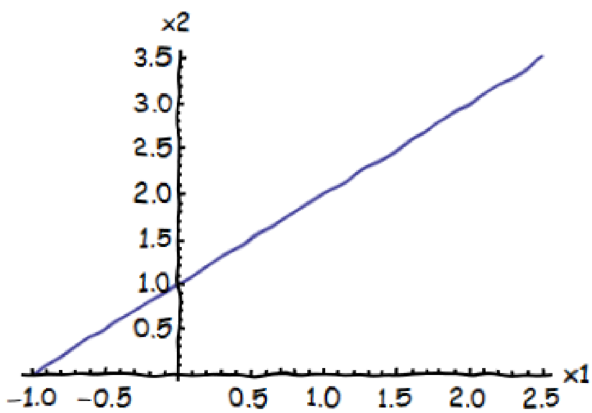
Example 10. Is $H = \left\{ \begin{bmatrix} x \\ x+1 \end{bmatrix} : x \in \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$? (i.e. does H satisfy the properties of a subspace?)

Another way to show that H is not a subspace of $\mathbb{R}^2$ is to check whether H is closed under addition (property 2).

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in H$$

but

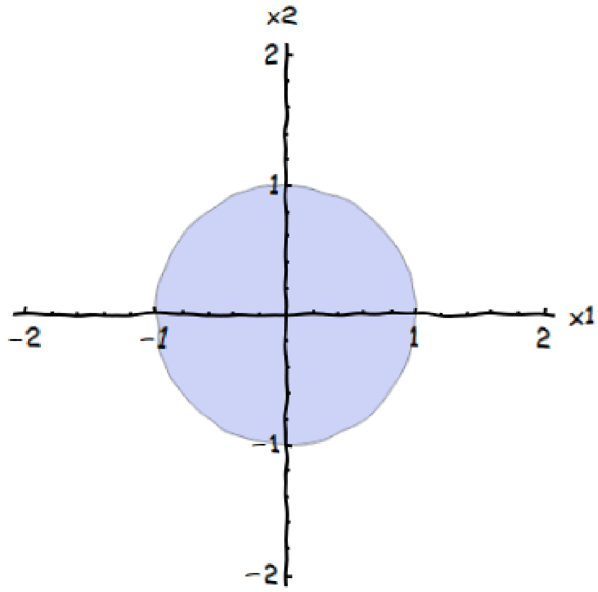
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} \notin H.$$

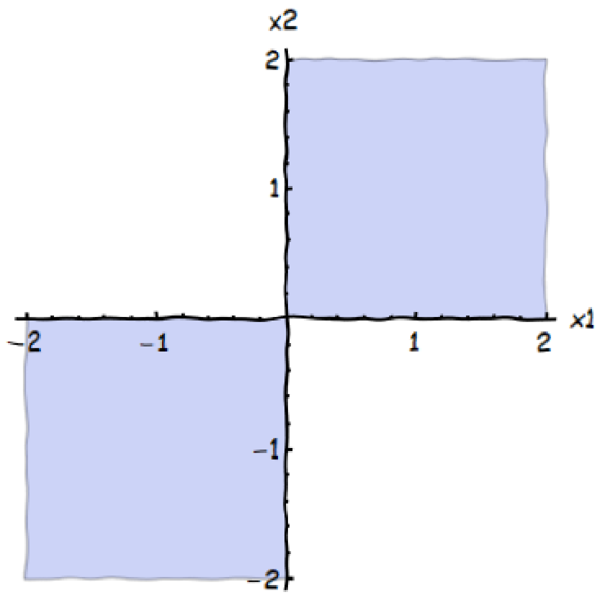


Problem 11. *Find as many subspaces in $\mathbb{R}^2$ as you can.*

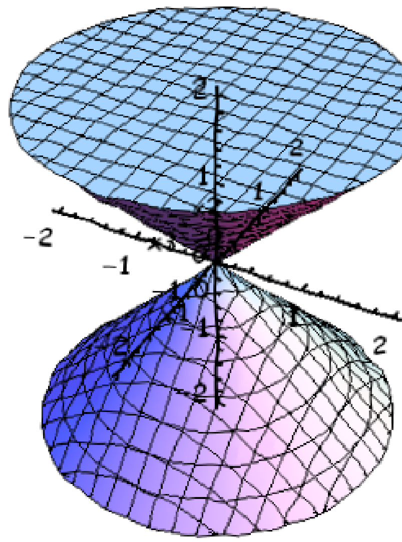
Think of this at home.

Example 12. Is one of the following a subspace of $\mathbb{R}^2$?





Example 13. Is this a subspace of $\mathbb{R}^3$?



Math 415 - Lecture 10

Span is a subspace, Null Space

Wednesday September 16th 2015

Textbook: Chapter 2.1, 2.2.

Suggested practice exercises: Chapter 2.1: 3, 21, 28. Chapter 2.2: 33 and additional exercises in this lecture note.

Khan Academy videos: Linear Subspaces, Introduction to the Null Space of a Matrix, Calculating the Null Space of a Matrix

1 Review of vector space and subspace

- A vector space is a set of vectors which can be added and scaled (without leaving the space!); subject to the "usual" rules.
- The set of all polynomials of degree up to 2 is a vector space. Why?

$$\begin{aligned}[a_0 + a_1t + a_2t^2] + [b_0 + b_1t + b_2t^2] &= \\ &[(a_0+b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2] \\ r[a_0 + a_1t + a_2t^2] &= [(ra_0) + (ra_1)t + (ra_2)t^2]\end{aligned}$$

  Note how it "works" just like $\mathbb{R}^3$.

- The set of all polynomials of degree exactly 2 is **not** a vector space. Why?

$$\underbrace{[1 + 4t + t^2]}_{\text{degree } 2} + \underbrace{[3 - t - t^2]}_{\text{degree } 2} = \underbrace{[4 + 3t]}_{\text{NOT degree } 2}$$

- Easy test: Is the zero vector in the set? (If not, then it's **not** a vector space.)

Example 1. Let V be the set of all function $f : \mathbb{R} \to \mathbb{R}$. Is V a vector space?

Solution. Yes! Adding of functions f and g:

$$f(x) + g(x) = (f + g)(x)$$

so $f(x) + g(x)$ is in V.

Note that, once more, this definition is "component-wise". Scalar multiplication works the same way.

Definition. A **subspace** of a vector space V is a subset H of V that has three properties:

1. The zero vector of V is in H.
2. For each $\mathbf{u}$ and $\mathbf{v}$ are in H, $\mathbf{u}+\mathbf{v}$ is in H. (In this case we say H is closed under vector addition.)
3. For each $\mathbf{u}$ in H and each scalar c, $c\mathbf{u}$ is in H. (In this case we say H is closed under scalar multiplication.)

Problem 2. *Find as many subspaces in $\mathbb{R}^2$ as you can.*

2 A Shortcut for Determining Subspaces

Definition. Recall that *span* $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ is the collection of all vectors that can be written as

$$x_1\mathbf{v_1} + x_2\mathbf{v_2} + \cdots + x_p\mathbf{v_p},$$

where $x_1, x_2, \ldots, x_p$ are scalars.

Theorem 1. *If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}$ are in a vector space V, then span $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ is a subspace of V.*

Example 3. Is $V = \left\{ \begin{bmatrix} a+2b \\ 2a-3b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$? Why or why not?

Solution. Write vectors in V as:

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

So $V = \mathit{span}\,\{\mathbf{v_1}, \mathbf{v_2}\}$ where

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

and therefore V is a subspace of $\mathbb{R}^2$ by the previous theorem.

Example 4. Is $H = \left\{ \begin{bmatrix} a+2b \\ a+1 \\ a \end{bmatrix} : a, b \in \mathbb{R} \right\}$ a subspace of $\mathbb{R}^3$? Why or why not?

Solution. No! H does not contain the zero vector. In other words,

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

cannot equal the zero vector for any choice of a or b.

Example 5. Is the set H of all matrices of the form $\begin{bmatrix} 2a & b \\ 3a+b & 3b \end{bmatrix}$ a subspace of M_{2x2}?

Solution. Yes!

$$H = span \left\{ \begin{bmatrix} 2 & 0 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 3 \end{bmatrix} \right\}.$$

Since H can be written as a span, it's a subspace of M_{2x2}.

Problem 6. *Determine which of the following sets are subspaces and give reasons:*

1. $W_1 = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a - 2b = c, 4a + 2c = 1 \right\}.$

2. $W_2 = \left\{ \begin{bmatrix} a-b \\ c \\ a+c \\ a-2b-c \end{bmatrix} : a, b, c \in \mathbb{R} \right\}.$

3. $W_3 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} : a \cdot b \geq 0 \right\}.$

3 Null Spaces

Definition. The **nullspace** of an $m \times n$ matrix A, written as $\text{Nul}(A)$, is the set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

$$\text{Nul}(A) = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}.$$

Theorem 2. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to the system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.*

Proof: $\text{Nul}(A)$ is a subset of $\mathbb{R}^n$ since A has n columns. We have to verify properties (a), (b), and (c) of the definition of a subspace.

Property (a): Show that $\mathbf{0}$ is in $\mathrm{Nul}(A)$.

$$A\mathbf{0} = \mathbf{0}.$$

and

$$A \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\text{in } \mathbb{R}^n} = \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{\text{in } \mathbb{R}^m}$$

Property (b): If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathrm{Nul}(A)$, show that $\mathbf{u}+\mathbf{v}$ is also in $\mathrm{Nul}(A)$. Suppose $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Then

$$A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Property (c): If $\mathbf{u}$ is in $\mathrm{Nul}(A)$ and c is a scalar, show that $c\mathbf{u}$ is also in $\mathrm{Nul}(A)$. Suppose $A\mathbf{u} = \mathbf{0}$. Then

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}.$$

Let's restate the theorem.

Theorem 3. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to the system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.*

Remark.
- Since properties (a), (b), and (c) hold, $\mathrm{Nul}(A)$ is a subspace of $\mathbb{R}^n$.
- Since $\mathrm{Nul}(A)$ is a subspace, it is closed under linear combinations. You can add solutions of $A\mathbf{x} = 0$ and get a new solution! This is very important. Not true for $A\mathbf{x} = \mathbf{b}$ for $b \neq 0$. Here you cannot add solutions!
- Solving $A\mathbf{x} = \mathbf{0}$ yields an explicit description of $\mathrm{Nul}(A)$.

Example 7. Find and explicit description of $\mathrm{Nul}(A)$ where

$$A = \begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 13 & 0 & 3 \end{bmatrix}.$$

Solution. We want to find all the solutions to $A\mathbf{x} = \mathbf{0}$. So we need to do Gaussian elimination on the augmented matrix $\left[A \mid \mathbf{0}\right]$.

$$\left[A \mid \mathbf{0}\right] = \left[\begin{array}{ccccc|c} 3 & 6 & 6 & 3 & 9 & 0 \\ 6 & 12 & 13 & 0 & 3 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccccc|c} 3 & 6 & 6 & 3 & 9 & 0 \\ 0 & 0 & 1 & -6 & -15 & 0 \end{array}\right] \longrightarrow$$

$$\left[\begin{array}{ccccc|c} 3 & 6 & 0 & 39 & 99 & 0 \\ 0 & 0 & 1 & -6 & -15 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccccc|c} 1 & 2 & 0 & 13 & 33 & 0 \\ 0 & 0 & 1 & -6 & -15 & 0 \end{array}\right] = \left[U \mid \mathbf{0}\right].$$

$$[U \mid \mathbf{0}\ ] = \left[\begin{array}{ccccc|c} 1 & 2 & 0 & 13 & 33 & 0 \\ 0 & 0 & 1 & -6 & -15 & 0 \end{array}\right].$$

This corresponds to the solution:

$$\begin{aligned} x_1 &= -2x_2 - 13x_4 - 33x_5 \\ x_3 &= 6x_4 + 15x_5. \end{aligned}$$

Write this as a linear combination:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_2 - 13x_4 - 33x_5 \\ x_2 \\ 6x_4 + 15x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix}.$$

So each vector in Nul(A) looks like:

$$x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix}.$$

Thus,

$$\text{Nul}(A) = \mathit{span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

In other words,

$$\text{Nul}\left(\begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 13 & 0 & 3 \end{bmatrix} \right) = \mathit{span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Remark. If Nul$(A) \neq \{\mathbf{0}\}$, then the number of vectors in the spanning set for Nul(A) equals the number of free variables in $A\mathbf{x} = \mathbf{0}$.
In this example, we had 3 free variables (x_2, x_4, and x_5) so there were 3 vectors in the spanning set for Nul(A). More about this later!

Math 415 - Lecture 11

Column space, Solution to $A\mathbf{x} = b$

Friday September 18th 2015

Textbook: Chapter 2.1, 2.2.

Suggested practice exercises: Chapter 2.1: 3, 21, 28. Chapter 2.2: 33 and additional exercises at the end of this lecture.

Khan Academy videos: Introduction to the Null Space of a Matrix, Calculating the Null Space of a Matrix, Column Space of a Matrix

1 Review

Definition. The **nullspace** of an $m \times n$ matrix A, written as $Nul(A)$, is the set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

$$Nul(A) = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}.$$

Theorem 1. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions $\mathbf{x}$ to the system $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$.*

For example

$$\text{Nul}\left(\begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 13 & 0 & 3 \end{bmatrix}\right) = \text{Nul}\left(\begin{bmatrix} 1 & 2 & 0 & 13 & 33 \\ 0 & 0 & 1 & -6 & -15 \end{bmatrix}\right).$$

This corresponds to the solution:

$$\begin{aligned} x_1 &= -2x_2 - 13x_4 - 33x_5 \\ x_3 &= 6x_4 + 15x_5. \end{aligned}$$

Write this as a linear combination:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_2 - 13x_4 - 33x_5 \\ x_2 \\ 6x_4 + 15x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix}.$$

This means that

$$\text{Nul}(A) = span \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Remark. If $\text{Nul}(A) \neq \{\mathbf{0}\}$, then the number of vectors in the spanning set for $\text{Nul}(A)$ equals the number of free variables in $A\mathbf{x} = \mathbf{0}$.
In this example, we had 3 free variables (x_2, x_4, and x_5) so there were 3 vectors in the spanning set for $\text{Nul}(A)$. More about this later!

2 Column Spaces

Definition. The **column space**, written as $Col(A)$, of an $m \times n$ matrix A is the set of all linear combinations of the columns of A. If $A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{bmatrix}$,

then $Col(A) = \text{Span}\{\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_n}\}$.

Example 1.

- If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $Col(A) = \text{Span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix})$. This is all of $\mathbb{R}^2$!
- If $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, then $Col(A) = \text{Span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}) = \text{Span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$. This is the x_1 axis in $\mathbb{R}^2$!
- If $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, then $Col(A) = \text{Span}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix})$. This is the zero subspace of $\mathbb{R}^2$!

Theorem 2. *The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.*

Why is it a subspace? Because it is a Span!

Remark. If A is $m \times n$ (m rows, n columns) then

- $Col(A)$ is a subspace of the output space $\mathbb{R}^m$.
- $Nul(A)$ is a subspace of the input space $\mathbb{R}^n$.

Theorem 3. *Let A be an $m \times n$ matrix.* $\mathbf{b}$ *is in $Col(A)$ iff there is an* $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ *in $\mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$.*

Proof. Suppose $A\mathbf{x} = \mathbf{b}$. Then

$$\mathbf{b} = A\mathbf{x} = \underbrace{x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_n\mathbf{a_n}}_{(\textit{lin. comb. of } \mathbf{a_1}, \ldots, \mathbf{a_n})}.$$

□

Example 2. Find a matrix A such that $W = Col(A)$ where

$$W = \left\{ \begin{bmatrix} x - 2y \\ 3y \\ x + y \end{bmatrix} : x, y \in \mathbb{R} \right\}.$$

Solution.

$$\begin{bmatrix} x - 2y \\ 3y \\ x + y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

So

$$W = \text{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} \right\} = Col \left(\begin{bmatrix} 1 & -2 \\ 0 & 3 \\ 1 & 1 \end{bmatrix} \right).$$

Therefore

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

3 $Nul(A)$ and solutions to $A\mathbf{x} = \mathbf{b}$

Theorem 4. *Let A be an $m \times n$ matrix, let $\mathbf{b} \in \mathbb{R}^m$, and let $\mathbf{x_p} \in \mathbb{R}^n$ such that*

$$A\mathbf{x_p} = \mathbf{b}.$$

Then the set of solutions $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ is exactly

$$\mathbf{x_p} + Nul(A).$$

So every solution of $A\mathbf{x} = \mathbf{b}$ is of the form

$$\mathbf{x_p} + \mathbf{x_n}$$

where $\mathbf{x_n}$ is some vector in $Nul(A)$.

Proof. Let $\mathbf{x_p} \in \mathbb{R}^n$ such that $A\mathbf{x_p} = \mathbf{b}$. Suppose $\mathbf{x}$ is also in $\mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$. Then

$$A(\mathbf{x} - \mathbf{x_p}) = A\mathbf{x} - A\mathbf{x_p} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Therefore, $\mathbf{x} - \mathbf{x_p} = \mathbf{x_n}$ is in $Nul(A)$, so $\mathbf{x} = \mathbf{x_p} + \mathbf{x_n}$. □

Remark. We often call $\mathbf{x_p}$ a *particular solution* of $A\mathbf{x} = \mathbf{b}$. The theorem then says that every solution to $A\mathbf{x} = \mathbf{b}$ is the sum of one particular solution $\mathbf{x_p}$ and all the solutions to $A\mathbf{x} = \mathbf{0}$ (the null space).

Example 3. Let $A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 5 \\ 5 \end{bmatrix}$. Solve $A\mathbf{x} = \mathbf{b}$.

Step 1 : Reduce $A\mathbf{x} = \mathbf{b}$ to $U\mathbf{x} = \mathbf{c}$.

$$\begin{bmatrix} A & | & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 1 \\ 2 & 6 & 9 & 7 & 5 \\ -1 & -3 & 3 & 4 & 5 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 1 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 6 & 6 & 6 \end{array}\right]$$

$$\to \left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 1 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

$$U\mathbf{x} = \mathbf{c}.$$

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

Step 2 : Find a particular solution to $U\mathbf{x} = \mathbf{c}$.

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

Could pick any value for the free variables (x_2 and x_4). Trick: Set them both to 0. Then

$$3x_3 = 3 \;\Rightarrow\; x_3 = 1.$$
$$x_1 + 3x_3 = 1 \;\Rightarrow\; x_1 = -2.$$

So $\mathbf{x_p} = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ is a particular solution to $A\mathbf{x} = \mathbf{b}$.

Step 3 : Find **all** the solutions to $A\mathbf{x} = \mathbf{0}$ to find $Nul(A)$.

$$\begin{bmatrix} U & | & \mathbf{0} \end{bmatrix} = \left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 0 \\ 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \to$$

$$\left[\begin{array}{cccc|c} 1 & 3 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Every vector in $Nul(A)$ is of the form

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

Step 4 : To find all the solutions to $A\mathbf{x} = \mathbf{b}$, add a particular solution $\mathbf{x_p}$ to the null space of A. So the set of solutions is

$$\mathbf{x_p} + Nul(A).$$

$$\begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \mathrm{Span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

and each solution $\mathbf{x}$ is of the form

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

Remark.

- If A is a matrix with echelon form U, then $Nul(A) = Nul(U)$. Why? Because $Nul(A)$ is the set of solutions of $Ax = 0$, which is the same as the space of solutions of $Ux = 0$ (That is the point of echelon form matrices!) which is $Nul(U)$.
- Not true that $Col(A) = Col(U)$! Why?

Example 4. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Then $U = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$.

$$Col(A) = \mathrm{Span}(\begin{bmatrix} 1 \\ 2 \end{bmatrix}), \quad Col(U) = \mathrm{Span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$$

Additional Exercises

1. True or false?
   (i) The solutions to $A\mathbf{x} = \mathbf{0}$ form a vector space. True. This is the null space $Nul(A)$.
   (ii) The solutions to $A\mathbf{x} = \mathbf{b}$ form a vector space. False, unless $\mathbf{b} = 0$.
2. Find an explicit description for $Nul(A)$ where

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

3. Show that the given set W is a subspace (by showing that W is the column space or null space of some matrix A) or find a specific example that shows that W is not a subspace.

   (i) $W_1 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 5x - 1 = y + 2z \right\}$.

   (ii) $W_2 = \left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : a = 2b + c,\ 2a = c - 3d \right\}$.

4. Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$. Find a smallest spanning set for $W = Col(A)$. Find a matrix B such that $W = Nul(B)$.

5. Let $B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Find a smallest spanning set for $W = Nul(A)$. Find a matrix B such that $W = Col(B)$

Math 415 - Lecture 12

Linear independence

Monday September 21st 2015

Textbook reading: Section 2.3

Suggested practice exercises: Section 2.3: 1, 2, 3, 4, 5,7, 8, 9

Khan Academy video: Introduction to Linear Independence, More on linear independence, Span and Linear Independence Example,

Strang lecture: Independence, Basis, and Dimension

* Exam 1 (7-8:15 pm Tuesday September 29):

* Rooms:

  - 213 Gregory Hall: AD3, ADG, ADU
  - 151 Loomis: ADC, ADD, ADL, ADM
  - 100 Gregory Hall: ADE, ADF, ADN, ADO
  - 66 Library: ADH, ADP, ADQ, ADX
  - 141 Loomis: AD1, AD2, ADS, ADT, ADW, ADZ
  - 100 MSEB: AD4, ADV, ADY, ADI, ADR
  - 150 ASL: ADA, ADB, ADJ, ADK

  MSEB is the Materials Science and Engineering Building. ASL is the Animal Science Lab.

* Conflicts: The conflict exams are at 8:00-9:20AM and 9:30-10:50AM on the same day. Email your TA with your reason for needing a conflict, and your preferred time to sign up for the conflict exam.

  The deadline for signing up for a conflict is a week before (September 22).

1 Linear independence

Review.

- Span $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_m}\}$ is the set of all linear combinations
$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_m\mathbf{v_m}.$$
- Span $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_m}\}$ is a vector space.
- $\text{Col}(A) = \text{Span}(\mathbf{a_1}, \mathbf{a_2}, \dots, \mathbf{a_n})$, if $A = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n}\end{bmatrix}$. In this case
$$\mathbf{b} \in \text{Col}(A) \iff \mathbf{b} = A\mathbf{x} \text{ for some} \mathbf{x} \in \mathbb{R}^n.$$

Today we want to think how *big* the span of a bunch of vectors is. Is it a line, or a plane or ....

Example 1. Is Span $\left\{ \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix} \right\}$ equal to $\mathbb{R}^2$?

Solution. To answer the question translate to linear systems. Recall that the span is equal to
$$\{b \colon b = A\mathbf{x}\} = \left\{ \begin{bmatrix}1 & 2\\1 & 2\end{bmatrix}\mathbf{x} : \mathbf{x} \in \mathbb{R}^2 \right\}.$$
Hence, the span is equal to $\mathbb{R}^2$ if and only if the system with augmented matrix
$$\left[\begin{array}{cc|c}1 & 2 & b_1\\1 & 2 & b_2\end{array}\right]$$
is consistent for all b_1, b_2.

To check consistency use Gaussian elimination:
$$\left[\begin{array}{cc|c}1 & 2 & b_1\\1 & 2 & b_2\end{array}\right] \longrightarrow \left[\begin{array}{cc|c}1 & 2 & b_1\\0 & 0 & b_2 - b_1\end{array}\right]$$
When is this system consistent? The system is only consistent if $b_2 - b_1 = 0$. Hence, the span does not equal all of $\mathbb{R}^2$. The span is a line instead of a plane!

Example 2. Is Span $\left\{ \begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\1\\3\end{bmatrix} \right\}$ equal to $\mathbb{R}^3$?

Solution. Recall that the span is equal to
$$\{b \colon b = A\mathbf{x}\} = \left\{ \begin{bmatrix}1 & 1 & -1\\1 & 2 & 1\\1 & 3 & 3\end{bmatrix}\mathbf{x} : \mathbf{x} \in \mathbb{R}^3 \right\}.$$
Hence, the span is equal to $\mathbb{R}^3$ if and only if the system with augmented matrix
$$\left[\begin{array}{ccc|c}1 & 1 & -1 & b_1\\1 & 2 & 1 & b_2\\1 & 3 & 3 & b_3\end{array}\right]$$
is consistent for all b_1, b_2, b_3.

To check consistency use Gaussian elimination:

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & b_1 \\ 1 & 2 & 1 & b_2 \\ 1 & 3 & 3 & b_3 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & -1 & b_1 \\ 0 & 1 & 2 & b_2 - b_1 \\ 0 & 2 & 4 & b_3 - b_1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & -1 & b_1 \\ 0 & 1 & 2 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{array}\right]$$

When is this system consistent? The system is only consistent if $b_3 - 2b_2 + b_1 = 0$. Hence, the span does not equal all of $\mathbb{R}^3$.

- What went wrong?
  Well, the three vectors that span satisfy a *relation*:

$$\begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = -3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

- Hence, $\text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}.$

- We are going to say that the three vectors are **linearly dependent** because they satisfy the (non trivial) relation

$$-3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \mathbf{0}.$$

Definition. Vectors $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are said to be **linearly independent** if the equation

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

has only the trivial solution (namely, $x_1 = x_2 = \cdots = x_p = 0$).

Likewise, $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are said to be **linearly dependent** if there exist coefficients $x_1, \ldots, x_p$, not all zero, such that

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

This is called a *non trivial relation* (when not all coefficient are zero.)

Example 3.

- Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ independent?
- If possible, find a linear dependence relation among them.

Solution. We need to check whether the equation

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

has a non trivial solution. The three vectors are independent if and only if there are no free variables for the system

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

To find out, we reduce the matrix to echelon form:

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

Since there is a column without a pivot, we do have a free variable. Hence, the three vectors are not linearly independent. To find a linear dependence relation we solve this system.

Initial steps of Gaussian elimination are as before:

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 0 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

x_3 is free. $x_2 = -2x_3$, and $x_1 = 3x_3$. Hence, for any x_3,

$$3x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2x_3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since we are only interested in one linear combination, we can set, say, $x_3 = 1$:

$$3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

2 Linear independence of matrix columns

- Note that a linear dependence relation, such as

$$3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \mathbf{0},$$

can be written in matrix form as

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

- Hence, each linear dependence relation among the columns of a matrix A corresponds to a solution to $A\mathbf{x} = \mathbf{0}$. The Null space determines (in)dependence!

Theorem 1. *Let A be an $m \times n$ matrix.*
The columns of A are linearly independent.
$\iff$ *$A\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = \mathbf{0}$.*
$\iff$ *$Nul(A) = \{\mathbf{0}\}$*
$\iff$ *A has n pivots.* $\iff$ *there are no free variables for $A\mathbf{x} = \mathbf{0}$.*

Example 4. Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$ independent?

Solution. Put the vectors in a matrix, and produce an echelon form:

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

Since each column contains a pivot, there are no free variables and the three vectors are independent. These vectors span $\mathbb{R}^3$.

Example 5. Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ independent?

Solution. Put the vectors in a matrix and produce an echelon form:

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

Since the last column does not contain a pivot, there is a free variable and the three vectors are linearly dependent. They span a plane.

3 Special cases

- A set of a single non-zero vector $\{\mathbf{v_1}\}$ is always linearly independent. Why? Because $x_1\mathbf{v_1} = \mathbf{0}$ only for $x_1 = 0$.

- A set of two vectors $\{\mathbf{v_1}, \mathbf{v_2}\}$ is linearly independent if and only if neither of the vectors is a multiple of the other.

  Why? Because if $x_1\mathbf{v_1} + x_2\mathbf{v_2} = \mathbf{0}$ with, say, $x_2 \neq 0$, then $\mathbf{v_2} = -\frac{x_1}{x_2}\mathbf{v_1}$.

- A set of vectors $\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ containing the zero vector is linearly dependent.

  Why? Because if, say, $\mathbf{v_1} = \mathbf{0}$, then $\mathbf{v_1} + 0\mathbf{v_2} + \cdots + 0\mathbf{v_p} = \mathbf{0}$.

- If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. In other words:

  Any set $\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ of vectors in $\mathbb{R}^n$ is linearly dependent if $p > n$.

  Why? Let A be the matrix with columns $\mathbf{v_1}, \dots, \mathbf{v_p}$. This is a $n \times p$ matrix.

  The columns are linearly independent if and only if each column contains a pivot.

  If $p > n$, then the matrix can have at most n pivots.

  Thus not all p columns can contain a pivot.

  In other words, the columns have to be linearly dependent.

Example 6. Let $A = \begin{bmatrix} * & * & * \\ * & * & * \end{bmatrix}$ be a two by three matrix. We want to count the free variables for $A\mathbf{x} = \mathbf{0}$. How many pivots can there be? How many free variables? Are the columns of A independent?

4 Additional exercises

With the least amount of work possible, decide which of the following sets of vectors are linearly independent.

(a) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 6 \\ 4 \end{bmatrix} \right\}$

Linearly independent, because the two vectors are not multiples of each other.

(b) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}$

Linearly independent, because it is a single non-zero vector.

(c) Columns of $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 8 & 7 & 6 \end{bmatrix}$

Linearly dependent, because these are more than 3 (namely, 4) vectors in $\mathbb{R}^3$.

(d) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$

Linearly dependent, because the set includes the zero vector.

Math 415 - Lecture 13

Basis and Dimension

Wednesday September 23rd 2015

Textbook reading: Chapter 2.3

Suggested practice exercises: Chapter 2.3 Exercise 1, 2, 3, 5, 6, 9, 11, 16, 19, 20, 22, 27.

Khan Academy video: Introduction to Linear Independence, More on linear independence, Span and Linear Independence Example, Basis of a Subspace

Strang lecture: Independence, Basis, and Dimension

* Exam 1 (7-8:15 pm Tuesday September 29):

* Rooms:

  - 213 Gregory Hall: AD3, ADG, ADU
  - 151 Loomis: ADC, ADD, ADL, ADM
  - 100 Gregory Hall: ADE, ADF, ADN, ADO
  - 66 Library: ADH, ADP, ADQ, ADX
  - 141 Loomis: AD1, AD2, ADS, ADT, ADW, ADZ
  - 100 MSEB: AD4, ADV, ADY, ADI, ADR
  - 150 ASL: ADA, ADB, ADJ, ADK

  MSEB is the Materials Science and Engineering Building. ASL is the Animal Science Lab.

* Conflicts: You should have signed up for a conflict exam by now.

* No Discussion Sections next week.

* No Class on Wednesday next week.

1 Review

- Vectors $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are *linearly **Dependent*** if

$$x_1\mathbf{v_1} + x_2\mathbf{v_2} + \cdots + x_p\mathbf{v_p} = \mathbf{0},$$

  and not all the coefficients are zero.

- The columns of A are linearly **IN**dependent $\iff$ each column of A contains a pivot $\iff$ there are no free variables for $A\mathbf{x} = \mathbf{0}$.

- Are the vectors $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\1\\3\end{bmatrix}$ independent?

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

  So no, they are dependent! (Coeff's for instance $x_3 = 1, x_2 = -2, x_1 = 3$)

- Any set of 11 vectors in $\mathbb{R}^{10}$ is linearly dependent. Why?

Definition 1. In a list of vectors $(\mathbf{v_1}, \ldots, \mathbf{v_p})$ in a vector space V we call $\mathbf{v_k}$ *redundant* if v_k is a linear combination of the previous vectors. In this case $\mathrm{Span}(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k-1}}, \mathbf{v_k}) = \mathrm{Span}(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k-1}})$, i.e., you can delete the redundant vector and get the same span.

Example 2. Let $\mathbf{v_1} = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix}1\\0\\2\end{bmatrix}$, $\mathbf{v_3} = \begin{bmatrix}2\\0\\3\end{bmatrix}$. Are there redunant vectors?

Solution. Since $\mathbf{v_3} = \mathbf{v_1} + \mathbf{v_2}$, $\mathbf{v_3}$ is redundant and $\mathrm{Span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}) = \mathrm{Span}(\mathbf{v_1}, \mathbf{v_2})$.

Today we are going to study sets of vectors without redundant elements.

2 A Basis of a Vector Space

Definition. A set of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ in V is a **basis** of V if

- $V = \mathrm{Span}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$, and
- the vectors $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are linearly independent.

Fact: $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ in V is a basis of V if and only if every vector $\mathbf{w}$ in V can be uniquely expressed as $\mathbf{w} = c_1\mathbf{v_1} + \cdots + c_p\mathbf{v_p}$.

Fact: A basis is a *minimal spanning set*: the elements of the basis span V but you cannot delete any of these elements and still get all of V. There are no redundant vectors.

Example 3. Let $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Show that $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$ is a basis of $\mathbb{R}^3$. (It is called the **standard basis**.)

Solution.

- Clearly, $\text{Span}\,\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\} = \mathbb{R}^3$.
- $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$ are independent, because $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has a pivot in each column, no free variables. Note that we can not delete one of $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ and still get all of $\mathbb{R}^3$.

Definition. V is said to have **dimension** p if it has a basis consisting of p vectors.

This definition makes sense because if V has a basis of p vectors, then every basis of V has p vectors. Why? (Think of $V = \mathbb{R}^3$.) A basis of $\mathbb{R}^3$ cannot have more than 3 vectors, because any set of 4 or more vectors in $\mathbb{R}^3$ is linearly dependent. A basis of $\mathbb{R}^3$ cannot have less than 3 vectors,because 2 vectors span at most a plane. (Challenge: can you think of an argument that is more "rigorous"?)

Example 4. $\mathbb{R}^3$ has dimension 3. Indeed, the standard basis

$$\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

has three elements. Likewise, $\mathbb{R}^n$ has dimension n.

Example 5. Not all vectors spaces have a finite basis. For instance, the vector space of all polynomials has *infinite dimension*. Its standard basis is $1, t, t^2, t^3, \ldots$

Why?

Solution. This is indeed a basis, because any polynomial can be written as a unique linear combination:

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n$$

for some n.

Recall that vectors in V form a **basis** of V if

- They span V.
- They are linearly independent.

These are two conditions. If we know the dimension of V, we only need to check *one* of these two conditions:

Theorem 1. *Suppose that V has dimension d.*

- *A set of d vectors in V are a basis if they span V.*
- *A set of d vectors in V are a basis if they are linearly independent.*

Why?

Solution.

- If the d vectors were not independent, then $d-1$ of them would still span V. In the end, we would find a basis of less than d vectors.
- If the d vectors would not span V, then we could add another vector to the set and have $d+1$ independent ones.

Example 6. Are the following sets a basis for $\mathbb{R}^3$?

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

(b) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}$

Solution.

(a) No, the set has less than 3 elements.

(b) No, the set has more than 3 elements.

Example 7. (c) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \right\}$. Is this a basis?

Solution. (c) The set has 3 elements. Hence, it is a basis if and only if the vectors are independent.

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

Since each column contains a pivot, the three vectors are independent. Hence, this is a basis for $\mathbb{R}^3$.

Example 8. Let P_2 be the space of polynomials of degree at most 2.

- What is the dimension of P_2?
- Is $\{t,\ 1-t,\ 1+t-t^2\}$ a basis of P_2?

Solution.

- The standard basis for P_2 is $\{1, t, t^2\}$. This is indeed a basis because every polynomial

$$a_0 + a_1 t + a_2 t^2$$

can clearly be written as a linear combination of $1, t, t^2$ in a unique way. Hence, P_2 has dimension 3.

- The set $\{t,\ 1-t,\ 1+t-t^2\}$ has 3 elements. Hence, it is a basis if and only if the three polynomials are linearly independent. We need to check whether
$$x_1 t + x_2(1-t) + x_3(1+t-t^2) = 0$$
has only the trivial solution $x_1 = x_2 = x_3 = 0$. We get the equations
$$\begin{aligned} x_2 + x_3 &= 0 \\ x_1 - x_2 + x_3 &= 0 \\ -x_3 &= 0 \end{aligned}$$
which clearly only have the trivial solution. (If you don't see it, solve the system!) Hence, $\{t, 1-t, 1+t-t^2\}$ is a basis of P_2.

3 Shrinking and Exanding Sets of Vectors

We can find a basis for $V = \text{Span}\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ by discarding, if necessary, some of the vectors in the spanning set.

Example 9. Produce a basis of $\mathbb{R}^2$ from the vectors

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}, \mathbf{v_3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Solution. Here, we notice that $\mathbf{v_2} = -2\mathbf{v_1}$. The remaining vectors $\{\mathbf{v_1}, \mathbf{v_3}\}$ are a basis for $\mathbb{R}^2$, because the two vectors are clearly linearly independent.

Example 10. Produce a basis of $\mathbb{R}^2$ from the vector

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Solution. $\mathbf{v_1}$ is independent. But is does not span $\mathbb{R}^2$. For instance $\mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is not in the span of $\mathbf{v_1}$. Let's add it! Then $\text{Span}(\mathbf{v_1}, \mathbf{v_2})$ is all of $\mathbb{R}^2$ and we found a basis.

4 Checking Our Understanding

Example 11. Subspaces of $\mathbb{R}^3$ can have dimension 0, 1, 2, or 3.

- The only 0-dimensional subspace is $\{\mathbf{0}\}$.
- A 1-dimensional subspace is of the form $\text{Span}\{\mathbf{v}\}$ where $\mathbf{v} \neq \mathbf{0}$. These subspaces are lines through the origin.
- A 2-dimensional subspace is of the form $\text{Span}\{\mathbf{v}, \mathbf{w}\}$ where $\mathbf{v}$ and $\mathbf{w}$ are not multiples of each other. These subspaces are planes through the origin.

- The only 3-dimensional subspace is $\mathbb{R}^3$ itself.

True or false?

1. Suppose that V has dimension n. Then any set in V containing more than n vectors must be linearly dependent. True.

2. The space P_n of polynomials of degree at most n has dimension $n+1$. True. A basis is $\left\{1, t, t^2, \ldots, t^n\right\}$.

3. The vector space of functions $f : \mathbb{R} \to \mathbb{R}$ is infinite-dimensional. True. A still-infinite-dimensional subspace are the polynomials.

4. Consider $V = \operatorname{Span}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$. If one of the vectors, say $\mathbf{v_k}$, in the spanning set is a linear combination of the remaining ones, then the remaining vectors still span V. True. $\mathbf{v_k}$ is not adding anything new.

Math 415 - Lecture 14

Null space and Column space basis

Friday September 25th 2015

Textbook reading: 2.4

Suggested practice exercises: Chapter 2.4 Exercise 1, 2, 3, 4, 21

Khan Academy video: Null Space and Column Space Basis, Dimension of the Null Space, Dimension of the Column Space

Strang lecture: Independence, Basis, and Dimension

* Exam 1 (7-8:15 pm Tuesday September 29):

* Rooms: look on Moodle.

* Conflicts: if you have a conflict you should have received an email about it. If not, talk to me after class.

* No Discussion Sections next week.

* No Class on Wednesday next week.

* The Exam will be part multiple choice. Bring pencils and erasers! Also bring ID.

* The material for the exam covers the lectures upto and including Lecture 12 (last Monday), and this weeks worksheet and quiz.

1 Review

- $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is a **basis** of V if the vectors
  - span V, and
  - are independent.
- The **dimension** of V is the number of elements in a basis.
- The columns of A are linearly independent $\iff$ each column of A contains a pivot. $\iff$ there are no free variables. $\iff$ $\text{Nul}(A) = 0$.

2 Warmup

Example 1. Find a basis and the dimension of

$$W = \left\{ \begin{bmatrix} a+b+2c \\ 2a+2b+4c+d \\ b+c+d \\ 3a+3c+d \end{bmatrix} : a,b,c,d \in \mathbb{R} \right\}.$$

Solution. First, note that

$$W = span \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Is $dim\ W = 4$? No, because the third vector is the sum of the first two.

Suppose we did not notice ...

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

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

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

Not a pivot in every column, hence the 4 vectors are dependent.

Remark. Not necessary here, but to get a relation, solve $A\mathbf{x} = \mathbf{0}$. Set free variable $x_3 = 1$. Then $x_4 = 0$, $x_2 = -x_3 = -1$ and $x_1 = -x_2 - 2x_3 = -1$. The relation is

$$-\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0}.$$

Precisely what we "noticed" to begin with.

Hence, a basis for W is $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $dim\ W = 3$. It follows from the echelon form that these vectors are independent.

Remark. Every set of linearly independent vectors can be extended to a basis.

In other words, let $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ be linearly independent vectors in V. If V has dimension d, then we can find vectors $\mathbf{v_{p+1}}, \ldots, \mathbf{v_d}$ such that $\{\mathbf{v_1}, \ldots, \mathbf{v_d}\}$ is a basis of V.

Example 2. Consider

$$H = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

- Give a basis for H. What is the dimension of H?
- Extend the basis of H to a basis of $\mathbb{R}^3$.

Solution.

- The vectors are independent. By definition, they span H. Therefore, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for H. In particular, $dim\ H = 2$.
- $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is not a basis for $\mathbb{R}^3$. Why? Because a basis for $\mathbb{R}^3$ needs to contain 3 vectors. Or because, for instance, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is not in H. So just add this (or any other) missing vector! By construction, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is independent. Hence, this is automatically a basis of $\mathbb{R}^3$.

3 Bases for Null Spaces

To find a basis for $Nul(A)$:

- find the parametric form of the solutions to $A\mathbf{x} = \mathbf{0}$.
- express solutions $\mathbf{x}$ as a linear combination of vectors with the free variables as coefficients;
- these vectors form a basis of $Nul(A)$.

Example 3. Find a basis for $Nul(A)$ with

$$A = \begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 15 & 0 & 3 \end{bmatrix}.$$

Solution.

$$\begin{bmatrix}3 & 6 & 6 & 3 & 9\\ 6 & 12 & 15 & 0 & 3\end{bmatrix} \longrightarrow \begin{bmatrix}3 & 6 & 6 & 3 & 9\\ 0 & 0 & 3 & -6 & -15\end{bmatrix}$$
$$\longrightarrow \begin{bmatrix}1 & 2 & 2 & 1 & 3\\ 0 & 0 & 1 & -2 & -5\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 2 & 0 & 5 & 13\\ 0 & 0 & 1 & -2 & -5\end{bmatrix}$$

The solutions to $A\mathbf{x} = \mathbf{0}$ are:

$$\mathbf{x} = \begin{bmatrix}-2x_2 - 5x_4 - 13x_5\\ x_2\\ 2x_4 + 5x_5\\ x_4\\ x_5\end{bmatrix} = x_2\begin{bmatrix}-2\\1\\0\\0\\0\end{bmatrix} + x_4\begin{bmatrix}-5\\0\\2\\1\\0\end{bmatrix} + x_5\begin{bmatrix}-13\\0\\5\\0\\1\end{bmatrix}$$

Hence, $Nul(A) = span\left\{\begin{bmatrix}-2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-5\\0\\2\\1\\0\end{bmatrix}, \begin{bmatrix}-13\\0\\5\\0\\1\end{bmatrix}\right\}$. These vectors are independent. (Can you see why?)

Hence, $\left\{\begin{bmatrix}-2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-5\\0\\2\\1\\0\end{bmatrix}, \begin{bmatrix}-13\\0\\5\\0\\1\end{bmatrix}\right\}$ is a basis for $Nul(A)$.

Remark. If A is a matrix, $Nul(A)$ has a basis vector for each free variable. So the *dimension* of $Nul(A)$ is equal to the number of free variables!

4 Basis for Column Space

Recall that the columns of A are independent $\iff$ $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (namely, $\mathbf{x} = \mathbf{0}$) $\iff$ A has no free variables.

Theorem 1. *A basis for $Col(A)$ is given by the pivot columns of A.*

Example 4. Find a basis for $Col(A)$ with

$$A = \begin{bmatrix}1 & 2 & 0 & 4\\ 2 & 4 & -1 & 3\\ 3 & 6 & 2 & 22\\ 4 & 8 & 0 & 16\end{bmatrix}.$$

Solution. $Col(A) = Span\left(\begin{bmatrix}1\\2\\3\\4\end{bmatrix}, \begin{bmatrix}2\\4\\6\\8\end{bmatrix}, \begin{bmatrix}0\\-1\\2\\0\end{bmatrix}, \begin{bmatrix}4\\3\\22\\16\end{bmatrix}\right)$. But there could be redundant vectors among these generators. Use row operations to find the redundant vectors.

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 2 & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

Note that for U we have column $\mathbf{u}_2 = 2\mathbf{u_1}$ and $\mathbf{u_4} = 4\mathbf{u_1} + 5\mathbf{u_3}$. The same is true for the columns of A! Therefore $\mathbf{a_2}$ and $\mathbf{a_4}$ are redundant. The leftover columns are independent. This are the pivot columns, the first and third. Hence, a basis for $Col(A)$ is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix} \right\}$. This argument works in general: the pivot columns of A form a basis for $Col(A)$. So the dimension of $Col(A)$ is the number of pivots.

Remark. If A has echelon form U then any relation for the columns of U:

$$x_1\mathbf{u_1} + \cdots + x_n\mathbf{u_n} = 0$$

also holds for the columns of A:

$$x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n} = 0,$$

for the *same* scalars x_i. **Why?**

Solution. Because the relation for the columns of U is in matrix form

$$Ux = 0,$$

but this is equivalent to $Ax = 0$, which is equivalent to the relation between the columns of A.

Warning : For the basis of $Col(A)$, you have to take the columns of A, not the columns of an echelon form. Row operations do not preserve the column space.

Example 5. Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$. Then the RREF of A is $U = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$.

The second column of both A and U are redundant, so

$$Col(A) = Span(\mathbf{a_1}, \mathbf{a_2}) = Span(\mathbf{a_1}) = Span(\begin{bmatrix} 1 \\ 2 \end{bmatrix}),$$

$$Col(U) = Span(\mathbf{u_1}, \mathbf{u_2}) = Span(\mathbf{u_1}) = Span(\begin{bmatrix} 1 \\ 0 \end{bmatrix})$$

So $Col(A)$ and $Col(U)$ are **NOT** equal. In contrast $Nul(A)$ and $Nul(U)$ **ARE** equal.

5 Checking Our Understanding

True or false?

1. Suppose that V has dimension n. Then any set in V containing more than n vectors must be linearly dependent. True.

2. The space P_n of polynomials of degree at most n has dimension $n+1$. True. A basis is $\left\{1, t, t^2, \ldots, t^n\right\}$.

3. The vector space of functions $f : \mathbb{R} \to \mathbb{R}$ is infinite-dimensional. True. A still-infinite-dimensional subspace are the polynomials.

4. Consider $V = \mathit{span}\left\{\mathbf{v_1}, \ldots, \mathbf{v_p}\right\}$. If one of the vectors, say $\mathbf{v_k}$, in the spanning set is a linear combination of the remaining ones, then the remaining vectors still span V. True. $\mathbf{v_k}$ is not adding anything new.

Math 415 - Lecture 15

The Four Fundamental Subspaces, the Fundamental Theorem of Linear Algebra, Linear Transformations

Monday September 28th 2015

Textbook: Chapter 2.4, 2.6

Suggested Practice Exercise: Chapter 2.4 Exercise 1, 2, 3, 4, 7, 10, 18, 20, 21, 22, 27, 32, 37 Chapter 2.6 Exercise 5, 6, 7, 36, 37

Khan Academy Video: Linear Transformation, Linear Transformations as Matrix Vector Products, Linear Transformation Examples: Rotations in $\mathbb{R}^2$

Strang lectures: Lecture 9: Independence, Basis, and Dimension Lecture 10: The Four Fundamental Subspaces Lecture 30: Linear Transformations

* Exam 1 (7-8:15 pm Tuesday September 29):

* Rooms: look on Moodle.

* Conflicts: if you have a conflict you should have received an email about it. If not, talk to me after class.

* No Discussion Sections next week.

* No Class on Wednesday next week.

* The Exam will be part multiple choice. Bring pencils and erasers! Also bring ID.

* The material for the exam covers the lectures upto and including Lecture 12 (last Monday), and this weeks worksheet and quiz.

1 Review

1.1 Basis for the Null Space

- $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is a basis of V if the vectors span V and are independent.

- To find a basis for $Nul(A)$, solve $A\mathbf{x} = \mathbf{0}$.

$$\begin{bmatrix} 3 & 6 & 6 & 3 \\ 6 & 12 & 15 & 0 \end{bmatrix} \xrightarrow[RREF]{} \begin{bmatrix} \boxed{1} & 2 & 0 & 5 \\ 0 & 0 & \boxed{1} & -2 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} -2x_2 - 5x_4 \\ x_2 \\ 2x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

So a basis for $Nul(A)$ is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}$

1.2 Basis for the Column space.

- To find a basis for $Col(A)$, take the pivot columns of A.

$$\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \longrightarrow \begin{bmatrix} \boxed{1} & 2 & 0 & 4 \\ 0 & 0 & \boxed{-1} & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So a basis for $Col(A)$ is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix} \right\}$

Question. Why do we take columns of A and not columns of the Echelon form?

1.3 The Column spaces of A and U.

Question. Why do we take columns of A and not columns of the Echelon form?

- Row operations do not preserve the column space. For example, $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \xrightarrow[R1 \leftrightarrow R2]{} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- On the other hand, row operations do preserve the null space. Why?
  Remember, we can do row operations to solve systems like $A\mathbf{x} = \mathbf{0}$.

2 Rank and Dimensions

2.1 Dimension of Column and Null Space

Definition. The rank of a matrix A is the number of pivots it has.

Theorem 1. ***Rank-Nullity Theorem*** *Let A be an $m \times n$ matrix of rank r. Then*

dim $Col(A) = r$ Why?

> *A basis for $Col(A)$ is given by the pivot columns of A.*

dim $Nul(A) = n - r$ is the number of free variables of A. Why?

> *In our method for finding a basis for $Nul(A)$, each free variable corresponds to an element in the basis.*

dim $Col(A) +$ dim $Nul(A) = n$ Why?

> *Each of the n columns of A either contains a pivot or corresponds to a free variable.*

3 The Four Fundamental Subpaces

3.1 Two Spaces we know

Let A be a matrix. We already know two fundamental subspaces:

- The column space of A and
- The null space of A

There are two more!

3.2 Row Space and Left Null Space

Definition.

- The row space of A is the column space of A^T. $Col(A^T)$ is spanned by the columns of A^T and these are the rows of A (but transposed, to turn into columns!).
- The left null space of A is the null space of A^T. Why is it called the "left" null space?

$$\begin{array}{lll} & \text{Suppose } \mathbf{x} \in Nul(A^T). & \text{Thus,} \\ \iff & A^T\mathbf{x} = \mathbf{0}. & \text{Take transposes of both sides:} \\ \iff & (A^T\mathbf{x})^T = \mathbf{0}^T. & \text{So,} \\ \iff & \mathbf{x}^T A = \mathbf{0}. & \end{array}$$

Therefore, $\mathbf{x} \in Nul(A^T) \iff \mathbf{x}^T A = \mathbf{0}$.

Example 1. Find a basis for $Col(A)$ and $Col(A^T)$ if

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix}.$$

Solution. We need to compute an echelon form of A to find a basis for $Col(A)$. Then we might compute an echelon form of A^T to find a basis for $Col(A^T)$. However, an echelon form of A will allow us to find a basis for both $Col(A)$ and $Col(A^T)$.

Instead of doing twice the work, we only need to find an echelon form of A.

$$\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 2 & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We identify the pivot columns:

$$\longrightarrow \begin{bmatrix} \boxed{1} & 2 & 0 & 4 \\ 0 & 0 & \boxed{\text{-1}} & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

So $r = 2$ for A and a basis for $Col(A)$ is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix} \right\}$.

Remark. Key idea: The row space is preserved by elementary row operations.

Remember, $Col(A) \neq Col(U)$ because we did row operations. However, the row spaces are the same! i.e.

$$Col(A^T) = Col(U^T)$$

$$U = \begin{bmatrix} \boxed{1} & 2 & 0 & 4 \\ 0 & 0 & \boxed{\text{-1}} & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad U^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 4 & -5 & 0 & 0 \end{bmatrix}$$

In particular, a basis for $Col(A^T)$ is given by $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ -5 \end{bmatrix} \right\}$.

3.3 Fundamental Theorem of Linear Algebra (Part 1)

Theorem 2. *Let A be an $m \times n$ matrix with rank r.*

- *dim* $Col(A) = r$ *(subspace of $\mathbb{R}^m$)*
- *dim* $Col(A^T) = r$ *(subspace of $\mathbb{R}^n$)*
- *dim* $Nul(A) = n - r$ *(subspace of $\mathbb{R}^n$)*
- *dim* $Nul(A^T) = m - r$ *(subspace of $\mathbb{R}^m$)*

Remark. The column and row space always have the same dimension. In other words, A and A^T have the same rank. (i.e. same number of pivots). Why?

It's easy to see this for a matrix in echelon form.

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 7 \end{bmatrix}, \quad A^T = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 2 & 7 \end{bmatrix}$$

(3 pivot columns in A, 3 non-zero columns in A^T.) But it's not as obvious for a random matrix.

4 Coordinates

4.1 Why Bases?

What is the point of having a *basis* for a vector space V?

- **Dimension!** If you have a basis $\mathcal{B} = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p})$ for V, you know that the dimension of V is p, so that you have an idea of the Size of V. In particular, if V has dimension 0 V is just the zero vector space.
- **Coordinates!** If $w \in V$ and $\mathcal{B} = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p})$ is a basis for V, we can express w in this basis. This means that we can write (uniquely!)

  $$w = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_p\mathbf{v_p}.$$

  We call the scalars $c_1, c_2, \ldots, c_p$ the *coordinates* of w with respect to the basis $\mathcal{B}$.

We are going to organize the coordinates in a convenient package.

4.2 Coordinate Vectors

Definition. If $w \in V$ and $\mathcal{B} = (\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_p})$ is a basis for V, the **coordinate vector** of w with respect to the basis $\mathcal{B}$ is

$$w_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_p \end{bmatrix}, \quad \text{if } w = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \dots + c_p\mathbf{v_p}.$$

So w is a vector in some vector space, but it's coordinate vector is always a column vector in $\mathbb{R}^p$, if $\dim(V) = p$. Why is the coordinate vector useful? Computations in V can be translated in computations in the familiar vector space $\mathbb{R}^p$.

Let $V = \mathbb{R}^2$, $\mathcal{B} = (\mathbf{b_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix})$ and $\mathbf{w} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. What is the coordinate vector of $\mathbf{w}$? Express in the basis:

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} = c_1\mathbf{b_1} + c_2\mathbf{b_2} = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Therefore

$$\mathbf{w}_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Geometrically: this means that to reach $\mathbf{w}$ walk 1 unit along the $\mathbf{b_1}$ basis vector and 2 units along the $\mathbf{b_2}$ basis vector.

4.3 Example with polynomials

Let $V = P_2$, the vector space of polynomials of the form $a_0 + a_1 t + a_2 t^2$. Let $\mathcal{B} = (\mathbf{b_1} = 1, \mathbf{b_2} = t, \mathbf{b_3} = t^2)$ be the obvious basis of P_2. Let $\mathbf{w} = 1 + 2t + 3t^2$. What is the coordinate vector of $\mathbf{w}$ with respect to basis $\mathcal{B}$? Express $\mathbf{w}$ in terms of the basis:

$$\mathbf{w} = c_1\mathbf{b_1} + c_2\mathbf{b_2} + c_3\mathbf{b_3} = c_1 1 + c_2 t + c_3 t^2 = 1 + 2t + 3t^2.$$

Hence

$$\mathbf{w}_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

What if we take another basis? Say take $\bar{\mathcal{B}} = (t^2, t, 1)$. (Different order!). Then

$$\mathbf{w}_{\bar{\mathcal{B}}} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

4.4 Standard Coordinate Vectors

Let $V = \mathbb{R}^3$ and let $E = (\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3})$ be the standard basis. If $\mathbf{w} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ what is the coordinate vector with respect to the standard basis? Express in the basis:

$$\mathbf{w} = c_1\mathbf{e_1} + c_2\mathbf{e_2} + c_3\mathbf{e_3} = c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Hence

$$\mathbf{w}_E = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = w!$$

So the coordinate vector with respect to the standard basis is just the vector itself!

5 Linear Transformations

Let V and W be vector spaces.

Definition. A map $T : V \to W$ is a linear transformation if

$$T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$ and all $c, d \in \mathbb{R}$. In other words, a linear transformation respects addition and scaling.

Remark. It follows immediately that

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$
- $T(c\mathbf{x}) = cT(\mathbf{x})$
- $T(\mathbf{0}) = \mathbf{0}$ (because $T(\mathbf{0}) = T(0 \cdot \mathbf{0}) = 0 \cdot T(\mathbf{0}) = \mathbf{0}$)

5.1 Some examples

Example 2. Let $V = \mathbb{R}, W = \mathbb{R}$. Then the map $f(x) = 3x$ is linear. Why?

If $x, y \in \mathbb{R}$, then $f(ax + by) = 3(ax + by) = a \cdot 3x + b \cdot 3y = af(x) + bf(y)$. What about the function $g(x) = 2x - 2$? Is this a linear transformation?

Example 3. Let A be an $m \times n$ matrix. Then the map $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. Why? Because matrix multiplication is linear.

$$A(c\mathbf{x} + d\mathbf{y}) = cA\mathbf{x} + dA\mathbf{y}$$

The left-hand side is $T(c\mathbf{x} + d\mathbf{y})$ and the right-hand side is $cT(\mathbf{x}) + dT(\mathbf{y})$.

Example 4. Let P_n be the vector space of all polynomials of degree at most n. Consider the map $T : P_n \to P_{n-1}$ given by

$$T(p(t)) = \frac{d}{dt}p(t).$$

This map is linear! Why?

Because differentiation is linear:

$$\frac{d}{dt}\left[ap(t) + bq(t)\right] = a\frac{d}{dt}p(t) + b\frac{d}{dt}q(t).$$

The left-hand side is $T(ap(t) + bq(t))$ and the right-hand side is $aT(p(t)) + bT(q(t))$.

Math 415 - Lecture 16

Linear Transformations

Friday October 2nd 2015

Textbook reading: Chapter 2.6

Suggested practice exercises: Chapter 2.6: 5, 6, 7, 36, 37

Khan Academy videos: Linear Transformations / Linear Transformations as Matrix Vector Products / Linear Transformation Examples: Rotations in $\mathbb{R}^2$

Strang lecture: Lecture 30: Linear Transformations

1 Review

If $\mathcal{B} = (\mathbf{b_1}, \dots, b_p)$ is a basis for a vector space V then the coordinate vector of a vector $\mathbf{w} \in V$ is the column vector

$$\mathbf{w}_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}, \quad \text{if } \mathbf{w} = c_1\mathbf{b_1} + c_2\mathbf{b_2} + \cdots + c_p\mathbf{b_p}$$

Example 1. Let $V = \mathbb{R}^2$, $\mathcal{B} = (\mathbf{b_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix})$ and $\mathbf{w} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

Solution. Then

$$\mathbf{w} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Longrightarrow \mathbf{w}_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Geometrically: this means that to reach $\mathbf{w}$ walk 1 unit along the $\mathbf{b_1}$ basis vector and 2 units along the $\mathbf{b_2}$ basis vector.

Example 2. Still $V = \mathbb{R}^2$, $\mathcal{B} = (\mathbf{b_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix})$ a basis for V. Suppose $\mathbf{w}_{\mathcal{B}} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ is a coordinate vector with respect to the basis $\mathcal{B}$. What is the vector $\mathbf{w}$, with respect to the standard basis?

Solution. $\mathbf{w}_{\mathcal{B}} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ means that you reach $\mathbf{w}$ by walking 4 units along $\mathbf{b_1}$ and 5 units along $\mathbf{b_2}$. So

$$\mathbf{w} = 4\mathbf{b_1} + 5\mathbf{b_2} = 4\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 9 \\ -1 \end{bmatrix}$$

Remark. Translating to the standard basis is always easy. To go from the standard basis to a new basis requires solving a system of equations, so is generally harder.

2 Linear Transformations

Let V and W be vector spaces.

Definition. A map $T : V \to W$ is a linear transformation if

$$T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$ and all $c, d \in \mathbb{R}$. In other words, a linear transformation respects addition and scaling.

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$
- $T(c\mathbf{x}) = cT(\mathbf{x})$

It also sends the zero vector in V to the zero vector in W:

- $T(\mathbf{0}) = \mathbf{0}$ (because $T(\mathbf{0}) = T(0 \cdot \mathbf{0}) = 0 \cdot T(\mathbf{0}) = \mathbf{0}$)

2.1 Some examples

Example 3. Let $V = \mathbb{R}, W = \mathbb{R}$. Then the map $f(x) = 3x$ is linear. Why?

Solution. If $x, y \in \mathbb{R}$, then $f(ax + by) = 3(ax + by) = a \cdot 3x + b \cdot 3y = af(x) + bf(y)$. What about the function $g(x) = 2x - 2$? Is this a linear transformation?

Example 4. Let A be an $m \times n$ matrix. Then the map $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. Why?

Solution. Because matrix multiplication is linear.

$$A(c\mathbf{x} + d\mathbf{y}) = cA\mathbf{x} + dA\mathbf{y}$$

The left-hand side is $T(c\mathbf{x} + d\mathbf{y})$ and the right-hand side is $cT(\mathbf{x}) + dT(\mathbf{y})$.

We will argue that all linear transformations are essentially matrix multiplication!

Example 5. Let P_n be the vector space of all polynomials of degree at most n. Consider the map $T : P_n \to P_{n-1}$ given by

$$T(p(t)) = \frac{d}{dt}p(t).$$

This map is linear! Why?

Solution. Because differentiation is linear:

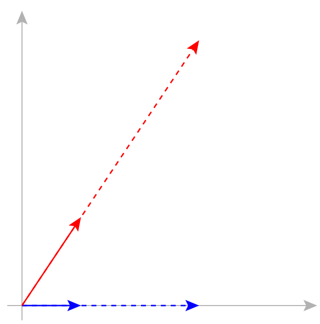
$$\frac{d}{dt}\left[ap(t) + bq(t)\right] = a\frac{d}{dt}p(t) + b\frac{d}{dt}q(t).$$

The left-hand side is $T(ap(t) + bq(t))$ and the right-hand side is $aT(p(t)) + bT(q(t))$.

3 Important Geometric Examples

Let's consider some linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ which are defined by matrix multiplication ($\mathbf{x} \mapsto A\mathbf{x}$). In fact, it turns out that all linear maps $\mathbb{R}^n \to \mathbb{R}^m$ are given by $\mathbf{x} \mapsto A\mathbf{x}$ for some $m \times n$ matrix A.

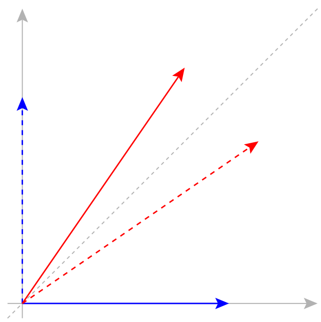
Example 6 (Stretching). The matrix $A = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$ gives the map $x \mapsto c\mathbf{x}$, It stretches every vector in $\mathbb{R}^2$ by a factor c.

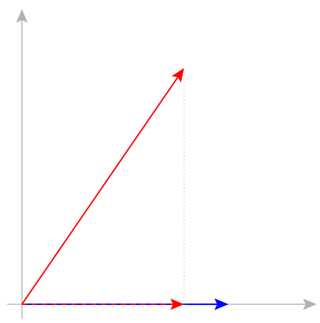


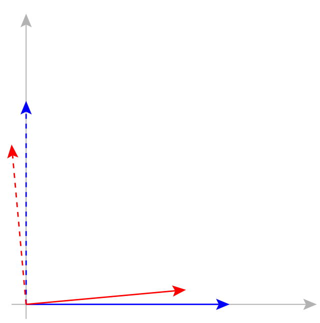
Example 7 (Reflection). The matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ x \end{bmatrix}$. It reflects every vector in $\mathbb{R}^2$ across the line $y = x$.

Example 8 (Projection.). The matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ 0 \end{bmatrix}$. It projects every vector in $\mathbb{R}^2$ onto the x-axis.

Example 9 (Rotation by 90°.). The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -y \\ x \end{bmatrix}$. It rotates every vector in $\mathbb{R}^2$ counter-clockwise by 90 degrees.







4 Representing linear maps by matrices

Motto
If you know T on a basis, you know T everywhere.

- Let $\mathbf{x_1}, \ldots, \mathbf{x_n}$ be an input basis, a basis for V. A linear map $T : V \to W$ is determined by the values $T(\mathbf{x_1}), \ldots, T(\mathbf{x_n})$.

- Why?

Take any $\mathbf{v} \in V$. It can be written as $\mathbf{v} = c_1\mathbf{x_1} + \cdots + c_n\mathbf{x_n}$ because $\{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ is a basis and hence spans V. Hence by the linearity of T:

$$T(\mathbf{v}) = T(c_1\mathbf{x_1} + \cdots + c_n\mathbf{x_n}) = c_1T(\mathbf{x_1}) + \cdots + c_nT(\mathbf{x_n}).$$

So we know how to write $T(\mathbf{v})$ as long as we know $T(\mathbf{x_1}), \ldots, T(\mathbf{x_n})$!

4.1 Standard Basis Coordinates

Example 10. Suppose $T : R^2 \to \mathbb{R}^3$ is a linear map so that

$$T\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix} \text{ and } T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\-3\end{bmatrix}$$

What is

$$T\begin{bmatrix}1\\2\end{bmatrix}?$$

Solution.

$$T\begin{bmatrix}1\\2\end{bmatrix} = T\begin{bmatrix}1\\0\end{bmatrix} + 2T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix} + 2\begin{bmatrix}0\\0\\-3\end{bmatrix} = \begin{bmatrix}1\\2\\-3\end{bmatrix}$$

Let us look at the example again. The linear transformation was given on the standard basis by

$$T\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix} \text{ and } T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\-3\end{bmatrix}$$

Let's take a general input vector for T:

$$T\begin{bmatrix}x\\y\end{bmatrix} = xT\begin{bmatrix}1\\0\end{bmatrix} + yT\begin{bmatrix}0\\1\end{bmatrix} = x\begin{bmatrix}1\\2\\3\end{bmatrix} + y\begin{bmatrix}0\\0\\-3\end{bmatrix}$$

A linear combination! Linear combination is matrix multiplication!

$$T\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}1 & 0\\2 & 0\\3 & -3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

Hence calculating T is multiplying by the matrix $A = \begin{bmatrix}1 & 0\\2 & 0\\3 & -3\end{bmatrix}$.

Summary: The linear transformation

$$T\colon \mathbb{R}^2 \to \mathbb{R}^3, \quad T\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix}, \quad T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\-3\end{bmatrix}$$

is the same as multiplying by the matrix

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

We say that the linear transformation T is represented by the matrix A, or that A is the *coordinate matrix* of the linear transformation T, (with respect to the standard bases).

Example 11. Let $T_\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the "rotation over α radians (counterclockwise)" map. So $T_\alpha(\mathbf{x})$ is the vector obtained by rotating $\mathbf{x}$ over angle α. Can you find a matrix so that $T_\alpha(\mathbf{x}) = A_\alpha\mathbf{x}$?

Solution. We just need to find what happens under rotation to the standard basic vectors. If you draw a picture you see that

$$T_\alpha\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}\cos(\alpha)\\ \sin(\alpha)\end{bmatrix}, \quad T_\alpha\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}-\sin(\alpha)\\ \cos(\alpha)\end{bmatrix},$$

So our matrix is $A_\alpha = \begin{bmatrix}\cos(\alpha) & -\sin(\alpha)\\ \sin(\alpha) & \cos(\alpha)\end{bmatrix}$. This is called the rotation matrix for angle α. It allows you to calculate the rotation of any vector!

Theorem 1 (Linear Transformation is Matrix Multiplication, Standard basis)**.** *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is a matrix A such that*

$$T(\mathbf{x}) = A\mathbf{x}, \quad \textit{for all } \mathbf{x} \in \mathbb{R}^n.$$

Explicitly,

$$A = \begin{bmatrix}T(e_1) & T(e_2) & \dots & T(e_n)\end{bmatrix},$$

where $e_1, e_2, \dots, e_n$ is the standard basis of $\mathbb{R}^n$.

Proof. We can write $\mathbf{x} = x_1e_1 + x_2e_2 + \cdots + x_ne_n$. Then

$$\begin{aligned} T(\mathbf{x}) &= T(x_1e_1 + x_2e_2 + \cdots + x_ne_n) = \\ &= x_1T(e_1) + x_2T(e_2) + \cdots + x_nT(e_n) = \\ &= A\mathbf{x}. \end{aligned}$$

□

Example 12. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}2a+3b-c\\ -a+b+2c\end{bmatrix}$. What is the matrix representing T (with respect to the standard bases)?

Solution. First think about the size of A. It must be 2×3.Then calculate the columns of A:

$$T\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix}, \quad \text{Why? } a = 1, b = c = 0,$$

$$T\begin{bmatrix}0\\1\\0\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}, \quad \text{Why? } a=0, b=1, c=0,$$

Example continued.

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

$$T\begin{bmatrix}0\\0\\1\end{bmatrix}=\begin{bmatrix}-1\\2\end{bmatrix}, \quad \text{Why? } a=0=b, c=1,$$

So $A=\begin{bmatrix}2&3&-1\\-1&1&2\end{bmatrix}$. Check:

$$A=\begin{bmatrix}2&3&-1\\-1&1&2\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix}=\begin{bmatrix}2a+3b-c\\-a+b+2c\end{bmatrix}=T\begin{bmatrix}a\\b\\c\end{bmatrix}$$

5 Nonstandard Bases

Untill now we have used the standard bases to describe $T\colon \mathbb{R}^n\to\mathbb{R}^m$. Often it is useful to use other bases.

Example 13. Let $T\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}3a+1b\\1a+3b\end{bmatrix}$. Then the matrix of T is $A=\begin{bmatrix}3&1\\1&3\end{bmatrix}$. But let us use, instead of the standard basis, another basis adapted to T. Put

$$\mathbf{b_1}=\begin{bmatrix}1\\-1\end{bmatrix}, \quad \mathbf{b_2}=\begin{bmatrix}1\\1\end{bmatrix}.$$

What is the coordinate matrix for T with respect to $\mathcal{B}=(\mathbf{b_1},\mathbf{b_2})$?

Solution. What do we want? We want to find a matrix B that relates the coordinate vectors (w.r.t. basis $\mathcal{B}$) of input vector $\mathbf{x}$ and and output vector $T(x)$:

$$T(x)_{\mathcal{B}}=Bx_{\mathcal{B}}.$$

This matrix B has columns $T(\mathbf{b_1})_{\mathcal{B}}$ and $T(\mathbf{b_2})_{\mathcal{B}}$. So let us calculate

$$T(\mathbf{b_1})=T\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}3&1\\1&3\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix}=2\begin{bmatrix}1\\-1\end{bmatrix}=2\mathbf{b_1},$$

$$T(\mathbf{b_2})=T\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}3&1\\1&3\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}=4\begin{bmatrix}1\\1\end{bmatrix}=4\mathbf{b_2}$$

This means that the coordinate matrix with respect to $\mathcal{B}$ is simply

$$B=\begin{bmatrix}2&0\\0&4\end{bmatrix}$$

Summary: The linear transformation $T\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3a+1b \\ 1a+3b \end{bmatrix}$ has with respect to the standard basis the coordinate matrix A, but with respect to the other basis $\mathcal{B}$ the coordinate B:

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

The linear transformation T is geometrically clear in the $\mathcal{B}$ basis: T is just stretching vectors by a factor 2 along $\mathbf{b_1}$ and by a factor 4 along $\mathbf{b_2}$. So using the standard basis T is an obscure operation on vectors, but using the basis $\mathcal{B}$ it becomes clear. You can say that $\mathcal{B}$ is a basis adapted to T.

6 Additional Problems

- Suppose $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 2 & 4 & 7 & 8 & 1 \end{bmatrix}$. Find the dimensions and a basis for all four fundamental subspaces of A.

- Suppose A is 5×5 and v is a vector in $\mathbb{R}^5$ which is not a linear combination of the columns of A. What can you say about the number of solutions to $A\mathbf{x} = \mathbf{0}$?

- Let T be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}, \quad T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

  What is $T\left(\begin{bmatrix} 0 \\ 4 \end{bmatrix}\right)$?

Math 415 - Lecture 17

Linear Transformations

Monday October 5th 2015

Textbook reading: Chapter 2.6

Suggested practice exercises: same as lecture 16

1 Review

- A map $T : V \to W$ between vector spaces is **linear** if
  - $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$.
  - $T(c\mathbf{x}) = cT(\mathbf{x})$.
- If $\mathbf{x_1}, \dots, \mathbf{x_n}$ is a basis for V, then T is determined by the values $T(\mathbf{x_1}), \dots, T(\mathbf{x_n})$:

$$T(\mathbf{v}) = T(c_1\mathbf{x_1} + \cdots + c_n\mathbf{x_n}) = c_1T(\mathbf{x_1}) + \cdots + c_nT(\mathbf{x_n}).$$

- Let A be an $m \times n$ matrix.
  - $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is linear.
  - *Every* linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is of the form $T(\mathbf{x}) = A\mathbf{x}$.
- $T : \mathbb{P}_n \to \mathbb{P}_{n-1}$ defined by $T(p(t)) = p'(t)$ is linear. What is its “matrix”?

2 Nonstandard Bases

Until now we have used the standard bases to describe $T\colon \mathbb{R}^n \to \mathbb{R}^m$. Often it is useful to use other bases.

Theorem 1 (Linear Transformation is Matrix Multiplication)**.** *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $\mathcal{B} := (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ be a basis of $\mathbb{R}^n$ and let $\mathcal{C} := (\boldsymbol{w}_1, \dots, \boldsymbol{w}_m)$ be a basis of $\mathbb{R}^m$. Then there is a matrix B such that*

$$T(\mathbf{x})_{\mathcal{C}} = B\mathbf{x}_{\mathcal{B}}, \quad \textit{for all } \mathbf{x} \in \mathbb{R}^n.$$

Explicitly,

$$B = \begin{bmatrix} T(\boldsymbol{v}_1)_{\mathcal{C}} & \dots & T(\boldsymbol{v}_n)_{\mathcal{C}}. \end{bmatrix},$$

Example 1. Let $T\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3a+1b \\ 1a+3b \end{bmatrix}$. Then the matrix of T is $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. But let us use, instead of the standard basis, another basis adapted to T. Put

$$\mathbf{b_1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{b_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

What is the coordinate matrix for T with respect to $\mathcal{B} = (\mathbf{b_1}, \mathbf{b_2})$?

Solution. What do we want? We want to find a matrix B that relates the coordinate vectors (w.r.t. basis $\mathcal{B}$) of input vector $\mathbf{x}$ and and output vector $T(x)$:

$$T(x)_{\mathcal{B}} = Bx_{\mathcal{B}}.$$

This matrix B has columns $T(\mathbf{b_1})_{\mathcal{B}}$ and $T(\mathbf{b_2})_{\mathcal{B}}$. So let us calculate

$$T(\mathbf{b_1}) = T\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2\mathbf{b_1},$$

$$T(\mathbf{b_2}) = T\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4\mathbf{b_2}$$

This means that the coordinate matrix with respect to $\mathcal{B}$ is simply

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

Summary: The linear transformation $T\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3a+1b \\ 1a+3b \end{bmatrix}$ has with respect to the standard basis the coordinate matrix A, but with respect to the other basis $\mathcal{B}$ the coordinate B:

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

The linear transformation T is geometrically clear in the $\mathcal{B}$ basis: T is just stretching vectors by a factor 2 along $\mathbf{b_1}$ and by a factor 4 along $\mathbf{b_2}$. So using the standard basis T is an obscure operation on vectors, but using the basis $\mathcal{B}$ it becomes clear. You can say that $\mathcal{B}$ is a basis adapted to T.

3 Matrices for... Polynomials?

Let P_n be the vector space of polynomials of degree at most n.

Example 2. Consider the map $T : P_2 \to P_1$ given by

$$T(p(t)) = \frac{d}{dt}p(t).$$

Describe T by a matrix.

Solution. Wait, what?! We can't multiply a polynomial by a matrix! Use coordinate vectors instead.

Pick bases $\mathcal{A} = (1, t, t^2)$ for P_2 and $\mathcal{B} = (1, t)$ for P_1. Find a matrix D that does to the coordinate vectors what T does to the polynomials.

$$T(2 + 3t + 4t^2) = 3 + 8t$$

$$D \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix}$$

$$T(t^2) = 2t$$

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

Formally,

$$D \cdot (f_{\mathcal{A}}) = T(f)_{\mathcal{B}}$$

From the equation

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

The third column of D is $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$. What are the remaining two columns?

$$T(1) = 0 \implies D \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$T(t) = 1 \implies D \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$T(t^2) = 2t \implies D \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

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

Check Take $f(t) = 2 - t + 3t^2$. Then the coordinate vector for $f(t)$ is

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

Then

$$D\begin{bmatrix}2\\-1\\3\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}\begin{bmatrix}2\\-1\\3\end{bmatrix} = \begin{bmatrix}-1\\6\end{bmatrix}.$$

On the other hand $T(f(t)) = f'(t) = -1 + 6t$, with coordinate vector $\begin{bmatrix}-1\\6\end{bmatrix}$.

4 Matrices for Linear Transformations

Let's organize this. Let $T : V \to W$ be a linear transformation, $\mathcal{A} = \{\mathbf{x_1}, \dots, \mathbf{x_n}\}$ be an *input basis* for V, and $\mathcal{B} = \{\mathbf{y_1}, \dots, \mathbf{y_m}\}$ an *output basis* for W. Each vector in V has a coordinate vector in $\mathbb{R}^n$, each vector in W has a coordinate vector in $\mathbb{R}^m$. T now corresponds to a matrix from $\mathbb{R}^n$ to $\mathbb{R}^m$.

$$\begin{tikzcd}
\text{vector in } V \arrow[r, "\text{apply } T"] \arrow[d, "\text{write in coordinates}"'] & \text{vector in } W \arrow[d, "\text{write in coordinates}"] \\
\text{coordinate vector in } \mathbb{R}^n \arrow[r, "\text{multiply by } A"] & \text{coordinate vector in } \mathbb{R}^m
\end{tikzcd}$$

In the last example this was

$$T(2 + 3t + 4t^2) = 3 + 8t$$

$$A\begin{bmatrix}2\\3\\4\end{bmatrix} = \begin{bmatrix}3\\8\end{bmatrix}$$

Definition. Let $\mathcal{A} = \{\mathbf{x_1}, \dots, \mathbf{x_n}\}$ be a basis for V, and $\mathcal{B} = \{\mathbf{y_1}, \dots, \mathbf{y_m}\}$ a basis for W. The matrix $T_{\mathcal{BA}}$ representing T with respect to these bases

- has n columns (one for each of the $\mathbf{x_j}$),
- the j-th column is the coordinate vector of $T(\mathbf{x_j})$ in the basis $\mathcal{B}$.

$$T_{\mathcal{BA}} = \begin{bmatrix}T(\mathbf{x_1})_{\mathcal{B}} & T(\mathbf{x_2})_{\mathcal{B}} & \dots & T(\mathbf{x_n})_{\mathcal{B}}\end{bmatrix}$$

Example 3. Give the matrix for $T : P_2 \to P_1$ given by

$$T(p(t)) = \frac{d}{dt}p(t).$$

in the bases $\mathcal{A} = (1, t, t^2)$ and $\mathcal{B} = (1, t)$.

Solution.

$$T_{\mathcal{BA}} = \begin{bmatrix}T(1)_{\mathcal{B}} & T(t)_{\mathcal{B}} & T(t^2)_{\mathcal{B}}\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}$$

Example 4. Recall the map T given by $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ x \end{bmatrix}$. (It reflects every vector in $\mathbb{R}^2$ across the line $y = x$.)

(a) Which matrix A represents T with respect to the standard bases?

(b) Which matrix B represents T with respect to the basis $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$?

Solution. (a) $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. So $A = \begin{bmatrix} 0 & * \\ 1 & * \end{bmatrix}$. $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. So $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

(b) $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. So $B = \begin{bmatrix} 1 & * \\ 0 & * \end{bmatrix}$. $T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. So $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.
Draw a picture!

Remark. If a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is represented by the matrix A with respect to the standard bases, then $T(\mathbf{x}) = A\mathbf{x}$. Matrix multiplication corresponds to function composition! That is, if T_1, T_2 are represented by A_1, A_2, then $T_1(T_2(\mathbf{x})) = (A_1 A_2)\mathbf{x}$.

Example 5. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix}.$$

What is the matrix B representing T with respect to the following bases?

$$\mathbf{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{x_2} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \text{ for } \mathbb{R}^2, \ \mathbf{y_2} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{y_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{y_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ for } \mathbb{R}^3.$$

Solution.

$$\begin{aligned} T(\mathbf{x_1}) = T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) &= T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 10 \end{bmatrix} \\ &= 5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ \implies B &= \begin{bmatrix} 5 & * \\ -3 & * \\ 5 & * \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
T(\mathbf{x_2}) = T\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) &= -T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= -\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \\ 11 \end{bmatrix} \\
&= 7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 9\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
\implies B &= \begin{bmatrix} 5 & 7 \\ -3 & -9 \\ 5 & 4 \end{bmatrix}
\end{aligned}
$$

Remark. A matrix representing T encodes in column j the coefficients of $T(\mathbf{x_j})$ expressed as a linear combination of $\mathbf{y_1}, \ldots, \mathbf{y_m}$.

5 Recap

What is the Point? Why write $T \colon V \to W$ as a matrix?

- Replace obscure computations in V and W by transparent computations with matrices.
- Even if $T : \mathbb{R}^n \to \mathbb{R}^m$ (already have standard coordinates), T may be simpler in a different coordinate system.

Summary: Given $\mathbf{v}$ in V, want to calculate $T(\mathbf{v})$ in W. Take an input basis $\mathcal{A} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ and and output basis $\mathcal{B} = (\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m})$.

- We know $\mathbf{v}$ if we know the coordinate vector $\mathbf{v}_{\mathcal{A}}$.
- We know $T(\mathbf{v})$ if we know the coordinate vector $T(\mathbf{v})_{\mathcal{B}}$.
- So we know T if we know the matrix $T_{\mathcal{B}\mathcal{A}}$:

$$T(\mathbf{v})_{\mathcal{B}} = T_{\mathcal{B}\mathcal{A}}\mathbf{v}_{\mathcal{A}}.$$

[-.5cm]The output coordinate vector equals the matrix for T times the input coordinate vector.

Example 6. Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Let T be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix}.$$

What is the matrix A representing T with respect to the standard bases? Use that to calculate $T\begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

Solution. The standard bases are

$$\mathbf{x_1} = \begin{bmatrix}1\\0\end{bmatrix}, \mathbf{x_2} = \begin{bmatrix}0\\1\end{bmatrix}, \quad \mathbf{y_1} = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{y_2} = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{y_3} = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\begin{aligned} T(\mathbf{x_1}) = \begin{bmatrix}1\\2\\3\end{bmatrix} &= 1\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\\0\end{bmatrix} + 3\begin{bmatrix}0\\0\\1\end{bmatrix} \\ &= 1\mathbf{y_1} + 2\mathbf{y_2} + 3\mathbf{y_3} \\ T(\mathbf{x_2}) = \begin{bmatrix}4\\0\\7\end{bmatrix} &= 4\mathbf{y_1} + 0\mathbf{y_2} + 7\mathbf{y_3} \\ \implies A &= \begin{bmatrix}1 & 4\\2 & 0\\3 & 7\end{bmatrix} \end{aligned}$$

So $T\begin{bmatrix}2\\3\end{bmatrix} = A\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}14\\4\\27\end{bmatrix}$

6 Additional Problems

- Suppose $A = \begin{bmatrix}1 & 2 & 3 & 4 & 1\\2 & 4 & 7 & 8 & 1\end{bmatrix}$. Find the dimensions and a basis for all four fundamental subspaces of A.

- Suppose A is 5×5 and v is a vector in $\mathbb{R}^5$ which is not a linear combination of the columns of A. What can you say about the number of solutions to $A\mathbf{x} = \mathbf{0}$?

- Let T be the linear map such that

$$T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}1\\0\\4\end{bmatrix}, \quad T\left(\begin{bmatrix}-1\\1\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\0\end{bmatrix}.$$

  What is $T\left(\begin{bmatrix}0\\4\end{bmatrix}\right)$?

Math 415 - Lecture 18

Inner Product and Orthogonality

Wednesday October 7th 2015

Textbook reading: Ch 3.1

Suggested practice exercises: 1, 2, 4, 5, 14, 16

Khan Academy video: Vector Dot Product and Vector Length

Strang lectures: Lecture 30: Linear Transformations / Lecture 14: Orthogonality

Applications: Information retrieval

1 Review

- A **linear map** $T : V \to W$ satisfies

$$T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y}).$$

- $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is linear. A is the matrix representing T in the standard bases. For example, $T(\mathbf{e_1}) = A\mathbf{e_1}$ = first column of A.

$$\left(\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)$$

- Any $T\colon V \to W$ can be represented by a matrix.

What is the Point? Why write $T\colon V \to W$ as a matrix?

- Replace obscure computations in V and W by transparent computations with matrices.
- Even if $T : \mathbb{R}^n \to \mathbb{R}^m$ (already have standard coordinates), T may be simpler in a different coordinate system.

Summary: Given $\mathbf{v}$ in V, want to calculate $T(\mathbf{v})$ in W. Take an input basis $\mathcal{A} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ and and output basis $\mathcal{B} = (\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m})$.

- The abstract input vector $\mathbf{v}$ and the coordinate vector $\mathbf{v}_{\mathcal{A}}$ determine each other.
- The abstract output vector $T(\mathbf{v})$ and the coordinate vector $T(\mathbf{v})_{\mathcal{B}}$ determine each other.
- So we know T if we know the matrix $T_{\mathcal{B}\mathcal{A}}$:

$$T(\mathbf{v})_{\mathcal{B}} = T_{\mathcal{B}\mathcal{A}}\mathbf{v}_{\mathcal{A}}.$$

The output coordinate vector equals the matrix for T times the input coordinate vector.

Formula For the Coordinate matrix. To write $T\colon V \to W$ as a matrix, take an input basis $\mathcal{A} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ and and output basis $\mathcal{B} = (\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m})$. Then

$$T_{\mathcal{B}\mathcal{A}} = \begin{bmatrix} T(\mathbf{x_1})_{\mathcal{B}} & T(\mathbf{x_2})_{\mathcal{B}} & \ldots & T(\mathbf{x_n})_{\mathcal{B}} \end{bmatrix}$$

Example 1. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be reflection across the x-y plane, $(x,y,z) \mapsto (x,y,-z)$. Determine the matrix representing T in the basis $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

$T\colon (x,y,z) \mapsto (x,y,-z)$. So calculate

$$T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = (-1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Hence

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

Example 2. Let $T : \mathbb{P}_3 \to \mathbb{P}_2$ be the linear map given by

$$T(p(t)) = \frac{d}{dt}p(t).$$

What's the matrix A representing T in the standard bases?

Solution. The standard bases are $\{1, t, t^2, t^3\}$ for $\mathbb{P}_2$ and $\{1, t, t^2\}$ for $\mathbb{P}_2$. The matrix A has 4 columns and 3 rows.

- $T(1) = 0$, so the first column is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

- $T(t) = 1$, so the second column is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

- $T(t^2) = 2t$, so the third column is $\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$.

- $T(t^3) = 3t^2$, so the last (fourth) column is $\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$.

So the matrix A representing T in the standard bases is

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

What is $Col(A)$ and $Nul(A)$ for $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$?

Solution. $Col(A) = \mathbb{R}^3$. Every quadratic polynomial is the derivative of some cubic polynomial.

$$Nul(A) = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

The corresponding polynomial is $p(t) = 1$. That makes sense because differentiation kills constant polynomials.

Let's try differentiating $7t^3 - t + 3$ using the matrix A.

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 21 \end{bmatrix}$$

We get $-1 + 0t + 21t^2$, which is indeed the derivative of $7t^3 - t + 3$.

2 Inner Product and Distances

Definition. The **inner product** (or **dot product**) of $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = v_1 w_1 + \cdots + v_n w_n.$$

Example 3.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = 1 - 2 - 6 = -7$$

Theorem 1. *Let* $\mathbf{u}, \mathbf{v}$ *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$*, and let* c *be any scalar. Then*

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

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

(c) $(c\mathbf{u}) \cdot \mathbf{v} = c\,(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

(d) $\mathbf{u} \cdot \mathbf{u} \geq \mathbf{0}$*, and* $\mathbf{u} \cdot \mathbf{u} = \mathbf{0}$ *if and only if* $\mathbf{u} = \mathbf{0}$*.*

Definition.

- The **norm** (or **length**) of a vector $\mathbf{v} \in \mathbb{R}^n$ is

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

- The **distance** between points $v, w \in \mathbb{R}^n$ is

$$dist\,(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|.$$

Example 4. (a)

$$\left\| \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \right\| = \sqrt{1^2 + (-1)^2 + 3^2} = \sqrt{11}$$

(b)

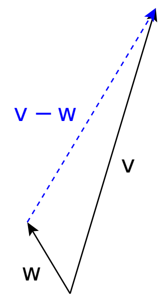
$$dist\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right) = \| \begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \end{bmatrix} \| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

3 Inner Product and Angles

We can use the dot product to compute angles too.

Theorem 2. *If* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *are linearly independent, they form an angle* θ*, and*

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$$



Example 5. What is the angle formed in $\mathbb{R}^3$ between the vectors

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}?$$

(A base jumper runs at a cliff at a 45° angle, then jumps straight away from the cliff and 45° downwards; what angle does he turn as he jumps?)

Solution.

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$$

$$\|v\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$$

$$\|w\| = \sqrt{0^2 + (-1)^2 + (-1)^2} = \sqrt{2}$$

$$v \cdot w = -1$$

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$$

$$-1 = 2\cos\theta \quad \Rightarrow \quad \cos\theta = -\frac{1}{2}$$

$$\theta = 120°$$

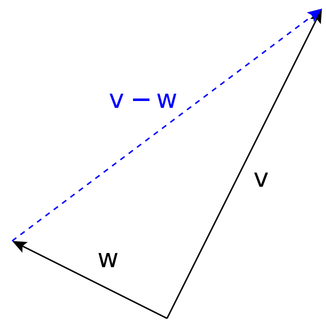
4 Orthogonal vectors

Definition. $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ are **orthogonal** if

$$\mathbf{v} \cdot \mathbf{w} = 0.$$

Remark. We write $\mathbf{v} \perp \mathbf{w}$ when $\mathbf{v}$ and $\mathbf{w}$ are orthogonal. Nonzero vectors $\mathbf{v}$, $\mathbf{w}$ are orthogonal if and only if they are perpendicular.

Nonzero vectors $\mathbf{v}$, $\mathbf{w}$ are orthogonal if and only if they are perpendicular. We can derive this from Pythagoras' theorem. $\mathbf{v} \perp \mathbf{w} \iff \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 = \|\mathbf{v} - \mathbf{w}\|^2 \iff \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) \iff \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{v} - 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} \iff \mathbf{v} \cdot \mathbf{w} = 0$



Example 6. Are the following vectors orthogonal?

(a) $\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-2\\1\end{bmatrix}$

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

So yes, they're orthogonal.

(b) $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-2\\1\\1\end{bmatrix}$

$$\begin{bmatrix}1\\2\\1\end{bmatrix} \cdot \begin{bmatrix}-2\\1\\1\end{bmatrix} = 1 \cdot (-2) + 2 \cdot 1 + 1 \cdot 1 = 1$$

So no, they're not orthogonal.

Example 7. Let $\mathbf{v} = \begin{bmatrix}2\\-1\\1\end{bmatrix}$. Is the set of vectors orthogonal to $\mathbf{v}$ a subspace of $\mathbb{R}^3$?

$$V = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{v} \cdot \mathbf{x} = 0\}$$

Solution.

$$\begin{aligned} &\mathbf{v} \cdot \mathbf{x} = 0 \\ \Leftrightarrow &\begin{bmatrix}2 & -1 & 1\end{bmatrix} \mathbf{x} = 0 \end{aligned}$$

V is just the null space of the matrix $\mathbf{v}^T = \begin{bmatrix}2 & -1 & 1\end{bmatrix}$. So yes, it is a subspace.

Definition. If V is a subspace of $\mathbb{R}^n$, a vector $\mathbf{x}$ is **orthogonal to** V if it is orthogonal to every vector in V.

Example 8. Let $V = \operatorname{span}\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$. Is $\mathbf{x} = \begin{bmatrix}-1\\1\end{bmatrix}$ orthogonal to V?

Solution.

$$\begin{bmatrix}-1\\1\end{bmatrix} \cdot \begin{bmatrix}a\\a\end{bmatrix} = -a + a = 0$$

So yes.

Math 415 - Lecture 19

Orthonormal basis, orthogonal complement

Friday October 9th 2015

Textbook reading: Ch 3.1

Suggested practice exercises: Ch 3.1: 7, 8, 9, 10, 11, 12, 14, 15, 17, 18, 19, 20, 22

Khan Academy videos: Introduction to orthonormal bases, Coordinates with respect to orthonormal bases

Strang lectures: Lec 10: The Four Fundamental Subspaces / Lec 14: Orthogonal Vectors and Subspaces

1 Review

- $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = v_1 w_1 + \cdots + v_n w_n$ is the **inner product** of $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.
  - The **length** of $\mathbf{v}$, $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
  - The **distance** between points $\mathbf{v}, \mathbf{w}$ is $\|\mathbf{v} - \mathbf{w}\|$.
- $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{v} \cdot \mathbf{w} = 0$.
  - This simple criterion is equivalent to Pythagoras' theorem.

2 Unit Vectors and Orthonormal basis

Definition. A vector $\mathbf{u} \in \mathbb{R}^n$ is called a *unit vector* if

- $\|\mathbf{u}\| = 1$, or, equivalently,
- $\mathbf{u} \cdot \mathbf{u} = 1$

Example 1. The standard basis vectors $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}$ of $\mathbb{R}^n$ are all unit vectors.

Example 2. If $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then is $\mathbf{x}$ a unit vector?

Solution. Since $\mathbf{x} \cdot \mathbf{x} = 5$ and $\|\mathbf{x}\| = \sqrt{5}$. However, $u = \frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{\mathbf{x}}{\sqrt{5}}$ *is* a unit vector. The unit vector $\mathbf{u}$ is called the *normalization* of $\mathbf{x}$.

Definition.
- A bunch of vectors $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_p}$ is called *orthogonal* if they are all nonzero and $\mathbf{x_i} \cdot \mathbf{x_j} = 0$ for $i \neq j$.
- Orthogonal vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_p}$ are called *orthonormal* if they are all unit vectors.

Example 3. Let $\mathbf{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{x_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then they are orthogonal but not orthonormal, since they are not unit vectors. We can normalize them to get a orthonormal set $\mathbf{u_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Let $\mathcal{B} := (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$ be an orthonormal basis for $\mathbb{R}^n$, so a basis consisting of unit vectors that are all perpendicular. Suppose we want to calculate the coordinates of $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x} = c_1 \mathbf{u_1} + c_2 \mathbf{u_2} + \cdots + c_n \mathbf{u_n}.$$

If this were an arbitrary basis, we would have to solve a system of equations to find the coordinates $c_1, \ldots, c_n$. Now we know that we have an orthonormal basis things are easier. Just calculate

$$\mathbf{u_1} \cdot \mathbf{x} = \mathbf{u_1} \cdot (c_1 \mathbf{u_1} + c_2 \mathbf{u_2} + \cdots + c_n \mathbf{u_n}) =$$

$$= c_1 \mathbf{u_1} \cdot \mathbf{u_1} + c_2 \mathbf{u_1} \cdot \mathbf{u_2} + \cdots + c_n \mathbf{u_1} \cdot \mathbf{u_n} = c_1$$

In the same way

$$\mathbf{u_2} \cdot \mathbf{x} = c_2, \ldots, \mathbf{u_n} \cdot \mathbf{x} = c_n$$

Example 4. $\mathbf{u_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an orthonormal basis for $\mathbb{R}^2$. Let $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

Solution. Then

$$\mathbf{x} = c_1 \mathbf{u_1} + c_2 \mathbf{u_2} = c_1 \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

for

$$c_1 = \mathbf{u_1} \cdot \mathbf{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{5}{\sqrt{2}}$$

$$c_2 = \mathbf{u_2} \cdot \mathbf{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{-1}{\sqrt{2}}$$

Theorem 1. *Let* $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ *be non-zero and mutually orthogonal. Then* $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ *are linearly independent.*

Solution. **Proof.** Suppose that

$$c_1\mathbf{v_1} + \cdots + c_n\mathbf{v_n} = \mathbf{0}.$$

Take the inner product of $\mathbf{v_1}$ on both sides.

$$\begin{aligned}\mathbf{0} &= \mathbf{v_1}\cdot(c_1\mathbf{v_1} + \cdots + c_n\mathbf{v_n}) \\ &= c_1\mathbf{v_1}\cdot\mathbf{v_1} + c_2\mathbf{v_1}\cdot\mathbf{v_2} + \cdots + c_n\mathbf{v_1}\cdot\mathbf{v_n} \\ &= c_1\mathbf{v_1}\cdot\mathbf{v_1} = c_1\|\mathbf{v_1}\|^2\end{aligned}$$

But $\|\mathbf{v_1}\| \neq 0$ and so $c_1 = 0$. Similarly, we find that $c_2 = 0, \ldots, c_n = 0$. Therefore, the vectors are independent.

3 Orthogonality and the Fundamental subspaces

Example 5. Let $A = \begin{bmatrix}1 & 2\\ 2 & 4\\ 3 & 6\end{bmatrix}$. Find $Nul(A)$ and $Col(A^T)$.

Solution. Note that $Nul(A)$ and $Col(A^T)$ both are subspace of $\mathbb{R}^2$.

$$Nul(A) = span\left\{\begin{bmatrix}-2\\ 1\end{bmatrix}\right\}, \quad Col(A^T) = span\left\{\begin{bmatrix}1\\ 2\end{bmatrix}\right\}$$

The basis vectors for the null and row space are orthogonal.

$$\begin{bmatrix}-2\\ 1\end{bmatrix}\cdot\begin{bmatrix}1\\ 2\end{bmatrix} = 0$$

Example 6. Repeat for $A = \begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 0\\ 3 & 6 & 0\end{bmatrix}$.

Solution.

$$\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 0\\ 3 & 6 & 0\end{bmatrix} \xrightarrow[RREF]{} \begin{bmatrix}1 & 2 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}$$

$$Nul(A) = span\left\{\begin{bmatrix}-2\\ 1\\ 0\end{bmatrix}\right\}, \quad Col(A^T) = span\left\{\begin{bmatrix}1\\ 2\\ 0\end{bmatrix}, \begin{bmatrix}0\\ 0\\ 1\end{bmatrix}\right\}$$

Again, the basis for the null space is orthogonal to the basis for the row space.

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

Since $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ is orthogonal to both basis vectors for the row space, it's orthogonal to *every* vector in the row space. It turns out this is true for the null and row space of any matrix A. That is, vectors in $Nul(A)$ are orthogonal to vectors in $Col(A^T)$ for *all* matrices A.

4 Fundamental Theorem of Linear Algebra (Revisited)

Definition. Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$.

- $\mathbf{v}$ is **orthogonal** to W if $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in W$. ($\iff$ $\mathbf{v}$ is orthogonal to each vector in a basis for W.)
- Another subspace V is **orthogonal** to W if every vector in V is orthogonal to W.
- The **orthogonal complement** of W is the space $W^\perp$ of all vectors that are orthogonal to W. (Show that the orthogonal complement of any vector space is also a vector space.)

Example 7. Let $V = \text{Span} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $W = \text{Span} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Then

- $V \perp W$, because every vector of V is perp to each vector in W.
- It is not true that $V^\perp = W$ since $V^\perp$ consists of *all* vectors in $\mathbb{R}^3$ perp to V. Which vectors are missing?
- $V^\perp = \text{Span}(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix})$

Example 8. In the last example, $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 3 & 6 & 0 \end{bmatrix}$. We found that

$$Nul(A) = span \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad Col(A^T) = span \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

are orthogonal subspaces. Indeed, $Nul(A)$ and $Col(A^T)$ are orthogonal complements. Why?

Solution. Because $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are orthogonal (so independent), and so they're a basis for all of $\mathbb{R}^3$.

Remark. In the last example, $Nul(A)$ and $Col(A)$ both happen to be subspaces of $\mathbb{R}^3$ (because A was a square 3×3 matrix).

$$Nul(A) = span \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad Col(A) = span \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

However, these spaces are **not** orthogonal. Why?

Solution.

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \neq 0$$

Theorem 2. *Let A be an $m \times n$ matrix of rank r.*

- *$dim\ Col(A) = r$* *(subspace of $\mathbb{R}^m$)*
- *$dim\ Col(A^T) = r$* *(subspace of $\mathbb{R}^n$)*
- *$dim\ Nul(A) = n - r$* *(subspace of $\mathbb{R}^n$)*
- *$dim\ Nul(A^T) = m - r$* *(subspace of $\mathbb{R}^m$)*
- *$Nul(A)^\perp = Col(A^T)$* *(both subspaces of $\mathbb{R}^n$) Note that $dim\ Nul(A) + dim\ Col(A^T) = n$.*
- *$Nul(A^T)^\perp = Col(A)$*

Example 9. Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$.

- $Col(A) = Span \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
- $Nul(A) = Span \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.
- $Col(A^T) = Span \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- $Nul(A^T) = Span \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

Math 415 - Lecture 20

Fundamental Theorem of Linear algebra, orthogonal complement of fundamental subspaces of a matrix

Monday October 12th 2015

Textbook reading: Chapter 3.1

Suggested practice exercises: Chapter 2.6, 5,6,7,36,37

Khan Academy video: Orthogonal complements

Strang lecture: Lecture 14: Orthogonal vectors and subspaces

1 Review

1.1 Orthogonality and FTLA

- $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are **orthogonal** iff $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = v_1 w_1 \cdots + v_n w_n = 0$.
  - Non-zero orthogonal vectors are independent.
- If V is a subspace of $\mathbb{R}^n$ then the *orthogonal complement* of V is

$$V^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{v} \cdot \mathbf{x} = 0, \text{for all } \mathbf{v} \in V\}$$

- If $W = V^\perp$ then $W^\perp = V$.
- In other words $(V^\perp)^\perp = V$.
- $\dim(V) + \dim(V^\perp) = \dim(\mathbb{R}^n) = n$

Example 1. Let V be the horizontal x-y-plane in $\mathbb{R}^3$ and W the vertical y-z-plane.

- Is W the orthogonal complement of V?
- Is it true that W is orthogonal to V?
- What is the orthogonal complement of V?

Example 2. Given

$$\mathrm{Nul}\left(\begin{bmatrix}1 & 2\\ 2 & 4\\ 3 & 6\end{bmatrix}\right) = \mathrm{Span}\left(\begin{bmatrix}2\\ -1\end{bmatrix}\right),$$

get

$$\mathrm{Col}\left(\begin{bmatrix}1 & 2\\ 2 & 4\\ 3 & 6\end{bmatrix}^T\right) = span\left(\begin{bmatrix}1\\ 2\end{bmatrix}\right)$$

Why?

Theorem 1 (Fundamental Theorem of Linear Algebra)**.** *Let A be a $m \times n$-matrix. Then*

- *$Nul(A)$ is the orthogonal complement of $Col(A^T)$ (in $\mathbb{R}^n$). Also, dim $Nul(A)+$ dim $Col(A^T) = (n-r)+r = n$.*
- *$Col(A)$ is the orthogonal complement of $Nul(A^T)$ (in $\mathbb{R}^m$).*

Why? Suppose $\mathbf{x} \in Nul(A)$. That is,

$$A\mathbf{x} = \mathbf{0}$$

What does this mean? (Think row-column rule).

- It means that the inner product of every row of A(transposed!) with $\mathbf{x}$ is zero. But that implies that $\mathbf{x}$ is **orthogonal to the row space.**

1.2 FLTA in action

Example 3. Find all vectors orthogonal to $\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}$ and $\begin{bmatrix}0\\ 1\\ 1\end{bmatrix}$.

Solution. This means: Find the orthogonal complement of $Col\left(\begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 1\end{bmatrix}\right)$.

Use the Fundamental Theorem: This is $Nul\left(\begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 1\end{bmatrix}^T\right) = Nul\left(\begin{bmatrix}1 & 1 & 1\\ 0 & 1 & 1\end{bmatrix}\right)$

Compute Nul space: $Nul\left(\begin{bmatrix}1 & 1 & 1\\ 0 & 1 & 1\end{bmatrix}\right) = Nul\left(\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\end{bmatrix}\right)$

Basis for Nul**:** $\left\{\begin{bmatrix}0\\ -1\\ 1\end{bmatrix}\right\}$

Final answer: the set of vectors orthogonal to $\begin{bmatrix}1\\1\\1\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$ is $span\left\{\begin{bmatrix}0\\-1\\1\end{bmatrix}\right\}$

Alternative solution. The FLTA is not magic! You can do this the old-fashioned way!

Looking for all $\mathbf{x}$ so that

$$\begin{bmatrix}1\\1\\1\end{bmatrix}\cdot\mathbf{x}=0 \text{ and } \begin{bmatrix}0\\1\\1\end{bmatrix}\cdot\mathbf{x}=0$$

Matrix form:

$$\begin{bmatrix}1&1&1\\0&1&1\end{bmatrix}\mathbf{x}=\begin{bmatrix}0\\0\end{bmatrix}$$

Get Null space:

$$\mathbf{x}\in Nul\left(\begin{bmatrix}1&1&1\\0&1&1\end{bmatrix}\right)$$

This is the same null space we obtained from the FTLA.

Example 4. Let $V=\left\{\begin{bmatrix}a\\b\\c\end{bmatrix}:a+b=2c\right\}$. Find a basis for the orthogonal complement of V.

Solution. Write as Null space: $V=Nul\left(\begin{bmatrix}1&1&-2\end{bmatrix}\right)$.

By FTLA: the orthogonal complement is $Col\left(\begin{bmatrix}1&1&-2\end{bmatrix}^T\right)$.

Basis for the orthogonal complement: $\begin{bmatrix}1\\1\\-2\end{bmatrix}$

Geometrically this makes sense: V is a plane with normal vector $\begin{bmatrix}1\\1\\-2\end{bmatrix}$

Alternative solution.

Notice that $a+b=2c \iff \begin{bmatrix}a\\b\\c\end{bmatrix}\cdot\begin{bmatrix}1\\1\\-2\end{bmatrix}=0.$

Interpret the above: V is actually defined as the orthogonal complement of

$$span\left\{\begin{bmatrix}1\\1\\-2\end{bmatrix}\right\}.$$

Example 5. Let $V = \left\{ \begin{bmatrix} 2a+b \\ -b \\ a+b \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Find the orthogonal complement of V.

Solution. Write as Column space: $V = span\left(\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right)$, so $V = Col\left(\begin{bmatrix} 2 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \right)$

By FTLA the orthogonal complement is $Nul\left(\begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \right)$

Get RREF to compute Null space: $\begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & -1 & \frac{1}{2} \end{bmatrix}$

Basis for the Null space: $\left\{ \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \right\}$

So the orthogonal complement to V is: $span\left\{ \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \right\}$

Directions and Equations. Let V be a subspace of $\mathbb{R}^n$. Then there are *two* ways of describing V.

By directions: If $V = Col(A)$ then you know that any vector $\mathbf{v}$ in V is a linear combination of the columns of A, so you know in which directions $\mathbf{v}$ can point.

By equations: If $V = Nul(B)$ then you know that any $\mathbf{v}$ in V satisfies the equations $\mathbf{R_i^T} \cdot \mathbf{v} = 0$, for all rows $\mathbf{R_i}$ of B.

Both descriptions are useful, and we will often switch between them, to answer any particular question we want to answer.

2 A new perspective on $A\mathbf{x} = \mathbf{b}$

To see if $A\mathbf{x} = \mathbf{b}$ has a solution, check that

Direct approach: $\mathbf{b} \in Col(A)$

Indirect approach: $\mathbf{b} \perp Nul(A^T)$

The indirect approach means:

$$\text{if } \underbrace{\mathbf{y}^T A = \mathbf{0}}_{\mathbf{y} \in Nul(A^T)}, \text{ then } \underbrace{\mathbf{y}^T \mathbf{b} = 0}_{\mathbf{b} \perp \mathbf{y}}.$$

Example 6. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}$. For which $\mathbf{b}$ does $A\mathbf{x} = \mathbf{b}$ have a solution?

Solution (old)**. Write augmented matrix, get Echelon form:**

$$\left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 3 & 1 & b_2 \\ 0 & 5 & b_3 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 0 & -3b_1 + b_2 + b_3 \end{array}\right]$$

When is this consistent? Whenever $-3b_1 + b_2 + b_3 = 0$.

Solution (new)**. Indirect approach says:** $A\mathbf{x} = \mathbf{b}$ solvable $\iff$ $\mathbf{b} \perp Nul(A^T)$.

Find basis for $Nul(A^T)$ **:**

$$\left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 2 & 1 & 5 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right]$$

so $Nul(A^T)$ has basis $\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$

Need $\mathbf{b} \perp Nul(A^T)$**:** $A\mathbf{x} = \mathbf{b}$ is solvable $\iff$ $\mathbf{b} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$

This is the same condition as before!

3 Motivation

3.1 How to find almost-solutions

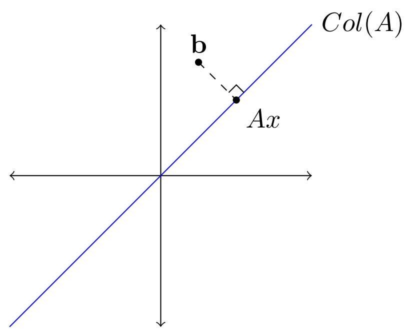
Why do we care about orthogonality? Not all linear systems have solutions. In fact, for many applications, data needs to be fitted and there is **no hope** for a perfect match. For example, $A\mathbf{x} = \mathbf{b}$ with

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

has no solution: $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ is not in $Col(A) = span \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$

Idea. Instead of giving up, we want the $\mathbf{x}$ which makes $A\mathbf{x}$ and $\mathbf{b}$ as close as possible.

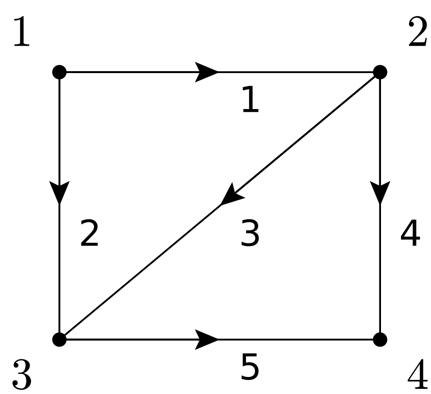
Such $\mathbf{x}$ is characterized by $A\mathbf{x}$ being **orthogonal** to the error $\mathbf{b} - A\mathbf{x}$.



4 Application: Directed graphs

4.1 Set up

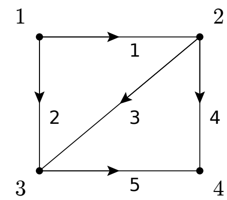
- Graphs appear in network analysis (e.g. internet) or circuit analysis.
- Arrow indicates direction of flow
- No edges from a node to itself



Definition 7. Let G be a graph with m edges and n nodes. The edge-node incidence matrix of G is the $m \times n$ matrix A with

$$A_{i,j} = \begin{cases} -1, & \text{if edge } i \text{ leaves node } j \\ +1, & \text{if edge } i \text{ enters node } j \\ 0, & \text{otherwise} \end{cases}$$

Example 8. Give the edge-node incidence matrix of our graph.



Solution.

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

- Each column represents a node
- Each row represents an edge

We are going to use linear algebra to study networks!

Math 415 - Lecture 21

Introduction

Wednesday October 14th 2015

Textbook reading: Chapter 2.5.

Suggested practice exercises: Chapter 2.5: 1, 2, 6.

Strang lecture: Lecture 12: Graphs, Networks, Incidence Matrices

1 Review

Recall that if $V \subset \mathbb{R}^n$ is a subspace, $V^\perp$ is the *orthogonal complement* of V, the subspace of all vectors $\mathbf{x}$ perp to all vectors of V.

Theorem 1. ***Fundamental Theorem of Linear Algebra.***

- $\dim(V) + \dim(V^\perp) = \dim(\mathbb{R}^n) = n$.
- $\mathrm{Col}(A)^\perp = \mathrm{Nul}(A^T)$.
- $\mathrm{Nul}(A)^\perp = \mathrm{Col}(A^T)$.

2 Directions and Equations

Directions and Equations. Let V be a subspace of $\mathbb{R}^n$. Then there are *two* ways of describing V.

By directions: If $V = Col(A)$ then you know that any vector $\mathbf{v}$ in V is a linear combination of the columns of A, so you know in which directions $\mathbf{v}$ can point.

By equations: If $V = \mathrm{Nul}(B)$ then you know that any $\mathbf{v}$ in V satisfies the equations $\mathbf{R_i^T} \cdot \mathbf{v} = 0$, for all rows $\mathbf{R_i}$ of B.

Both descriptions are useful, and we will often switch between them, to answer any particular question we want to answer.

3 A new perspective on $A\mathbf{x} = \mathbf{b}$

To see if $A\mathbf{x} = \mathbf{b}$ has a solution, check that

Direct approach: $\mathbf{b} \in Col(A)$

Indirect approach: $\mathbf{b} \perp Nul(A^T)$

The indirect approach means:

$$\text{if } \underbrace{\mathbf{y}^T A = \mathbf{0}}_{\mathbf{y} \in Nul(A^T)}, \text{ then } \underbrace{\mathbf{y}^T \mathbf{b} = 0}_{\mathbf{b} \perp \mathbf{y}}.$$

Example 2. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}$. For which $\mathbf{b}$ does $A\mathbf{x} = \mathbf{b}$ have a solution?

Solution (old)**. Write augmented matrix, get Echelon form:**

$$\left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 3 & 1 & b_2 \\ 0 & 5 & b_3 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 0 & -3b_1 + b_2 + b_3 \end{array}\right]$$

When is this consistent? Whenever $-3b_1 + b_2 + b_3 = 0$.

Solution (new)**. Indirect approach says:** $A\mathbf{x} = \mathbf{b}$ solvable $\iff$ $\mathbf{b} \perp Nul(A^T)$.

Find basis for $Nul(A^T)$:

$$\left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 2 & 1 & 5 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right]$$

so $Nul(A^T)$ has basis $\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$

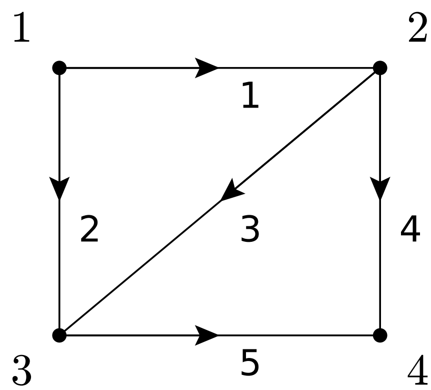
Need $\mathbf{b} \perp Nul(A^T)$: $A\mathbf{x} = \mathbf{b}$ is solvable $\iff$ $\mathbf{b} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$

This is the same condition as before!

4 Application: Directed graphs

4.1 Set up

- Graphs appear in network analysis (e.g. internet) or circuit analysis.
- Arrow indicates direction of flow
- No edges from a node to itself

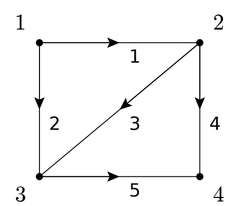


Definition 3. Let G be a graph with m edges and n nodes. The edge-node incidence matrix of G is the $m \times n$ matrix A with

$$A_{i,j} = \begin{cases} -1, & \text{if edge } i \text{ leaves node } j \\ +1, & \text{if edge } i \text{ enters node } j \\ 0, & \text{otherwise} \end{cases}$$

So each row (describing an edge=arrow) contains a single -1 (the tail of the arrow), a single $+1$ (the head of the arrow), and for the rest zeroes.

Example 4. Give the edge-node incidence matrix of our graph.



Solution.

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

4.2 Meaning of the null space

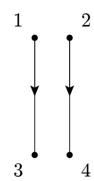
The $\mathbf{x}$ in $A\mathbf{x}$ assigns values to each node. (Think: assigning potentials)

$$A\mathbf{x} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_1 + x_2 \\ -x_1 + x_3 \\ -x_2 + x_3 \\ -x_2 + x_4 \\ -x_3 + x_4 \end{bmatrix}$$

Idea. $Ax = 0 \iff$ nodes connected by an edge are assigned the same value.

For our graph, $Nul(A)$ has basis $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$ (i.e. $x_1 = x_2 = x_3 = x_4$.) This always happens as long as G is connected.

Example 5. Give a basis for $Nul(A)$ for this graph:



Solution. If $A\mathbf{x} = 0$, then

$$\underbrace{x_1 = x_3}_{\text{connected by an edge}} \qquad \text{and} \qquad \underbrace{x_2 = x_4}_{\text{connected by an edge}}$$

So, $Nul(A)$ has basis: $\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\end{bmatrix}$

Just to make sure, the edge-node incidence matrix is:

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

Theorem 6. *$dim(Nul(A))$ is the number of connected subgraphs.*

- For large graphs, disconnection may not be visually apparent
- But, we can always find out by computing $dim(Nul(A))$ using Gaussian elimination!

4.3 Meaning of left null space

The $\mathbf{y}$ in $\mathbf{y}^T A$ is assigning values to each edge. (Think: assigning currents to edges, so that $\mathbf{y}$ describes a *flow pattern*.)

$$A = \begin{bmatrix}-1 & 1 & 0 & 0\\ -1 & 0 & 1 & 0\\ 0 & -1 & 1 & 0\\ 0 & -1 & 0 & 1\\ 0 & 0 & -1 & 1\end{bmatrix}, A^T = \begin{bmatrix}-1 & -1 & 0 & 0 & 0\\ 1 & 0 & -1 & -1 & 0\\ 0 & 1 & 1 & 0 & -1\\ 0 & 0 & 0 & 1 & 1\end{bmatrix}$$

$$A^T\mathbf{y} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 - y_4 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{bmatrix}$$

$$A^T\mathbf{y} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 - y_4 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{bmatrix}$$

Idea. So: $A^T\mathbf{y} = 0 \iff$ at each node, (directed) values assigned to edges add to zero.

When thinking of currents, this is Kirchhoff's first law: at each node, incoming and outgoing currents balance. **Flow in = Flow out.**

What is the simplest way to balance current?

Assign current in a loop! We have two loops:

$$edge_1 \to edge_3 \to -edge_2 \text{ and } edge_3 \to edge_5 \to -edge_4$$

Example 7. Solve $A^T\mathbf{y} = 0$ for our graph. Recall

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

Solution. Get RREF:

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \overset{RREF}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The parametric solution is: $\begin{bmatrix} y_3 - y_5 \\ -y_3 + y_5 \\ y_3 \\ -y_5 \\ y_5 \end{bmatrix}$

So a basis for $Nul(A^T)$ **is:** $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$

Observation: These two basis vectors correspond to loops.

Note: get the "simpler" loop $\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ as $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$

Theorem 8. *In general, $dim(Nul(A^T))$ is the number of (independent) loops.*

For large graphs, we now have a nice way to computationally find all loops.

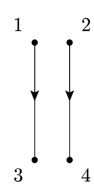
5 Summary/Outlook

* We described a network by using a matrix A.
* The Null space $Nul(A)$ has as dimension the number of connected components of the network.
* The Left Null Space $Nul(A^T)$ has as dimension the number of independent loops.
* The column space $Col(A)$ and row space $Col(A^T)$ also have "geometric" meaning in terms of the network, see the book and Strang's lecture.

6 Practice problems

6.1 Problem 1

Example 9. Give a basis for $Nul(A^T)$ for the following graph:



Solution. This graph contains no loops, so $Nul(A^T)$ =??. $Nul(A^T)$ has the empty set as basis.

To check, the incidence matrix is :

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

Indeed, $Nul(A^T) = \{0\}$.

6.2 Problem 2

Example 10. Draw the graph with edge-node incidence matrix

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

Give a basis for $Nul(A)$ and $Nul(A^T)$.

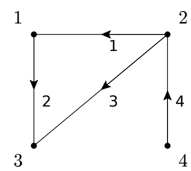


Figure 1: The graph

Solution.

If $A\mathbf{x} = 0$**,** then $x_1 = x_2 = x_3 = x_4$ (all connected by edges.)

So, $Nul(A)$ has basis $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

(This graph is connected, so only 1 connected subgraph, so $dim(Nul(A)) = 1$.)

Loops: This graph has one loop: $edge_1 \to edge_2 \to -edge_3$. Assign values $y_1 = 1, y_2 = 1, y_3 = -1$ along the edges of that loop.

$Nul(A^T)$ has basis $\begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$

(The graph has 1 loop, so $dim(Nul(A^T)) = 1$.)

Math 415 - Lecture 22

Orthogonal projection

Friday October 16th 2015

Textbook reading: Chapter 3.2.

Suggested practice exercises: Chapter 3.2: 9, 10, 17, 19.

Strang lecture: Lecture 15: Projections onto Subspaces

1 Review/Outlook

- We can deal with complicated linear systems $Ax = b$ (maybe with help of a computer), but what to do when there is no exact solution?
- $Ax = b$ had no solution if b is not in $Col(A)$.
- Idea: make $Ax - b$ as small as possible (when we vary x).
- How? *Project* b on the column space $Col(A)$.
- Recall: If $v_1, v_2, \dots, v_n$ are orthogonal (and non zero) they are independent.
- Recall: To calculate coordinates for orthogonal vectors is easy: this uses

$$v_1 \cdot (c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) = c_1 v_1 \cdot v_1.$$

2 Orthogonal Bases

2.1 Orthogonal Basis

Definition 1. A basis $v_1, v_2, \dots, v_n$ of $\mathbb{R}^n$ is called *orthogonal* if the vectors are pairwise orthogonal, $v_i \cdot v_j = 0$ if $i \neq j$.

Example 2. The standard basis $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an orthogonal basis for $\mathbb{R}^3$. Similarly, the standard basis $e_1, e_2, \dots, e_n$ is an orthogonal basis for $\mathbb{R}^n$.

Example 3. Are the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ an orthogonal basis for $\mathbb{R}^3$?

Solution. Just check:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = ??, \quad \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = ??, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = ??$$

So this *is* an orthogonal basis. Note that we don't have to check it is a basis: orthogonality implies independence, and 3 independent vectors form a basis in $\mathbb{R}^3$.

Example 4. Suppose $v_1, v_2, \ldots, v_n$ form an orthogonal basis of $\mathbb{R}^n$, $w \in \mathbb{R}^n$. Find the coordinates of w. That is, find constants $c_1, c_2, \ldots, c_n$ so that

$$w = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

Solution. Take dot product with v_1 on both sides:

$$v_1 \cdot w = v_1 \cdot (c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) = c_1 v_1 \cdot v_1.$$

Hence $c_1 = \frac{v_1 \cdot w}{v_1 \cdot v_1}$ and more generally $c_i = \frac{v_i \cdot w}{v_i \cdot v_i}$.

Easy (and Important) Formula
If $v_1, v_2, \ldots, v_p$ form an orthogonal basis of $V \subset \mathbb{R}^n$, $w \in V$, then $w = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p$, with

$$c_i = \frac{v_i \cdot w}{v_i \cdot v_i}.$$

Special Case
If $v_1, v_2, \ldots, v_p$ is orthonormal then

$$c_i = v_i \cdot w.$$

Example 5. Express $w = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in the basis $v_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Solution.

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} =$$

We use the formula for the coordinates:

$$c_1 = ??c_2 = ??c_3 = ??,$$

Warning
The easy formula for the coordinates only works for *orthogonal* bases.

Example 6. Take the basis $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}$ and the vector $w = \begin{bmatrix}4\\9\end{bmatrix}$. Then

$$\begin{bmatrix}4\\9\end{bmatrix} = 1\begin{bmatrix}1\\0\end{bmatrix} + 3\begin{bmatrix}1\\3\end{bmatrix},$$

and the coefficients are *not* the numbers you get from the easy formula. To find them you need to solve a system of equations.

Example 7. The standard basis $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$ is orthonormal. Find the coordinates of $\begin{bmatrix}3\\7\\4\end{bmatrix}$ in the standard basis.

Solution. This is trivial of course,

$$\begin{bmatrix}3\\7\\4\end{bmatrix} = ?? \qquad \begin{bmatrix}1\\0\\0\end{bmatrix} + ?? \qquad \begin{bmatrix}0\\1\\0\end{bmatrix} + ?? \qquad \begin{bmatrix}0\\0\\1\end{bmatrix}$$

But ....

Solution (continued)**.** note that the coordinates are dot products with orthonormal vectors:

$$\begin{bmatrix}3\\7\\4\end{bmatrix} \cdot \begin{bmatrix}1\\0\\0\end{bmatrix} = ?? \qquad , \qquad \begin{bmatrix}3\\7\\4\end{bmatrix} \cdot \begin{bmatrix}0\\1\\0\end{bmatrix} = \qquad , \qquad \begin{bmatrix}3\\7\\4\end{bmatrix} \cdot \begin{bmatrix}0\\0\\1\end{bmatrix} = ?? \qquad .$$

Example 8. The vectors $\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$ form an orthogonal basis. Produce from it an *orthonormal* basis.

Solution. We just divide by the lengths of these vectors (this will keep them orthogonal).

$$\begin{bmatrix}1\\-1\\0\end{bmatrix} \text{ has length ?? } \qquad = ?? \qquad \text{, normalized: ??} \qquad \begin{bmatrix}1\\-1\\0\end{bmatrix}$$

Solution (continued)**.**

$$\begin{bmatrix}1\\1\\0\end{bmatrix} \text{ has length ?? } \qquad = ?? \qquad \text{, normalized: ??} \qquad \begin{bmatrix}1\\1\\0\end{bmatrix}$$

and

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

is already normalized. So we get as orthonormal basis

$$?? \qquad \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad ?? \qquad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad ?? \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Example 9. Express $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in the orthonormal basis $(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix})$.

Solution. Just calculate dot products:

$$c_1 = \ = ?? \qquad , \quad c_2 = ?? \qquad = ?? \qquad ,$$

Solution (continued)**.**

$$c_3 = ?? \qquad = ??$$

so that

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = ?? \qquad \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + ?? \qquad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + ?? \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} =$$

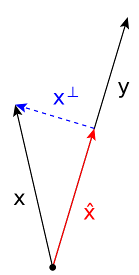
$$??$$

3 Orthogonal Projection

Definition 10 (Orthogonal Projection)**.** The **orthogonal projection** of vector $\mathbf{x}$ on vector $\mathbf{y}$ is

$$\hat{\mathbf{x}} = \frac{\mathbf{y} \cdot \mathbf{x}}{\mathbf{y} \cdot \mathbf{y}} \mathbf{y}.$$

The **error** is $\mathbf{x}^\perp = \mathbf{x} - \hat{\mathbf{x}}$.



- The projection $\hat{\mathbf{x}}$ is the *closest point* to $\mathbf{x}$ on the line through $\mathbf{y}$.
- The error $\mathbf{x}^\perp = \mathbf{x} - \hat{\mathbf{x}}$ is characterized by the property that it is orthogonal to $Span(\mathbf{y})$.
- We have a decomposition $\mathbf{x} = \hat{\mathbf{x}} + \mathbf{x}^\perp$. The **projection** $\hat{\mathbf{x}}$ is in $Span(\mathbf{y})$ and $\mathbf{x}^\perp$ is orthogonal to $Span(\mathbf{y})$.

Summary: the projection formula is

$$\hat{\mathbf{x}} = \frac{\mathbf{y} \cdot \mathbf{x}}{\mathbf{y} \cdot \mathbf{y}} \mathbf{y}.$$

Why?

Solution.

- We know $\hat{\mathbf{x}}$ is in the direction of $\mathbf{y}$, so $\hat{\mathbf{x}} = c\mathbf{y}$ for some constant c.
- The error $\mathbf{x} - \hat{\mathbf{x}}$ orthogonal to $\mathbf{y}$.
- So $0 = \mathbf{y} \cdot (\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{y} \cdot (\mathbf{x} - c\mathbf{y}) = \mathbf{y} \cdot \mathbf{x} - c\mathbf{y} \cdot \mathbf{y}$.
- Solving for c gives $c = \frac{\mathbf{y}\cdot\mathbf{x}}{\mathbf{y}\cdot\mathbf{y}}$.

Example 11. Find the orthogonal projection of $\mathbf{x} = \begin{bmatrix} -8 \\ 4 \end{bmatrix}$ onto $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

Solution.

$$\hat{\mathbf{x}} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}}\mathbf{y} = \frac{-8.3 + 4.1}{3^2 + 1^2} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = -2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ -2 \end{bmatrix}.$$

The error is

$$\mathbf{x}^{\perp} = \mathbf{x} - \hat{\mathbf{x}} = \begin{bmatrix} -8 \\ 4 \end{bmatrix} - \begin{bmatrix} -6 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}.$$

Note that vector $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and error $\mathbf{x}^{\perp} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$ are orthogonal.

Example 12. What is the projection of $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ onto each of the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$?

Solution.

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \text{ onto } \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} : \frac{2.1 + 1.(-1) + 1.0}{1^2 + (-1)^2 + 0^2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Solution.

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \text{ onto } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} : \frac{2.0 + 1.0 + 1.1}{0^2 + 0^2 + 1^2} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note that these sum up to $\frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \mathbf{x}$. Why? because ...

Theorem 1. *If $v_1, \ldots, v_n$ is orthogonal basis of V and $w \in V$ then*

$$w = c_1 v_1 + \cdots + c_n v_n, \quad \textit{with } c_j = \frac{w \cdot v_j}{v_j \cdot v_j}.$$

So the terms in this sum are precisely the projections onto each basis vector.

4 Projection Matrix

If $\mathbf{y}$ is a fixed nonzero vector, we get from any vector $\mathbf{x}$ the projection $\hat{\mathbf{x}}$. There is a matrix that turns $\mathbf{x}$ into $\hat{\mathbf{x}}$. How? Rewrite the formula for $\hat{\mathbf{x}}$.

$$\hat{\mathbf{x}} = \frac{\mathbf{y}\cdot\mathbf{x}}{\mathbf{y}\cdot\mathbf{y}}\mathbf{y} = \frac{1}{\mathbf{y}\cdot\mathbf{y}}\mathbf{y}(\mathbf{y}^T\mathbf{x}) = \left(\frac{1}{\mathbf{y}\cdot\mathbf{y}}\mathbf{y}\mathbf{y}^T\right)\mathbf{x} = P\mathbf{x},$$

where $P = \frac{1}{\mathbf{y}\cdot\mathbf{y}}\mathbf{y}\mathbf{y}^T$. P is called the *projection matrix* on the subspace $Span(\mathbf{y})$.

Example 13. Let $\mathbf{y} = \begin{bmatrix}1\\1\end{bmatrix}$. Find the projection matrix P for $\mathbf{y}$ and use it to calculate the projections of $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}$ on $\mathbf{y}$.

Solution.

$$P = \frac{1}{\mathbf{y}\cdot\mathbf{y}}\mathbf{y}\mathbf{y}^T = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}.$$

Then

- If $\mathbf{x} = \begin{bmatrix}1\\0\end{bmatrix}$ then $\hat{\mathbf{x}} = P\mathbf{x} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}$.
- If $\mathbf{x} = \begin{bmatrix}1\\1\end{bmatrix}$ then $\hat{\mathbf{x}} = P\mathbf{x} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix} = \mathbf{x}$! Why?
- If $\mathbf{x} = \begin{bmatrix}1\\-1\end{bmatrix}$ then $\hat{\mathbf{x}} = P\mathbf{x} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$! Why?

Math 415 - Lecture 23

Projections on subspaces

Monday October 19th 2015

Textbook reading: Chapter 3.2, 3.3, 3.4

Suggested practice exercises: Chapter 3.2 Exercise 17, 18, 24, Chapter 3.4 Exercise 2, 3 and see exercise at the end of this notes

Khan Academy video: Projections onto subspaces, Visualizing a projection onto a plane, Projection is closest vector in subspace

1 Review

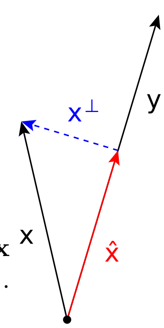
Last time

- **Orthogonal projection** of x onto y:

$$\hat{\mathbf{x}} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}} \mathbf{y}$$

"Error" $\mathbf{x}^{\perp} = \mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to $\mathbf{y}$.

- If $\mathbf{y}_1, \ldots, \mathbf{y}_n$ is an **orthogonal basis** of V, and $\mathbf{x}$ is in V, then $\mathbf{x} = c_1\mathbf{y}_1 + \ldots + c_n\mathbf{y}_n$ with $c_j = \frac{\mathbf{x} \cdot \mathbf{y}_j}{\mathbf{y}_j \cdot \mathbf{y}_j}$.



Remark. $\mathbf{x}$ decomposes as the sum of its projections onto each vector in the orthogonal basis.

Remark. The formulas simplify when you project on *unit* vectors: all denominators are then 1.

Example 1. Express $\underbrace{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{x}}$ in terms of the basis $\underbrace{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\mathbf{y}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{y}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{y}_3}$

Solution.
Notice that $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ is an orthogonal basis of $\mathbb{R}^3$.

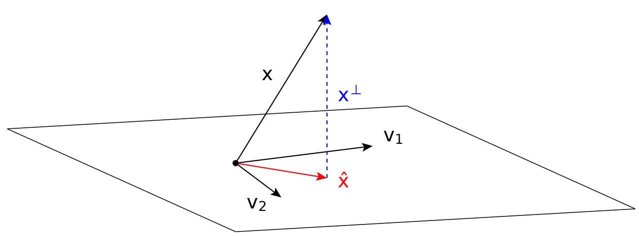
$$\begin{aligned}
\begin{bmatrix}2\\1\\1\end{bmatrix} &= c_1 \begin{bmatrix}1\\-1\\0\end{bmatrix} + c_2 \begin{bmatrix}1\\1\\0\end{bmatrix} + c_3 \begin{bmatrix}0\\0\\1\end{bmatrix} \\
&= \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix} \cdot \begin{bmatrix}1\\-1\\0\end{bmatrix}}{\begin{bmatrix}1\\-1\\0\end{bmatrix} \cdot \begin{bmatrix}1\\-1\\0\end{bmatrix}} \begin{bmatrix}1\\-1\\0\end{bmatrix}}_{\text{projection of } \mathbf{x} \text{ onto } \mathbf{y}_1} + \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix} \cdot \begin{bmatrix}1\\1\\0\end{bmatrix}}{\begin{bmatrix}1\\1\\0\end{bmatrix} \cdot \begin{bmatrix}1\\1\\0\end{bmatrix}} \begin{bmatrix}1\\1\\0\end{bmatrix}}_{\text{projection of } \mathbf{x} \text{ onto } \mathbf{y}_2} + \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix} \cdot \begin{bmatrix}0\\0\\1\end{bmatrix}}{\begin{bmatrix}0\\0\\1\end{bmatrix} \cdot \begin{bmatrix}0\\0\\1\end{bmatrix}} \begin{bmatrix}0\\0\\1\end{bmatrix}}_{\text{projection of } \mathbf{x} \text{ onto } \mathbf{y}_3} \\
&= \frac{1}{2} \begin{bmatrix}2\\1\\1\end{bmatrix} + \frac{3}{2} \begin{bmatrix}1\\1\\0\end{bmatrix} + \begin{bmatrix}0\\0\\1\end{bmatrix}
\end{aligned}$$

2 Orthogonal projection on subspaces

2.1 Projecting onto a subspace

Theorem 2. *Let W be a subspace of $\mathbb{R}^n$. Then, each $\mathbf{x}$ in $\mathbb{R}^n$ can be uniquely written as*

$$\mathbf{x} = \underbrace{\hat{\mathbf{x}}}_{\text{in } W} + \underbrace{\mathbf{x}^{\perp}}_{\text{in } W^{\perp}}$$



$\hat{\mathbf{x}}$ is the **orthogonal projection** of $\mathbf{x}$ onto W.

- $\hat{\mathbf{x}}$ is the point in W closest to $\mathbf{x}$. For any other $\mathbf{y}$ in W, $\text{dist}(\mathbf{x}, \hat{\mathbf{x}}) < \text{dist}(\mathbf{x}, \mathbf{y})$.
- If $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is an orthogonal basis of W, then

$$\hat{\mathbf{x}} = \left(\frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right) \mathbf{v}_1 + \ldots + \left(\frac{\mathbf{x} \cdot \mathbf{v}_m}{\mathbf{v}_m \cdot \mathbf{v}_m}\right) \mathbf{v}_m$$

  Once $\hat{\mathbf{x}}$ is determined, $\mathbf{x}^{\perp} = \mathbf{x} - \hat{\mathbf{x}}$.

  (This is also the orthogonal projection of $\mathbf{x}$ onto $W^{\perp}$.)

Example 3. Let $W = span\left\{\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$, and $\mathbf{x} = \begin{bmatrix}0\\3\\10\end{bmatrix}$.

- Find the orthogonal projection of $\mathbf{x}$ onto W. (Or: find the vector in W which is closest to $\mathbf{x}$)
- Write $\mathbf{x}$ as a vector in W plus a vector orthogonal to W.

Solution.

Note that $\mathbf{w}_1 = \begin{bmatrix}3\\0\\1\end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$ are an orthogonal basis for W.
(We will soon learn how to construct orthogonal bases ourselves).
Hence, the orthogonal projection of $\mathbf{x}$ onto W is:

$$\begin{aligned}\hat{\mathbf{x}} &= \frac{\mathbf{x}\cdot\mathbf{w}_1}{\mathbf{w}_1\cdot\mathbf{w}_1}\mathbf{w}_1 + \frac{\mathbf{x}\cdot\mathbf{w}_2}{\mathbf{w}_2\cdot\mathbf{w}_2}w_2 \\ &= \frac{\begin{bmatrix}0\\3\\10\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}{\begin{bmatrix}3\\0\\1\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}\begin{bmatrix}3\\0\\1\end{bmatrix} + \frac{\begin{bmatrix}0\\3\\10\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}\begin{bmatrix}0\\1\\0\end{bmatrix} \\ &= \frac{10}{10}\begin{bmatrix}3\\0\\1\end{bmatrix} + 3\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}3\\3\\1\end{bmatrix}\end{aligned}$$

Warning
This calculation only works for *orthogonal* $\mathbf{w_1}, \mathbf{w_2}$!

$\hat{\mathbf{x}}$ is the vector in W which best approximates $\mathbf{x}$.
Orthogonal projection of $\mathbf{x}$ onto the orthogonal complement of W:

$$\mathbf{x}^{\perp} = \begin{bmatrix}0\\3\\10\end{bmatrix} - \begin{bmatrix}3\\3\\1\end{bmatrix} = \begin{bmatrix}-3\\0\\9\end{bmatrix}$$

Hence,

$$\mathbf{x} = \begin{bmatrix}0\\3\\10\end{bmatrix} = \underbrace{\begin{bmatrix}3\\3\\1\end{bmatrix}}_{\text{in } W} + \underbrace{\begin{bmatrix}-3\\0\\9\end{bmatrix}}_{\text{in } W^{\perp}}$$

Indeed, $\begin{bmatrix}-3\\0\\9\end{bmatrix}$ is orthogonal to $\mathbf{w}_1 = \begin{bmatrix}3\\0\\1\end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$.

2.2 The matrix of a projection

Definition 4. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be an orthogonal basis of W, a subspace of $\mathbb{R}^n$. Note that the projection map $\pi_W : \mathbb{R}^n \to \mathbb{R}^n$ that sends $\mathbf{x}$ to

$$\hat{\mathbf{x}} = \left(\frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\right)\mathbf{v}_1 + \ldots + \left(\frac{\mathbf{x}\cdot\mathbf{v}_m}{\mathbf{v}_m\cdot\mathbf{v}_m}\right)\mathbf{v}_m$$

is a linear map. The matrix P representing π_W with respect to the standard basis is the **projection matrix**.

Example 5. Find the projection matrix P for the orthogonal projection onto

$$W = span\left\{\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$$

in $\mathbb{R}^3$.

Solution. Standard basis: $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$.

The first column of P encodes the projection of $\begin{bmatrix}1\\0\\0\end{bmatrix}$:

$$\frac{\begin{bmatrix}1\\0\\0\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}{\begin{bmatrix}3\\0\\1\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}\begin{bmatrix}3\\0\\1\end{bmatrix} + \frac{\begin{bmatrix}1\\0\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}\begin{bmatrix}0\\1\\0\end{bmatrix} = \frac{3}{10}\begin{bmatrix}3\\0\\1\end{bmatrix}.$$

Hence $P = \begin{bmatrix}\frac{9}{10} & * & *\\ 0 & * & *\\ \frac{3}{10} & * & *\end{bmatrix}$.

The second column of P encodes the projection of $\begin{bmatrix}0\\1\\0\end{bmatrix}$:

$$\frac{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}{\begin{bmatrix}3\\0\\1\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}}\begin{bmatrix}3\\0\\1\end{bmatrix} + \frac{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}}\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Hence $P = \begin{bmatrix} \frac{9}{10} & 0 & * \\ 0 & 1 & * \\ \frac{3}{10} & 0 & * \end{bmatrix}$.

The third column of P encodes the projection of $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$:

$$\frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

Hence $P = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix}$.

Let's do the earlier example again using the matrix P.

Example 6. Let $W = span \left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$, and $\mathbf{x} = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{x}$ onto W.

Solution.

$$\hat{\mathbf{x}} = P\mathbf{x} = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix},$$

as in the previous example.

Example 7. Compute P^2 when

$$P = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix}.$$

Explain why the answer makes sense.

Solution.

$$\begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix}$$

$$P^2 = P$$

Once we have projected down onto W, projecting onto W again does not change anything!

3 Practice problems

3.1 Practice problems

Example 8. Find the closest point to $\mathbf{x}$ in $span\{\mathbf{v}_1, \mathbf{v}_2\}$ where

$$\mathbf{x} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -2 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Solution.

$$\hat{\mathbf{x}} = \frac{6}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \frac{-2}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -1 \\ -1 \end{bmatrix}$$

Example 9. If P is the projection matrix for projecting on W, what is the projection matrix Q for projecting on $W^{\perp}$?

Solution. $Q = 1 - P$!

Example 10. Let P be the projection matrix for projecting on W, and let $\mathbf{x}$ be some vector.

- Suppose $P\mathbf{x} = \mathbf{x}$. What can you say about $\mathbf{x}$?
- Suppose $P\mathbf{x} = 0$. What can you say about $\mathbf{x}$?

Math 415 - Lecture 24

Least squares

Wednesday October 21st 2015

Textbook reading: Chapter 3.3

Suggested practice exercises: Exercises 3, 5, 6, 13, 24, 25

Khan Academy video: Least Squares Approximation, Least Squares Examples, Another Least Squares Example

1 Review

Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ in $\mathbb{R}^n$ (but maybe not in W). Let $\mathbf{x}_W$ be the orthonormal projection of $\mathbf{x}$ onto W. (vector in W as close as possible to $\mathbf{x}$)

- If $\mathbf{v}_1, \cdots, \mathbf{v}_m$ is an orthogonal basis of W then

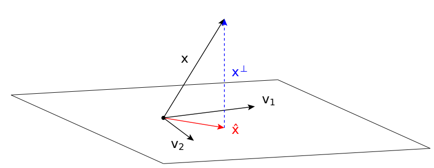
$$\mathbf{x}_W = \underbrace{\left(\frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\right)\mathbf{v_1}}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_1} + \cdots + \underbrace{\left(\frac{\mathbf{x}\cdot\mathbf{v}_m}{\mathbf{v}_m\cdot\mathbf{v}_m}\right)\mathbf{v_m}}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_m}.$$

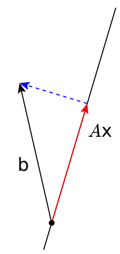
- The decomposition $\mathbf{x} = \underbrace{\mathbf{x}_W}_{\text{in } W} + \underbrace{\mathbf{x}^{\perp}}_{\text{in } W^{\perp}}$ is unique.

2 Least squares

2.1 Least squares

Definition. $\hat{\mathbf{x}}$ is a **least squares solution** of the system $A\mathbf{x} = \mathbf{b}$ if $\hat{\mathbf{x}}$ is such that $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible.





- If $A\mathbf{x} = \mathbf{b}$ is consistent, then a least squares solution $\hat{\mathbf{x}}$ is just an ordinary solution. (in that case, $A\hat{\mathbf{x}} - \mathbf{b} = 0$)
- Interesting case: $A\mathbf{x} = \mathbf{b}$ is inconsistent. (in other words: the system is overdetermined)

Idea. $A\mathbf{x} = \mathbf{b}$ is consistent $\iff$ $\mathbf{b}$ is in $\mathrm{Col}(A)$

So if $A\mathbf{x} = \mathbf{b}$ is inconsistent we

- replace $\mathbf{b}$ with its projection $\hat{\mathbf{b}}$ onto $\mathrm{Col}(A)$,
- and solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.(consistent by construction!)

Example 1. Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

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

Solution. Note the columns of A are orthogonal. Otherwise, we could not proceed in the same way. Hence the projection of $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\mathbf{b}} = \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

We have already solved $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ in the process: $\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$. **Question:** What to do when the columns of A are not orthogonal?

3 The normal equations

Theorem 1. $\hat{\mathbf{x}}$ *is a least squares solution of* $A\mathbf{x} = \mathbf{b}$ $\iff$ $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$

Proof. $\iff$ $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible $\iff$ $A\hat{\mathbf{x}} - \mathbf{b}$ is orthogonal to $\mathrm{Col}(A)$ $\overset{FTLA}{\iff}$ $A\hat{\mathbf{x}} - \mathbf{b}$ is in $\mathrm{Nul}(A^T)$ $\iff$ $A^T(A\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0}$ $\iff$ $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ $\square$

Example 2 (again). Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

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

Solution.

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

$$A^T \mathbf{b} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

The normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ are $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Solving, we find (again) $\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

Example 3. Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

What is the projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$?

Note that the columns of A are not orthogonal.

Solution.

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

The normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ are $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$. Solving, we find $\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. The projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is $A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Just to make sure: why is $A\hat{\mathbf{x}}$ the projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$? Because, for a least square solution $\hat{\mathbf{x}}$, $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible.

The projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\mathbf{b}} = A\hat{\mathbf{x}}, \text{ with } \hat{\mathbf{x}} \text{ such that } A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

If A has full column rank, (so the columns of A are independent,) this is

$$\hat{\mathbf{b}} = A(A^TA)^{-1}A^T\mathbf{b}.$$

(In this case A^TA is invertible.)

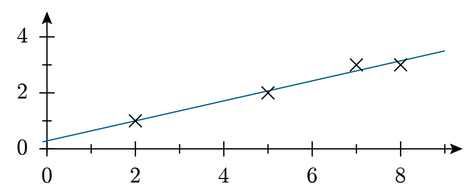
Hence, the projection matrix for projecting onto $\mathrm{Col}(A)$ is

$$P = A(A^TA)^{-1}A^T.$$

4 Applications

4.1 Least square regression lines

Experimental data: (x_i, y_i), for $i = 1, 2, 3, \ldots..$ Wanted: parameters β_1, β_2 such that $y_i \approx \beta_1 + \beta_2 x_i$ for all i



The approximation should be so that

$$\underbrace{\mathrm{SS}_{\mathrm{res}} = \sum_i [y_i - (\beta_1 + \beta_2 x_i)]^2}_{\text{residue sum of squares}} \text{ is as small as possible.}$$

Example 4. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2, 1), (5, 2), (7, 3), (8, 3)$.

Solution. The equations $y = \beta_1 + \beta_2 x$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \mathbf{y}}$$

Here, we need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

Calculate X^TX and $X^T\mathbf{y}$ to get the normal equation:

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$. Hence the least squares line is $y = \frac{2}{7} + \frac{5}{14}x$.

Example 5. Blood is drawn form volunteers to determine the effects of a new experimental drug designed to lower cholesterol levels. The following data shows the results of varying the dosage from 0 unit to 1 units in step of 0.2 of a unit. Find a line $C = \beta_1 D + \beta_2$ that best fits the data. What drug usage would you recommend if you want to accomplish a Cholesterol level of 215?

Drug Dosage: D	0.0	0.2	0.4	0.6	0.8	1
Cholesterol: C	289	273	254	226	213	189

Solution.

$$\underbrace{\begin{bmatrix} D_1 & 1 \\ D_2 & 1 \\ D_3 & 1 \\ D_4 & 1 \\ D_5 & 1 \\ D_6 & 1 \end{bmatrix}}_{\text{design matrix } D} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{bmatrix}}_{\text{observation vector } \mathbf{c}}$$

Here, we need to find a least squares solution to $D\beta = c$ or

$$\begin{bmatrix} 0 & 1 \\ 0.2 & 1 \\ 0.4 & 1 \\ 0.6 & 1 \\ 0.8 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 289 \\ 273 \\ 254 \\ 226 \\ 213 \\ 189 \end{bmatrix}.$$

Calculate D^TD and $D^T\mathbf{c}$ to get the normal equation:

$$D^TD = \begin{bmatrix} 0 & .2 & .4 & .6 & .8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ .2 & 1 \\ .4 & 1 \\ .6 & 1 \\ .8 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2.2 & 3 \\ 3 & 6 \end{bmatrix}$$

$$D^T\mathbf{c} = \begin{bmatrix} 0 & .2 & .4 & .6 & .8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 289 \\ 273 \\ 254 \\ 226 \\ 213 \\ 189 \end{bmatrix} = \begin{bmatrix} 651.2 \\ 1444 \end{bmatrix}$$

Solving $\begin{bmatrix} 2.2 & 3 \\ 3 & 6 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 651.2 \\ 1444 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -\frac{65708}{7} \\ \frac{14116}{21} \end{bmatrix}$. Hence the least squares line is $c = -\frac{65708}{7}d + \frac{14116}{21}$.

Math 415 - Lecture 25

Multiple linear regression, Gram Schmidt and Orthogonal matrices

Monday October 26th 2015

Textbook reading: Chapters 3.3,3.4

Suggested practice exercises: Chapter 3.3, 3,5,6,13,22,24,25,26 and Chapter 3.4, 10,11,13,14,16,26

Khan Academy video: Another Least Squares Example, Gram-Schmidt Example

Strang lecture: Orthogonal Matrices and Gram-Schmidt Process.

1 Review

$\hat{\mathbf{x}}$ **is a least squares solution** of the system $A\mathbf{x} = \mathbf{b}$

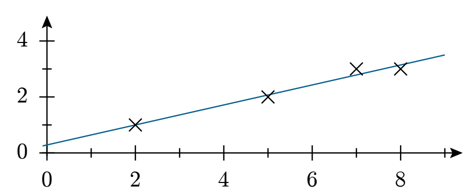
$\iff$ $\hat{\mathbf{x}}$ is such that $A\hat{\mathbf{x}} - \mathbf{b}$ is as small as possible

$\overset{FTLA}{\iff}$ $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ (the **normal equations**)

2 Application: fitting data

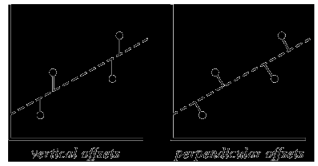
2.1 Least square lines

Example 1. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2, 1), (5, 2), (7, 3), (8, 3)$.



Comment

As usual in practice, we are minimizing the (the sum of the squares of the) vertical offsets.



2.2 Solution

The equations $y = \beta_1 + \beta_2 x$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \mathbf{y}}$$

Here, we need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Hence the least squares line is $y = \frac{2}{7} + \frac{5}{14}x$.

2.3 Fitting to other curves

What happens if the data just lie close to any line? We can also fit the experimental data using other curves. Try to find $\beta_1, \beta_2, \beta_3$ such that $y = \beta_1 + \beta_2 x + \beta_3 x^2$ fits the data. To fit: $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$. The equations $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\text{observation vector } \mathbf{y}}$$

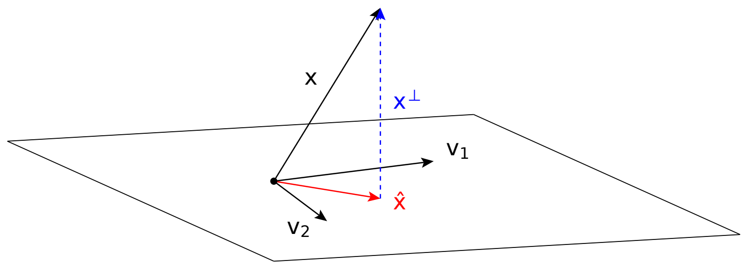
Given data (x_i, y_i), we then find the least squares solution to $X\beta = \mathbf{y}$.

2.4 Multiple linear regression

Of course, sometimes the variable y might not just depend on a single variable x, but on two variables, say u and v. So, here you have find the least-squares solution of

$$\underbrace{\begin{bmatrix} 1 & v_1 & w_1 \\ 1 & v_2 & w_2 \\ 1 & v_3 & w_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\text{observation vector}}$$

And we again proceed by finding a least squares solution.



3 Review

Suppose $\mathbf{v}_1, \dots, \mathbf{v}_m$ is an orthonormal basis of W. The **orthogonal projection** of $\mathbf{x}$ onto W is :

$$\hat{\mathbf{x}} = \underbrace{\langle \mathbf{x}, \mathbf{v}_1\rangle \mathbf{v}_1}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_1} + \dots + \underbrace{\langle \mathbf{x}, \mathbf{v}_m\rangle \mathbf{v}_m}_{\text{proj. of } \mathbf{x} \text{ onto } \mathbf{v}_m}$$

(To stay agile, we are writing $\langle \mathbf{x}, \mathbf{v}_1\rangle = \mathbf{x} \cdot \mathbf{v}_1$ for the inner product.)

4 Gram-Schmidt

4.1 Our goal

* In calculating projections we used an *orthogonal basis* and the easy formula for the coefficients.

* What if we are given an arbitrary basis, not orthogonal?

* Turn the starting basis into an orthogonal (or orthonormal) basis.

* **Gram-Schmidt Process.**

Recipe. (Gram-Schmidt orthonormalization)
Given a basis $\mathbf{a}_1, \dots, \mathbf{a}_n$, produce a orthogonal basis $\mathbf{b}_1, \dots, \mathbf{b}_n$ and an orthonormal basis $\mathbf{q}_1, \dots, \mathbf{q}_n$.

$$\begin{array}{ll}
\mathbf{b}_1 = \mathbf{a}_1, & \mathbf{q}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} \\
\mathbf{b}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1\rangle \mathbf{q}_1, & \mathbf{q}_2 = \frac{\mathbf{b}_2}{\|\mathbf{b}_2\|} \\
\mathbf{b}_3 = \mathbf{a}_3 - \langle \mathbf{a}_3, \mathbf{q}_1\rangle \mathbf{q}_1 - \langle \mathbf{a}_3, \mathbf{q}_2\rangle \mathbf{q}_2, & \mathbf{q}_3 = \frac{\mathbf{b}_3}{\|\mathbf{b}_3\|} \\
\dots & \dots
\end{array}$$

Example 2. Find an orthonormal basis for $V = \text{Span}\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\}$.

Solution.

$$\mathbf{b}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \qquad \mathbf{q}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

$$\mathbf{b}_2 = \begin{bmatrix}2\\1\\0\\0\end{bmatrix} - \langle \begin{bmatrix}2\\1\\0\\0\end{bmatrix}, \mathbf{q}_1\rangle \mathbf{q}_1 = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \qquad \mathbf{q}_2 = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

$$\mathbf{b}_3 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \langle \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \mathbf{q}_1\rangle \mathbf{q}_1 - \langle \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \mathbf{q}_2\rangle \mathbf{q}_2 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}, \quad \mathbf{q}_3 = \tfrac{1}{\sqrt{2}} \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

We have obtained an orthonormal basis for V: $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$.

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

Why does Gram-Schmidt work? Recall, if W is a subspace, $\mathbf{b}$ any vector, then

$$\hat{\mathbf{b}} \rightsquigarrow \text{ projection to } W,\ \mathbf{b}^{\perp} = \mathbf{b} - \hat{\mathbf{b}} \rightsquigarrow \text{ orth. to } W$$

Recipe. (Gram-Schmidt orthonormalization)
Given a basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$, produce a orthogonal basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ and an orthonor-

mal basis $\mathbf{q}_1, \ldots, \mathbf{q}_n$.

$$\mathbf{b}_1 = \mathbf{a}_1, \qquad \mathbf{q}_1 = \underbrace{\frac{\mathbf{b}_1}{\|\mathbf{b}_1\|}}_{normalize}$$

$$\mathbf{b}_2 = \mathbf{a}_2 - \underbrace{\langle \mathbf{a}_2, \mathbf{q}_1\rangle \mathbf{q}_1}_{\hat{\mathbf{a_2}} \rightsquigarrow \text{projection to Span}\{\mathbf{q_1}\}} \; \underbrace{\mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1\rangle \mathbf{q}_1}_{\mathbf{a}_2^{\perp} \rightsquigarrow \text{orth. to Span}\{\mathbf{q_1}\}}, \qquad \mathbf{q}_2 = \underbrace{\frac{\mathbf{b}_2}{\|\mathbf{b}_2\|}}_{normalize}$$

$$\mathbf{b}_3 = \underbrace{\mathbf{a}_3 - (\langle \mathbf{a}_3, \mathbf{q}_1\rangle \mathbf{q}_1 + \langle \mathbf{a}_3, \mathbf{q}_2\rangle \mathbf{q}_2)}_{\mathbf{a}_3^{\perp} \text{ orth. to Span}\{\mathbf{q_1},\mathbf{q_2}\}} \qquad \underbrace{\mathbf{q}_3 = \frac{\mathbf{b}_3}{\|\mathbf{b}_3\|}}_{normalize}$$

$$\ldots \qquad \ldots$$

Example 3. Let $V = \text{Span}\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \}$. Find an orthonormal basis for V. Check that your basis is actually orthonormal.

4.2 Orthogonal matrices

Theorem 1. *Let $A = \begin{bmatrix} \mathbf{a_1}, \ldots . \mathbf{a_n} \end{bmatrix}$ be ay matrix. Then A^TA is the matrix of dot products of the columns of A:*

$$A^TA = \begin{bmatrix} \mathbf{a_1} \cdot \mathbf{a_1} & \mathbf{a_1} \cdot \mathbf{a_2} & \mathbf{a_1} \cdot \mathbf{a_3} & \ldots \\ \mathbf{a_2} \cdot \mathbf{a_1} & \mathbf{a_2} \cdot \mathbf{a_2} & \mathbf{a_2} \cdot \mathbf{a_3} & \ldots \\ \mathbf{a_3} \cdot \mathbf{a_1} & \mathbf{a_3} \cdot \mathbf{a_2} & \mathbf{a_3} \cdot \mathbf{a_3} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

What happens if the columns of A are orthonormal?

Theorem 2. *The columns of Q are orthonormal $\iff$ $Q^TQ = I$*

Proof. Let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ be the columns of Q. They orthonormal if and only if $\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$ All these products are packaged in $Q^TQ = I$:

$$\begin{bmatrix} -- & \mathbf{q}_1^T & -- \\ -- & \mathbf{q}_2^T & -- \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

□

Definition. An **orthogonal matrix** is a square matrix Q with orthonormal columns.

It is historical convention to restrict to square matrices, and to say orthogonal matrix even though "orthonormal matrix" might be better.

An $n \times n$ matrix Q is orthogonal $\iff$ $Q^TQ = I$ In other words, $Q^{-1} = Q^T$.

Example 4. $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is orthogonal.Why?

Why is $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ orthogonal?

Solution. Because their columns are a permutation of the standard basis. And so we always have $P^T P = I$. So what is P^{-1}?

Example 5. $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal, Why?

Solution.

- $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ is an orthonormal basis of $\mathbb{R}^2$. Just to make sure: why length 1? Because $\| \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$.

- Alternatively: $Q^T Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

  So what is Q^{-1}?

Example 6. Is $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ orthogonal?

Solution. No, the columns are orthogonal but not normalized. But $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is an orthogonal matrix.

Example 7. (Just for fun) an $n \times n$ matrix with entries ± 1 whose columns are orthogonal is called a *Hadamard matrix* of size n. A size 4 example: $\begin{bmatrix} H & H \\ H & -H \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$ Continuing this construction, we get examples of size $8, 16, 32, \ldots$. It is believed that Hadamard matrices exist for all sizes $4n$. But, no example of size 668 is known yet. If you find one you will be famous!

Math 415 - Lecture 26

Orthogonal Matrices and QR Decomposition

Monday October 26th 2015

Textbook reading: Chapter 3.4

Suggested practice exercises: 3.4: 13, 16, 17, 18. 13,

Khan Academy video: Gram-Schmidt Example

Strang lecture: Orthogonal Matrices and Gram-Schmidt Process.

1 Review

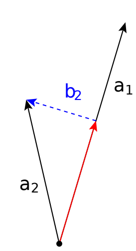
- Vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are orthonormal if

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$

- **Gram-Schmidt** orthonormalization: input: basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ for V. output: orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_n$ for V.

$$\begin{aligned} \mathbf{b}_1 = \mathbf{a}_1, &\quad \mathbf{q}_1 = \tfrac{\mathbf{b}_1}{\|\mathbf{b}_1\|} \\ \mathbf{b}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1, &\quad \mathbf{q}_2 = \tfrac{\mathbf{b}_2}{\|\mathbf{b}_2\|} \\ \mathbf{b}_3 = \mathbf{a}_3 - \langle \mathbf{a}_3, \mathbf{q}_1 \rangle \mathbf{q}_1 - \langle \mathbf{a}_3, \mathbf{q}_2 \rangle \mathbf{q}_2, &\quad \mathbf{q}_3 = \tfrac{\mathbf{b}_3}{\|\mathbf{b}_3\|} \\ \cdots &\quad \cdots \end{aligned}$$

[-1cm]



Fact 1. *if A is any matrix A^TA is the matrix of dot products of the columns of A: Write $A = \begin{bmatrix}\mathbf{a_1}, \dots . \mathbf{a_n}\end{bmatrix}$ then*

$$A^TA = \begin{bmatrix} \mathbf{a_1}\cdot\mathbf{a_1} & \mathbf{a_1}\cdot\mathbf{a_2} & \mathbf{a_1}\cdot\mathbf{a_3} & \dots \\ \mathbf{a_2}\cdot\mathbf{a_1} & \mathbf{a_2}\cdot\mathbf{a_2} & \mathbf{a_2}\cdot\mathbf{a_3} & \dots \\ \mathbf{a_3}\cdot\mathbf{a_1} & \mathbf{a_3}\cdot\mathbf{a_2} & \mathbf{a_3}\cdot\mathbf{a_3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Theorem 1. *The columns of Q are orthonormal $\iff Q^TQ = I$*

Definition. An **orthogonal matrix** is a square matrix Q with orthonormal columns.

2 The QR decomposition

In linear algebra "everything" is a matrix factorization.

- Gaussian elimination in terms of matrices: $A = LU$
- Gram-Schmidt in terms of matrices $A = QR$

Theorem 2 (QR decomposition)**.** *Let A be an $m \times n$ matrix of rank n. There is is a orthogonal matrix $m \times n$-matrix Q and an upper triangular $n \times n$ invertible matrix R such that*

$$A = QR.$$

Idea. Gram-Schmidt on the columns of A to get columns of Q.

Recipe
In general, to obtain $A = QR$:

- Gram-Schmidt on (columns of) A, to get (columns of) Q.
- Then $R = Q^TA$.

The resulting R is indeed upper triangular, and we get:

$$\begin{bmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} | & | & \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T\mathbf{a}_1 & \mathbf{q}_1^T\mathbf{a}_2 & \mathbf{q}_1^T\mathbf{a}_3 & \cdots \\ & \mathbf{q}_2^T\mathbf{a}_2 & \mathbf{q}_2^T\mathbf{a}_3 & \\ & & \mathbf{q}_3^T\mathbf{a}_3 & \\ & & & \ddots \end{bmatrix}$$

It should be noted that, no extra work is needed for computing R: all the inner products in R have been computed during Gram-Schmidt. (Just the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram-Schmidt.)

Example 3. Find the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$.

Solution. We apply Gram-Schmidt to the columns of A:

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = \mathbf{q}_1$$
$$\begin{bmatrix}2\\0\\3\end{bmatrix} - \langle \begin{bmatrix}2\\0\\3\end{bmatrix}, \mathbf{q}_1\rangle \mathbf{q}_1 = \begin{bmatrix}0\\0\\3\end{bmatrix}, \quad \begin{bmatrix}0\\0\\1\end{bmatrix} = \mathbf{q}_2$$
$$\begin{bmatrix}4\\5\\6\end{bmatrix} - \langle \begin{bmatrix}4\\5\\6\end{bmatrix}, \mathbf{q}_1\rangle \mathbf{q}_1 - \langle \begin{bmatrix}4\\5\\6\end{bmatrix}, \mathbf{q}_2\rangle \mathbf{q}_2 = \begin{bmatrix}0\\5\\0\end{bmatrix}, \quad \begin{bmatrix}0\\1\\0\end{bmatrix} = \mathbf{q}_3$$

Solution (continued)**.** Hence: $Q = [\mathbf{q}_1\ \mathbf{q}_2\ \mathbf{q}_3] = \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}$. Note Q is a permutation matrix and so orthogonal. Why? Q has orthonormal columns so $Q^TQ = 1$! To find R in $A = QR$, note that $Q^TA = Q^TQR = R$.

$$R = Q^TA = \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}1&2&4\\0&0&5\\0&3&6\end{bmatrix} = \begin{bmatrix}1&2&4\\0&3&6\\0&0&5\end{bmatrix}$$

Note R is upper triangular. Summarizing, we have

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

Example 4. Find the QR decomposition of $A = \begin{bmatrix}1&1&2\\0&0&1\\1&0&0\end{bmatrix}$.

3 Applications of $A = QR$

3.1 Using QR to solve systems of equations

QR decomposition can be used to solve systems of linear equations.

$$\begin{aligned} A\mathbf{x} = \mathbf{b} \quad &\iff \quad QR\mathbf{x} = \mathbf{b} \\ &\iff \quad R\mathbf{x} = Q^T\mathbf{b} \end{aligned}$$

$R\mathbf{x} = Q^T\mathbf{b}$ is triangular, so solve it by back substitution. QR is a little slower than LU, but makes up in numerical stability.

Theorem 2. *Let A be matrix with linear independent columns. Suppose $A\mathbf{x} = \mathbf{b}$ has no solution. Then the solution of $R\mathbf{x} = Q^T\mathbf{b}$ is the least square solution of $A\mathbf{x} = \mathbf{b}$.*

Proof.

$$\begin{aligned}
A^T A\hat{\mathbf{x}} = A^T\mathbf{b} \quad &\iff \quad \underbrace{(QR)^T QR}_{=R^TQ^TQR}\,\hat{\mathbf{x}} = (QR)^T\mathbf{b}\\
&\iff \quad R^T R\hat{\mathbf{x}} = R^T Q^T\mathbf{b}\\
&\iff \quad R\hat{\mathbf{x}} = Q^T\mathbf{b}
\end{aligned}$$

Again, this is triangular, solved by back substitution.

$\hat{\mathbf{x}}$ is a least square solution of $A\mathbf{x} = \mathbf{b} \iff R\hat{\mathbf{x}} = Q^T\mathbf{b}$ (where $A = QR$)

□

Remark. $R\mathbf{x} = Q^T\mathbf{b}$ always gives the best possible solution to $A\mathbf{x} = \mathbf{b}$.

Example 5. Let $A = \begin{bmatrix} 1 & 1\\ 2 & 3\\ 2 & 1\end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1\\ 0\\ 0\end{bmatrix}$.Find the least square solution of $A\mathbf{x} = \mathbf{b}$ using QR-decomposition.

Solution. Let us first apply Gram-Schmidt to the columns of A. (We are going to work first with unnormalized vectors, and normalize at the end. Check that this also works!) We have $\mathbf{b_1} = \begin{bmatrix} 1\\ 2\\ 2\end{bmatrix}$ and $\mathbf{b_2} = \mathbf{a_2} - \frac{\langle \mathbf{a_2}, \mathbf{b_1}\rangle}{\langle \mathbf{b_1}, \mathbf{b_1}\rangle}\mathbf{b_1} = \begin{bmatrix} 1\\ 3\\ 1\end{bmatrix} - \frac{9}{9}\begin{bmatrix} 1\\ 2\\ 2\end{bmatrix} = \begin{bmatrix} 0\\ 1\\ -1\end{bmatrix}$.

Solution (continued)**.** Normalizing we get

$$Q = \begin{bmatrix} 1/3 & 0\\ 2/3 & 1/\sqrt{2}\\ 2/3 & -1/\sqrt{2}\end{bmatrix}.$$

We have $A = \begin{bmatrix} 1 & 1\\ 2 & 3\\ 2 & 1\end{bmatrix}$, and $Q = \begin{bmatrix} 1/3 & 0\\ 2/3 & 1/\sqrt{2}\\ 2/3 & -1/\sqrt{2}\end{bmatrix}$. Then

$$R = \begin{bmatrix} \mathbf{q_1}\cdot\mathbf{a_1} & \mathbf{q_1}\cdot\mathbf{a_2}\\ 0 & \mathbf{q_2}\cdot\mathbf{a_2}\end{bmatrix} = \begin{bmatrix} 3 & 3\\ 0 & \sqrt{2}\end{bmatrix}.$$

Now $A\mathbf{x} = \mathbf{b}$ is not consistent.

Solution. So we do least squares, but in this case ($A = QR$) we know the normal equations are

$$R\hat{\mathbf{x}} = Q^T\mathbf{b}, \qquad \begin{bmatrix} 3 & 3\\ 0 & \sqrt{2}\end{bmatrix}\hat{\mathbf{x}} = \begin{bmatrix} 1/3 & 2/3 & 2/3\\ 0 & 1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}\begin{bmatrix} 1\\ 0\\ 0\end{bmatrix} = \begin{bmatrix} 1/3\\ 0\end{bmatrix}$$

So $\hat{\mathbf{x}} = \begin{bmatrix} 1/9 \\ 0 \end{bmatrix}$, and $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = 1/9 \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

Math 415 - Lecture 27

An application of QR-decomposition, Change of basis

Friday October 30th 2015

Textbook reading: Chapter 3.4, Chapter 2.6

Suggested practice exercises: Chapter 2.6: Exercises 36, 37, 38,39, 40,43

Khan Academy video: Change of basis

Strang lecture: Change of basis; image compression

1 Review

Theorem 1 (QR decomposition)**.** *Let A be a $m \times n$ matrix of rank n with linear independent columns. There is an orthogonal matrix $m \times n$-matrix Q and an upper triangular $n \times n$ invertible matrix R such that*

$$A = QR.$$

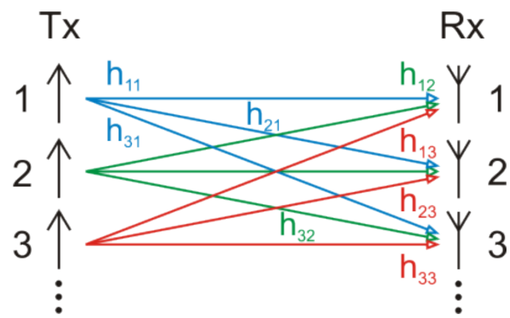
Theorem 2. *Let A be a matrix with linear independent columns. Suppose $A\mathbf{x} = \mathbf{b}$ has no solution. Then the solution of $R\mathbf{x} = Q^T\mathbf{b}$ is the least square solution of $A\mathbf{x} = \mathbf{b}$.*

2 An application in wireless communication

In multiple-input and multiple-output (short: MIMO) systems, a transmitter sends multiple streams by multiple transmit antennas. Let us suppose there are n transmitters and m receivers. This can modelled using Linear Algebra:

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{\text{received vector } \mathbf{y}} = \underbrace{\begin{bmatrix} h_{1,1} & \dots & h_{1,n} \\ \vdots & \ddots & \vdots \\ h_{m,1} & \dots & h_{m,n} \end{bmatrix}}_{\text{channel matrix } H} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{transmitted vector } \mathbf{x}} .$$

Suppose that the channel matrix H is known both to person A who sending information and to person B who is receiving the information.



Let us try and understand the engineering meaning of some of the linear algebra of the matrix H and the equation $y = Hx$. Remember: the x vector describes what the transmiter is sending out and y is the vector describing what is received.

We want to understand

$$y = Hx.$$

- What is the first column of H?
- What is Nul(H)? If the signal x belongs to the nullspace, what signal y will be received?
- In a well designed system you want Dim(Nul(H)) =?
- What is $Col(H)$?

When B receives the signal, she wants to reconstruct the vector $\mathbf{x}$. Optimally, she would just solve the linear system

$$H\mathbf{x} = \mathbf{y}.$$

Unfortunately, almost always B received $\mathbf{y} + \epsilon$ instead of $\mathbf{y}$, where $\epsilon \in \mathbb{R}^m$ is noise.

So B would try to solve

$$H\mathbf{x} = \mathbf{y} + \epsilon.$$

instead. However, that system might not have a solution. So B has to find the least square solution! Because B receives many messages from A, she will have to find the least square solution many times. Luckily, H does not change, and has independent columns (Nul(H) = 0). So B determines the QR-decomposition of H

$$H = QR,$$

once, and then just solves

$$R\mathbf{x} = Q^T(\mathbf{y} + \epsilon)$$

each time she receives a new message. This is easy to do, since R is upper triangular.

3 Linear transformation revisited

Recall the notion of coordinate vectors. If $\mathcal{B} = \{b_1, b_2\}$ is a basis for $\mathbb{R}^2$, and x some vector then the coordinate vector of x is $x_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ precisely if $x = c_1 b_1 + c_2 b_2$. We want to understand how to relate coordinate vectors $x_{\mathcal{B}}$ and $x_{\mathcal{C}}$ for different bases $\mathcal{B}$ and $\mathcal{C}$. We will see that there is for every two bases a matrix $I_{\mathcal{C},\mathcal{B}}$ so that

$$x_{\mathcal{C}} = I_{\mathcal{C},\mathcal{B}} x_{\mathcal{B}}.$$

Remember Theorem 1 of Lecture 17? Here it is again.

Theorem 3. *Let $\mathcal{B}$ be a basis of $\mathbb{R}^m$ and $\mathcal{C}$ be a basis of $\mathbb{R}^n$ and let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then there is a $n \times m$ matrix $T_{\mathcal{C},\mathcal{B}}$ such that for every $\mathbf{v} \in \mathbb{R}^m$*

$$T(\mathbf{v})_{\mathcal{C}} = T_{\mathcal{C},\mathcal{B}} \mathbf{v}_{\mathcal{B}}.$$

and

$$T_{\mathcal{C},\mathcal{B}} = \begin{bmatrix} T(\mathbf{v_1})_{\mathcal{C}} & T(\mathbf{v_2})_{\mathcal{C}} & \dots & T(\mathbf{v_m})_{\mathcal{C}} \end{bmatrix}$$

where $\mathcal{B} = \left(\mathbf{v_1}; \dots; \mathbf{v_m}\right)$.

We will use this first in the special case $T = I$, where $I(v) = v$ (seemingly boring!).

Example 4. Consider $\mathcal{E} := \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$ and $\mathcal{B} := \{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$. Let $I : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation (the Identity!)

$$I(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Find the matrix $I_{\mathcal{E},\mathcal{B}}$ that represents I with respect to the input basis $\mathcal{B}$ and output basis $\mathcal{E}$.

Solution. By definition the matrix $I_{\mathcal{E},\mathcal{B}}$ has as first column b_1 expressed in the standard basis, and as second column b_2 also expressed in the standard basis. But for any vector $x \in \mathbb{R}^n$ we have $x_{\mathcal{E}} = x$! So

$$I_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}.$$

Example 5. Given $\mathbf{v} \in \mathbb{R}^2$ what is $I_{\mathcal{E},\mathcal{B}} \mathbf{v}_{\mathcal{B}}$?

Solution. Let $\mathbf{v}_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$. Then

$$I_{\mathcal{E},\mathcal{B}} \mathbf{v}_{\mathcal{B}} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = c_1 b_1 + c_2 b_2 = v!$$

Suppose $\mathbf{v}$ is a vector in $\mathbb{R}^n$, and we have two bases in $\mathbb{R}^n$. so that we get two coordinate vectors $\mathbf{v}_{\mathcal{C}}$ and $\mathbf{v}_{\mathcal{B}}$. How are they related?

Theorem 6. *Let $\mathcal{B}$ be a basis of $\mathbb{R}^n$ and $\mathcal{C}$ be another basis of $\mathbb{R}^n$ and let $I : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation such that $I(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v} \in \mathbb{R}^n$. Then*

$$\mathbf{v}_{\mathcal{C}} = I_{\mathcal{C},\mathcal{B}}\mathbf{v}_{\mathcal{B}}.$$

We call the matrix $I_{\mathcal{C},\mathcal{B}}$ a **change of base matrix**, it transforms coordinate vectors from the $\mathcal{B}$ to the $\mathcal{C}$ basis.

Example 7. Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^n$ and $\mathcal{B}$ be another basis of $\mathbb{R}^n$. What is $I_{\mathcal{E},\mathcal{B}}$?

Solution. The columns of $I_{\mathcal{E},\mathcal{B}}$ are the basic vectors $b_1, b_2, \dots$ expressed in the standard basis. So

$$I_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}$$

So this is the *easy* change of basis matrix: you just write down the $\mathcal{B}$ basis as columns of your matrix. It has the property that

$$v = v_{\mathcal{E}} = I_{\mathcal{E},\mathcal{B}} v_{\mathcal{B}}$$

Example 8. Let $\mathcal{B}$ be a basis of $\mathbb{R}^n$ and $\mathcal{C}$ be a basis of $\mathbb{R}^n$. What is $I^{-1}_{\mathcal{C},\mathcal{B}}$?

Solution. $I_{\mathcal{C},\mathcal{B}}$ is the matrix with columns the $\mathcal{B}$ basis vectors expressed in the $\mathcal{C}$ basis, and $I^{-1}_{\mathcal{C},\mathcal{B}}$ is the inverse of this matrix. These matrices have the property that

$$v_{\mathcal{C}} = I_{\mathcal{C},\mathcal{B}} v_{\mathcal{B}}, \quad v_{\mathcal{B}} = I^{-1}_{\mathcal{C},\mathcal{B}} v_{\mathcal{C}}.$$

Example 9. As before, let $\mathcal{E} := \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$ and $\mathcal{B} := \{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$. What is $I_{\mathcal{B},\mathcal{E}}$?

Solution. We know what $I_{\mathcal{E},\mathcal{B}}$ is, it is just $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Then $I_{\mathcal{B},\mathcal{E}}$ is the transition matrix going the other way, so it is the inverse of the *easy* matrix, so

$$I_{\mathcal{B},\mathcal{E}} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Example 10. Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^n$ and $\mathcal{C}$ be a orthonormal basis of $\mathbb{R}^n$. Then $I_{\mathcal{C},\mathcal{E}} = I^T_{\mathcal{E},\mathcal{C}}$. Why?

Solution. $I_{\mathcal{E},\mathcal{C}}$ the matrix with orthonormal columns, so it is an orthogonal matrix. $I_{\mathcal{C},\mathcal{E}}$ is the inverse. But the inverse of an orthogonal matrix is easy, just the transpose.

Theorem 1. *Let $\mathcal{B} := (\mathbf{u_1}, \dots, \mathbf{u_n})$ be an orthonormal basis of $\mathbb{R}^n$ and $U = \begin{bmatrix} \mathbf{u_1} & \dots & \mathbf{u_n} \end{bmatrix}$. Then for every $\mathbf{v} \in \mathbb{R}^n$*

$$v_{\mathcal{B}} = U^T v.$$

4 Change of basis

Theorem 11. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation and $\mathcal{A}$ and $\mathcal{B}$ be two bases of $\mathbb{R}^m$ and $\mathcal{C}, \mathcal{D}$ be two bases of $\mathbb{R}^n$. Then*

$$T_{\mathcal{C},\mathcal{A}} = I_{\mathcal{C},\mathcal{D}} T_{\mathcal{D},\mathcal{B}} I_{\mathcal{B},\mathcal{A}}.$$

$$\begin{tikzcd}
(\mathbb{R}^m, \mathcal{A}) \arrow[r, "\text{apply } T_{\mathcal{C},\mathcal{A}}"] \arrow[d, "I_{\mathcal{B},\mathcal{A}}"'] & (\mathbb{R}^n, \mathcal{C}) \\
(\mathbb{R}^m, \mathcal{B}) \arrow[r, "\text{apply } T_{\mathcal{D},\mathcal{B}}"] & (\mathbb{R}^n, \mathcal{D}) \arrow[u, "I_{\mathcal{C},\mathcal{D}}"]
\end{tikzcd}$$

Example 12. Consider $\mathcal{B} := \mathcal{D} := \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$ and $\mathcal{A} := \mathcal{C} := \{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$ as before. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be again the linear transformation that

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto A \begin{bmatrix} x \\ y \end{bmatrix}, \text{ where } A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Determine $T_{\mathcal{C},\mathcal{C}}$.

Solution. By Theorem 11

$$T_{\mathcal{C},\mathcal{C}} = I_{\mathcal{C},\mathcal{B}} T_{\mathcal{B},\mathcal{B}} I_{\mathcal{B},\mathcal{C}}.$$

In Lecture 27, we already calculated that

$$I_{\mathcal{C},\mathcal{B}} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, I_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Since $\mathcal{B}$ is the standard basis,

$$T_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Therefore

$$T_{\mathcal{C},\mathcal{C}} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 1 & 5 \end{bmatrix}$$

.

Example 13. Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^n$, let $\mathcal{B} := (\mathbf{u_1}, \dots, \mathbf{u_n})$ be an orthonormal basis of $\mathbb{R}^n$ and $U = \begin{bmatrix} \mathbf{u_1} & \dots & \mathbf{u_n} \end{bmatrix}$ Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then

$$T_{\mathcal{B},\mathcal{B}} = U^T T_{\mathcal{E},\mathcal{E}} U.$$

Why?

Solution.

Math 415 - Lecture 28

Change of base, Image Compression

Monday November 2nd 2015

Textbook reading: Notes by Strang

Suggested practice exercises:

Khan Academy video:

Strang lecture: Change of basis; image compression

1 Review

* If $\mathcal{B}, \mathcal{C}$ are bases of $\mathbb{R}^n$, get *coordinate vectors* $x_{\mathcal{B}}, x_{\mathcal{C}}$ for any $x \in \mathbb{R}$.

* *Change of basis matrices*: $I_{\mathcal{C},\mathcal{B}}, I_{\mathcal{B},\mathcal{C}}$ such that

$$x_{\mathcal{B}} = I_{\mathcal{B},\mathcal{C}} x_{\mathcal{C}}, \quad x_{\mathcal{C}} = I_{\mathcal{C},\mathcal{B}}.$$

* Inverses: $I_{\mathcal{C},\mathcal{B}}^{-1} = I_{\mathcal{B},\mathcal{C}}$.

* Easy case: If $\mathcal{E}$ is the standard basis: then

$$I_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}, \quad I_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}^{-1}.$$

Theorem 1. *Let $\mathcal{U} := (\mathbf{u_1}, \dots, \mathbf{u_n})$ be an orthonormal basis of $\mathbb{R}^n$ and $U = \begin{bmatrix} \mathbf{u_1} & \dots & \mathbf{u_n} \end{bmatrix}$. Then for every $\mathbf{v} \in \mathbb{R}^n$*

$$v_{\mathcal{U}} = U^T v.$$

Why? $I_{\mathcal{E},\mathcal{U}} = \begin{bmatrix} \mathbf{u_1} & \dots & \mathbf{u_n} \end{bmatrix} = U$. But U has orthonormal columns, so $I_{\mathcal{U},\mathcal{E}} = U^{-1} = U^T$.

Example 1. Let $\mathcal{U} := (\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix})$. Determine $\begin{bmatrix} 2 \\ 4 \end{bmatrix}_{\mathcal{U}}$.

Solution. We have $U = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$. This is the change of basis matrix from the $\mathcal{U}$ basis to the standard basis. So to go the other direction take the inverse. In this case inverse is transpose, so

$$\begin{bmatrix}2\\4\end{bmatrix}_{\mathcal{U}} = U^T\begin{bmatrix}2\\4\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}6\\-2\end{bmatrix}.$$

Check:

$$\begin{bmatrix}2\\4\end{bmatrix} = \frac{6}{2}\begin{bmatrix}1\\1\end{bmatrix} + \frac{-2}{2}\begin{bmatrix}1\\-1\end{bmatrix}$$

Example 2. Let $\mathcal{B} := (\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix})$. Let $A = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2}\end{bmatrix}$. How can you easily compute $A_{\mathcal{B}} := \begin{bmatrix}\mathbf{a_1}_{\mathcal{B}} & \mathbf{a_2}_{\mathcal{B}}\end{bmatrix}$, ie the matrix whose are $\mathcal{B}$-coordinates of the columns of A?

Solution. To get the $\mathcal{B}$ coordinate vectors, multiply each column of A by U^T, where $U = \begin{bmatrix}u_1 & u_2\end{bmatrix}$. So the wanted matrix is $A_{\mathcal{B}} = U^T A$.

Theorem 2. *Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^n$, let $\mathcal{B} := (\mathbf{u_1}, \dots, \mathbf{u_n})$ be an orthonormal basis of $\mathbb{R}^n$ and $U = \begin{bmatrix}\mathbf{u_1} & \dots & \mathbf{u_n}\end{bmatrix}$. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then*

$$T_{\mathcal{B},\mathcal{B}} = U^T T_{\mathcal{E},\mathcal{E}} U,$$

or equivalently,

$$T_{\mathcal{E},\mathcal{E}} = U T_{\mathcal{B},\mathcal{B}} U^T.$$

Example 3. Let $\mathcal{B} := (\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix})$. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformation given by

$$T(\mathbf{v}) = \begin{bmatrix}\frac{1}{2} & \frac{3}{2}\\ \frac{3}{2} & \frac{1}{2}\end{bmatrix}\mathbf{v}.$$

Determine $T_{\mathcal{B},\mathcal{B}}$!

Solution. $T_{\mathcal{B},\mathcal{B}} = I_{\mathcal{B}\mathcal{E}} A I_{\mathcal{E}\mathcal{B}}$, where $A = T_{\mathcal{E}\mathcal{E}}$ is the matrix of T with respect to the standard basis. Now $I_{\mathcal{E}\mathcal{B}} = \begin{bmatrix}b_1 & b_1\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = U$ So

$$T_{\mathcal{B},\mathcal{B}} = U^T A U = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{2} & \frac{3}{2}\\ \frac{3}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \begin{bmatrix}2 & 0\\ 0 & -1\end{bmatrix}$$

Use $T_{\mathcal{B},\mathcal{B}}$ to calculate $T(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix})$.

Solution.

$$T(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix})_{\mathcal{B}} = T_{\mathcal{B}\mathcal{B}}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}_{\mathcal{B}} = \begin{bmatrix}2 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}.$$

This means that $T(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}) = 2\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$.

We will call such vectors eigenvectors and the number 2 will be called an eigenvalue. More about this soon!

2 Data compression

Let consider the following basis $\mathcal{H}$ of $\mathbb{R}^8$:

$$\left(\begin{bmatrix} \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \right)$$

(i) Is $\mathcal{H}$ orthogonal?

(ii) Is $\mathcal{H}$ orthonormal?

This basis $\mathcal{H}$ is called **Haar Wavelet basis**. We will see in the following that $\mathcal{B}$ is much more effective than the standard basis (at least for certain applications).

Example 4. Find the coordinate vector of $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 88 \\ 90 \\ 92 \\ 93 \\ 92 \\ 92 \\ 94 \\ 95 \end{bmatrix}$ with respect to $\mathcal{H}$?

Solution.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{H}} = \begin{bmatrix} \sqrt{8} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 88 \\ 90 \\ 92 \\ 93 \\ 92 \\ 92 \\ 94 \\ 95 \end{bmatrix}_{\mathcal{H}} = \begin{bmatrix} 260.2 \\ -3.5 \\ -3.5 \\ -2.5 \\ -1.4 \\ -0.7 \\ 0 \\ -0.7 \end{bmatrix}.$$

How could one use that for (lossy) data compression?

- Pick $\epsilon > 0$, and set all entries of the vectors with absolute value at most ϵ to 0.
- The $\mathcal{H}$-coordinate vector has more of these small entries, so more values become 0.

Let's do it with pictures!

2.1 Image compression

Consider 8×8-matrix, i.e., a 8×8-grayscale picture:

$$A = \begin{bmatrix} 88 & 88 & 89 & 90 & 92 & 94 & 96 & 97 \\ 90 & 90 & 91 & 92 & 93 & 95 & 97 & 97 \\ 92 & 92 & 93 & 94 & 95 & 96 & 97 & 97 \\ 93 & 93 & 94 & 95 & 96 & 96 & 96 & 96 \\ 92 & 93 & 95 & 96 & 96 & 96 & 96 & 95 \\ 92 & 94 & 96 & 98 & 99 & 99 & 98 & 97 \\ 94 & 96 & 99 & 101 & 103 & 103 & 102 & 101 \\ 95 & 97 & 101 & 104 & 106 & 106 & 105 & 105 \end{bmatrix}$$

Let suppose we want to replace each column of A by its $\mathcal{H}$-coordinate. By Theorem, we have to calculate $H^T A$, where

$$H = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

We get

$$H^T A = \begin{bmatrix} 260.22 & 263.4 & 267.29 & 268.35 & 268.35 & 273.3 & 282.49 & 289.56 \\ -3.54 & -6.72 & -4.95 & -3.18 & -2.47 & -4.6 & -6.72 & -8.84 \\ -3.5 & -1.5 & -1.5 & -1.5 & -3. & -4. & -5. & -6.5 \\ -2.5 & -3. & -1.5 & 0. & 0.5 & 1.5 & 1.5 & 1. \\ -1.41 & 0. & 0. & 0. & -0.71 & -1.41 & -1.41 & -1.41 \\ -0.71 & -0.71 & -0.71 & -0.71 & -0.71 & -1.41 & -1.41 & -2.12 \\ 0. & -1.41 & -0.71 & 0. & 0. & 0. & 0. & 0. \\ -0.71 & 0. & 0. & 0. & 0.71 & 0.71 & 0.71 & 0. \end{bmatrix}$$

How could one use that for (lossy) image compression?

- Pick $\epsilon > 0$, and set all entries of $H^T A$ with absolute value at most ϵ to 0.
- $H^T A$ has more of these small entries, so more values become 0.

Already good, but we can do even better! Replace the rows of $H^T A$ by their $\mathcal{H}$-coordinates. For that we just need to calculate $H^T AH$! Why? We calculate

$$H^T AH = \begin{bmatrix} 768.25 & -19.25 & -6.01 & -15.2 & -2.25 & -0.75 & -3.5 & -5. \\ -14.5 & 1.5 & -1.06 & 4.24 & 2.25 & -1.25 & 1.5 & 1.5 \\ -9.37 & 3.71 & -1. & 2.25 & -1.41 & 0. & 0.71 & 1.06 \\ -0.88 & -4.07 & -2. & -0.25 & 0.35 & -1.06 & -0.71 & 0.35 \\ -2.25 & 1.25 & -0.71 & 0.35 & -1. & 0. & 0.5 & 0. \\ -3. & 1. & 0. & 0.71 & 0. & 0. & 0.5 & 0.5 \\ -0.75 & -0.75 & -0.35 & 0. & 1. & -0.5 & 0. & 0. \\ 0.5 & -1. & -0.35 & 0.35 & -0.5 & 0. & 0. & 0.5 \end{bmatrix}.$$

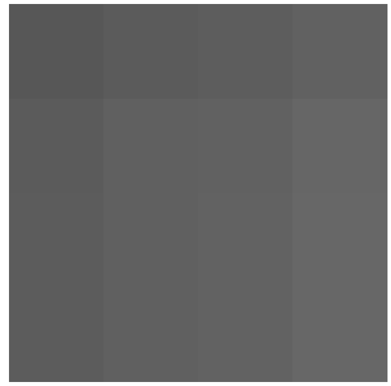
This just a change of base: $H^T AH = A_{\mathcal{H},\mathcal{H}}$. Image compression algorithms usually then divide every entry by an integers (this process is called quantization) and then rounds each entries to the nearest integer. Say we divide here by 12, then our matrix becomes the matrix B:

$$B = \begin{bmatrix} 64 & -2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

To recover an image, we have to reverse the process. How do you do that? So let's calculate $H(12B)H^T$:

$$\begin{bmatrix} 87.3 & 87.3 & 91.5 & 91.5 & 93.3 & 93.3 & 97.5 & 97.5 \\ 87.3 & 87.3 & 91.5 & 91.5 & 93.3 & 93.3 & 97.5 & 97.5 \\ 91.5 & 91.5 & 95.7 & 95.7 & 97.5 & 97.5 & 101.7 & 101.7 \\ 91.5 & 91.5 & 95.7 & 95.7 & 97.5 & 97.5 & 101.7 & 101.7 \\ 92.4 & 92.4 & 96.6 & 96.6 & 98.4 & 98.4 & 102.6 & 102.6 \\ 92.4 & 92.4 & 96.6 & 96.6 & 98.4 & 98.4 & 102.6 & 102.6 \\ 92.4 & 92.4 & 96.6 & 96.6 & 98.4 & 98.4 & 102.6 & 102.6 \\ 92.4 & 92.4 & 96.6 & 96.6 & 98.4 & 98.4 & 102.6 & 102.6 \end{bmatrix}$$

Let's compare the images. The original is on the left, the compressed image on the right:

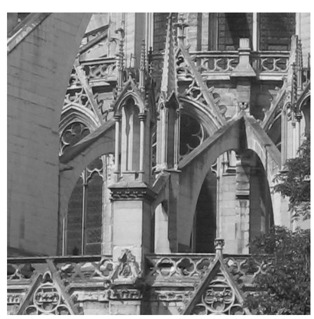
The compression ratio of an image is the ratio of the non-zero elements in the original matrix to the non-zero elements in the matrix representing the compressed image. The matrix

$$\begin{bmatrix} 64 & -2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

has only 6 non-zero entry, while matrix A has 64. So the compression ratio is 64/6. That's pretty high!

2.2 JPEG

So does JPEG works? Given an image, let's say a 512×512 pixel grayscale image of the flying buttresses of the Notre Dame Cathedral in Paris:



This picture is split into blocks of 8×8-pixels. The block in top left corner is given by our matrix A. As the next step the JPEG algorithm does precisely what we did above.

Math 415 - Lecture 29

Determinants

Wednesday November 4th 2015

Textbook reading: Chapters 4.2, 4.3

Suggested practice exercises: Chapter 4.2, # 1, 2, 4, 5, 10, 14, 15, 17, 18, 19, 20, 22, 23

Khan Academy video: 3×3 Determinant, $n \times n$ Determinant, Determinants along other rows/ columns,

Strang lecture: Lecture 18: Properties of determinants, Lecture 19: Determinant formulas and cofactors

1 Determinants

For the next few lectures, all matrices are square! Recall that

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

The determinant of

- a 2×2 matrix is $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$,
- a 1×1 matrix is $\det([a]) = a$.
- What is the determinant of an $n \times n$ matrix?

Goal (Point of the determinant)
A is invertible $\iff \det(A) \neq 0$

Notation: We will write both $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$ and $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ for the determinant.

We want to define for every $n \times n$ matrix A a number $\det(A)$.

Definition. The determinant is characterized by:

- the normalization $\det I_{n\times n} = 1$,
- and how it is affected by elementary row operations:
  - **(Replacement)** Add a multiple of one row to another row. Does not change the determinant.
  - **(Interchange)** Interchange two rows. Reverses the sign of the determinant.
  - **(Scaling)** Multiply all entries in a row by s. Multiplies the determinant by s.

This allows us to compute the determinant using just row operations!

Important Fact
The determinant of a triangular matrix is the product of the diagonal entries.

Example

$$\det \begin{bmatrix} 2 & 3 & 7 \\ 0 & 4 & 8 \\ 0 & 0 & 6 \end{bmatrix} = 2 \cdot 4 \cdot 6.$$

Why? Take out the diagonal entries, and then use row operations to get the identity matrix.

Example 1 (Generic matrix). Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$.

Solution.

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} &\overset{\substack{R2 \to R2 - 3R1 \\ R3 \to R3 - 2R1}}{=} \begin{vmatrix} 1 & 2 & 0 \\ 0 & -7 & 2 \\ 0 & -4 & 1 \end{vmatrix} \\ &\overset{R3 \to R3 - \frac{4}{7}R2}{=} \begin{vmatrix} 1 & 2 & 0 \\ 0 & -7 & 2 \\ 0 & 0 & -\frac{1}{7} \end{vmatrix} \\ &= 1 \cdot (-7) \cdot (-\frac{1}{7}) = 1 \end{aligned}$$

Example 2 (Reality check). Discover the formula for $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

Solution.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \overset{R2 \to R2 - \frac{c}{a}R1}{=} \begin{vmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{vmatrix} = a\left(d - \frac{c}{a}b\right) = ad - bc$$

NB: this only works if $a \neq 0$. What do you do if $a = 0$?

Example 3 (Larger matrix). Compute $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix}$.

Solution. The matrix looks complicated, but we can do this in two steps!

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix} \overset{R4 \to R4 - \frac{3}{2}R3}{=} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & \frac{7}{2} \end{vmatrix} = 1 \cdot 2 \cdot 2 \cdot \frac{7}{2} = 14$$

The following important properties follow from behavior under row operations.

Important properties

- $\det(A) = 0 \iff A$ is not invertible. Why? Because $\det(A) = 0$, if and only if, in an echelon form, a diagonal entry is zero (that is, a pivot is missing).
- $\det(AB) = \det(A)\det(B)$ **Challenge**: Figure out why! (Matrix multiplication can be seen as linear combinations of rows)
- $\det(A^{-1}) = \frac{1}{\det(A)}$ Why? Because $AA^{-1} = I$. Since $\det(AA^{-1}) = \det(I) = 1$ and $\det(AA^{-1}) = \det(A)\det(A^{-1})$, we have $\det(A)\det(A^{-1}) = 1$.
- $\det(A^T) = \det(A)$. (Think about why this works at home.)

Remark. $\det(A^T) = \det(A)$ means that everything you know about determinants in terms of *rows* of A is also true for the columns.For instance:

- If you exchange two *columns* in a determinant you get a minus sign.
- You can add a multiple of a *column* to another column without changing the determinant.
- If your matrix has equal *columns* the determinant is zero.
- If your matrix has a zero *column* the determinant is zero.

Example 4. Recall that $AB = \mathbf{0}$, then it does not follow that $A = \mathbf{0}$ or $B = \mathbf{0}$. However, show that $\det(A) = 0$ or $\det(B) = 0$.

Solution. If $AB = \mathbf{0}$, then $\det(AB) = \det(\mathbf{0}) = 0$. Follows from $\det(AB) = \det(A)\det(B)$.

2 A "bad" way to compute determinants, Cofactor expansion

Idea. The determinant is linear in each row or column.

(In some texts this is one of the basic assumptions about det.)

Fact 5.

$$\det\begin{bmatrix} a & b & c \\ * & * & * \\ * & * & * \end{bmatrix} = \det\begin{bmatrix} a & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix} + \det\begin{bmatrix} 0 & b & 0 \\ * & * & * \\ * & * & * \end{bmatrix} + \det\begin{bmatrix} 0 & 0 & c \\ * & * & * \\ * & * & * \end{bmatrix}$$

We can use this idea to calculate an $n \times n$ determinant in terms of n determinants of smaller matrices.

Example 6. What is the determinant $\begin{bmatrix} a & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix}$? What about $\begin{bmatrix} 0 & b & 0 \\ * & * & * \\ * & * & * \end{bmatrix}$?

Solution.

$$\det\begin{bmatrix} a & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix} = a\det\begin{bmatrix} 1 & 0 & 0 \\ * & * & * \\ * & * & * \end{bmatrix} = a\det\begin{bmatrix} 1 & 0 \\ * & B \end{bmatrix} = a\det\begin{bmatrix} B \end{bmatrix},$$

where B is the 2×2 right lower block. Same way, with a twist:

$$\det\begin{bmatrix} 0 & b & 0 \\ v_1 & v_2 & v_3 \end{bmatrix} = \boxed{-1} b\det\left(\begin{bmatrix} 1 & 0 & 0 \\ v_2 & v_1 & v_3 \end{bmatrix}\right) = -b\det\begin{bmatrix} v_1 & v_3 \end{bmatrix}.$$

We can use this idea to calculate an $n \times n$ determinant in terms of n determinants of $(n-1) \times (n-1)$ matrices. Then repeat ....

Example 7. Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ by **cofactor expansion**.

Solution. We expand by the first row:

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

$$\overset{\text{i.e.}}{=} 1 \cdot \begin{vmatrix} -1 & 2 \\ 0 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} + 0 \cdot \begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} = 1 \times (-1) - 2 \cdot (-1) + 0 = 1$$

Each term in the cofactor expansion is ± 1 times an entry times a smaller determinant (row and column of entry deleted). The $\pm$ is assigned to each entry according to

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

There is nothing special about the first row. We can use any other row or column.

For example, let's use the second column:

Solution.

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

$$= -2 \cdot (-1) + (-1) \cdot 1 - 0 = 1$$

Same answer!

Let use the third column:

Solution.

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

$$= 0 - 2 \cdot (-4) + 1 \cdot (-7) = 1$$

Same answer!

Why not cofactor expansion

Why is the method of cofactor expansion not practical (except when there are lots of zeroes in your matrix.)? Because to compute a large $n \times n$ matrix,

- one reduces to n determinants of size $(n-1) \times (n-1)$,
- then $n(n-1)$ determinants of size $(n-2) \times (n-2)$,
- and so on.

In the end, we have $n! = n(n-1) \cdots 3 \cdot 2 \cdot 1$ many numbers to add.WAY TOO MUCH WORK! Already

$$25! = 15511210043330985984000000 \approx 1.55 \cdot 10^{25}.$$

Context: today's fastest computer, Tianhe-2, runs at 34 pflops ($3.4 \cdot 10^{16}$ operations per second).By the way: "fastest" is measured by computing LU decompositions!

3 Practice Problems

3.1

Example 8. Compute $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 0 & 0 \\ 2 & 7 & 6 & 10 \\ 2 & 9 & 7 & 11 \end{vmatrix}$. Use your favorite method (or a mix of methods!)

Solution. The final answer should be -10.

- What's **wrong**?!

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

The correct calculation is:

$$\det \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(ad-bc)^2}\det\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}$$

Example 9. Suppose A is a 3×3 matrix with $\det(A) = 5$. What is $\det(2A)$?

Solution. A has three rows. Multiplying all 3 of them produces $2A$. Hence, $\det(2A) = 2^3\det(A) = 40$.

Imaginary unit and Fibonacci numbers

Example 10. First off, say hello to our new friend: i, the **imaginary unit**. It is infamous for $\boxed{i^2=-1}$. Let us calculate some determinants.

$$\begin{aligned}
|1| &= 1 \\
\begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} &= 1 - i^2 = 2 \\
\begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} &= 1\begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} - i\begin{vmatrix} i & 0 \\ i & 1 \end{vmatrix} = 2 - i^2 = 3 \\
\begin{vmatrix} 1 & i & & \\ i & 1 & i & \\ & i & 1 & i \\ & & i & 1 \end{vmatrix} &= 1\begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} - i\begin{vmatrix} i & 0 & \\ i & 1 & i \\ & i & 1 \end{vmatrix} = 3 - i^2\begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} = 5
\end{aligned}$$

Example 11 (continued).

$$\begin{vmatrix} 1 & i & & & \\ i & 1 & i & & \\ & i & 1 & i & \\ & & i & 1 & i \\ & & & i & 1 \end{vmatrix} = 1\begin{vmatrix} 1 & i & & \\ i & 1 & i & \\ & i & 1 & i \\ & & i & 1 \end{vmatrix} - i^2\begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} = 5 + 3 = 8$$

The Fibonacci numbers!

Do you know about the connection of Fibonacci numbers and rabbits? If not, Google is your friend. 

Math 415 - Lecture 30

Eigenvectors and Eigenvalues

Friday November 6th 2015

Textbook reading: Chapter 5.1

Suggested practice exercises: 12, 20, 21, 22, 36

Khan Academy video: Introduction to Eigenvalues and Eigenvectors, Proof of formula for determining Eigenvalues, Finding Eigenvectors and Eigenspaces example

Strang lecture: Lecture 21: Eigenvalues and eigenvectors

1 Review

Definition. The determinant is characterized by:

- the normalization $\det I_{n\times n} = 1$,
- and how it is affected by elementary row operations:
  - **(Replacement)** Add a multiple of one row to another row. Does not change the determinant.
  - **(Interchange)** Interchange two rows. Reverses the sign of the determinant.
  - **(Scaling)** Multiply all entries in a row by s. Multiplies the determinant by s.
- For triangular A the determinant is just product of the diagonal entries.

This allows us to compute the determinant using just row operations!. Bring A into echelon form= triangular form, keeping track how the determinant changes under the row operations you are using.

- What's **wrong**?!

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

The correct calculation is:

$$\det \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(ad-bc)^2}\det \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}$$

Example 1. Suppose A is a 3×3 matrix with $\det(A) = 5$. What is $\det(2A)$?

Solution 2. *A has three rows. Multiplying all 3 of them produces $2A$. Hence, $\det(2A) = 2^3 \det(A) = 40$.*

2 Eigenvectors and eigenvalues

2.1 Definition

Throughout, A will be an $n \times n$ matrix.

Definition. An **eigenvector** of A is a nonzero $\mathbf{x}$ such that

$$A\mathbf{x} = \lambda\mathbf{x} \text{ for some scalar } \lambda.$$

The scalar λ is the corresponding **eigenvalue**.

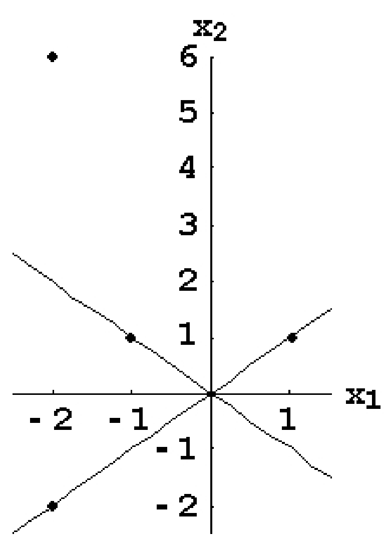
In words, eigenvectors are those $\mathbf{x}$, for which $A\mathbf{x}$ is parallel to $\mathbf{x}$.

Example 3. Verify that $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$. Is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ an eigenvector?

Solution.

$$A\mathbf{x} = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix} = -2\mathbf{x}$$

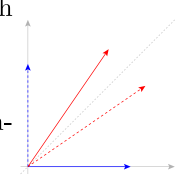
Hence, $\mathbf{x}$ is an eigenvector of A with eigenvalue -2.



2.2 Geometric interpretation

Example 4. Use your geometric understanding to find the eigenvectors and the eigenvalues of $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Solution. $A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$, i.e. multiplication with A is reflection through the line $y = x$.



- $A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. So $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 1$.

- $A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ So $\mathbf{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = -1$.

Example 5. Use your geometric understanding to find the eigenvectors and the eigenvalues of $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution. $A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$, i.e. multiplication with A is projection on the x-axis.

- $A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. So $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 1$.

- $A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ So $\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 0$.

Summary

* Eigenvectors $\mathbf{x}$ get stretched by eigenvalue λ under multiplication by A:

$$A\mathbf{x} = \lambda\mathbf{x}.$$

* Eigenvectors $\mathbf{x}$ **CANNOT** be zero. Why? $A\mathbf{0} = \lambda\mathbf{0}$ for any λ. Not useful!

* Eigenvalues λ **CAN** be zero. See the projection example.

Problems

* How to find possible eigenvalues for A? This uses determinants.

* How to find eigenvectors? This uses null spaces.

3 Eigenspaces

Definition. The **eigenspace** of A corresponding to λ is the set of all $\mathbf{x}$ satisfying $A\mathbf{x} = \lambda\mathbf{x}$. It consists of all the eigenvectors of A with eigenvalue λ, and also the zero vector.

Example 6. We saw the projection matrix P of the projection onto a subspace V has two eigenvalues $\lambda = 0, 1$.

- The eigenspace of $\lambda = 1$ is V.
- The eigenspace of $\lambda = 0$ is $V^\perp$.

4 How to solve $A\mathbf{x} = \lambda\mathbf{x}$

Key observation: $\mathbf{x} \neq 0$ is an eigenvector means:

$$\begin{aligned} & A\mathbf{x} = \lambda\mathbf{x} \\ \iff & A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \\ \iff & (A - \lambda I)\mathbf{x} = \mathbf{0} \end{aligned}$$

This $\mathbf{x}$ is a non trivial solution! This can happen $\iff$ the square matrix $A - \lambda I$ is not invertible $\iff$ $\det(A - \lambda I) = 0$

Recipe
To find the eigenvectors and eigenvalues of A:

- First, find the eigenvalues using λ is an eigenvalue $\iff$ $\det(A - \lambda I) = 0$
- Then, for each eigenvalue λ, find the corresponding eigenvectors by solving $(A - \lambda I)\mathbf{x} = \mathbf{0}$. So you need to find the null space $\mathrm{Nul}(A - \lambda I)$.

4.1 The characteristic polynomial

Example 7. Find the eigenvectors and eigenvalues of

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

Solution.

- $A - \lambda I = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}$
- $\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1$

  $$= \lambda^2 - 6\lambda + 8 = 0 \implies \lambda_1 = 2, \lambda_2 = 4$$

  This is the **characteristic polynomial** of A. Its roots are the eigenvalues of A.

- Next, find the eigenvectors with eigenvalue $\lambda_1 = 2$:

$$A - \lambda_1 I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \left(A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \right)$$

  Solutions to $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$ have basis $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. So: $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda_1 = 2$. All other eigenvectors with eigenvalue $\lambda = 2$ are multiples of $\mathbf{x}_1$. Span $\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ is the **eigenspace** for the eigenvalue $\lambda = 2$.

- Find the eigenvectors with eigenvalue $\lambda_2 = 4$:

$$A - \lambda_2 I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \left(A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \right)$$

  Solutions to $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0}$ have basis $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. So: $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda_2 = 4$. The eigenspace of $\lambda = 4$ is Span $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

4.2 Triangular matrices

Example 8. Find the eigenvectors and eigenvalues of

$$A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 6 & 10 \\ 0 & 0 & 2 \end{bmatrix}$$

Solution.

- The characteristic polynomial is:

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 2 & 3 \\ 0 & 6-\lambda & 10 \\ 0 & 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(6-\lambda)(2-\lambda)$$

- A has eigenvalues $2, 3, 6$.

  $\boxed{\text{The eigenvalues of a triangular matrix are its diagonal entries.}}$

- $\lambda_1 = 2$:

$$(A - \lambda_1 I)\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 10 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_1 = \begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix}$$

- $\lambda_2 = 3$:

$$(A - \lambda_2 I)\mathbf{x} = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 3 & 10 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

- $\lambda_3 = 6$:

$$(A - \lambda_3 I)\mathbf{x} = \begin{bmatrix} -3 & 2 & 3 \\ 0 & 0 & 10 \\ 0 & 0 & -4 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

- Each of those matrices had a one-dimensional null space. So our eigenvectors are not unique. They are unique up to scaling.
- In summary, $A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 6 & 10 \\ 0 & 0 & 2 \end{bmatrix}$ has eigenvalues $2, 3, 6$ with corresponding eigenvectors

$$\begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2/3 \\ 1 \\ 0 \end{bmatrix}.$$

  These three vectors are independent. By the next result, this is always so.

4.3 Independent eigenvectors

Theorem 1. *If $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are eigenvectors of A corresponding to different eigenvalues, then they are independent.*

Proof. Suppose, for contradiction, that $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are dependent. By kicking out some vectors, we may assume that there is (up to multiples) only one linear relation: $c_1\mathbf{x}_1 + \ldots + c_m\mathbf{x}_m = \mathbf{0}$. In other words, the matrix with columns $\mathbf{x}_1, \ldots, \mathbf{x}_m$ has one-dimensional null space. Now multiply this relation with A:

$$A(c_1\mathbf{x}_1 + \ldots + c_m\mathbf{x}_m) = c_1\lambda_1\mathbf{x}_1 + \ldots + c_m\lambda_m\mathbf{x}_m = \mathbf{0}$$

This is a second independent relation! In other words, two independent vectors living in a one-dimensional vector space. Contradiction. □

5 Relations between eigenvalues

5.1 Product of Eigenvalues

If A is $n \times n$ get in principle n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. How are these eigenvalues related?

Theorem 2. *The product of eigenvalues $\lambda_1\lambda_2 \ldots \lambda_n$ is equal to the determinant of A.*

Proof. The characteristic polynomial $p(\lambda) = \det(A - \lambda I)$ has constant term $\det(A)$. On the other hand $p(\lambda)$ factors, because the roots are the eigenvalues we get $p(\lambda) = (\pm 1)(\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n)$, which has constant term $\lambda_1\lambda_2 \ldots \lambda_n$. □

Example 9. Let $A = \begin{bmatrix} \lambda_1 & b \\ 0 & \lambda_2 \end{bmatrix}$. Then the eigenvalues are λ_1, λ_2 and $\det(A) = \lambda_1\lambda_2$.

5.2 Sum of Eigenvalues

What other relations are there between the eigenvalues?

Definition 10. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$ be $n \times n$. Then the **TRACE** of A is the sum of the diagonal entries: $\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$.

Theorem 3. *Let A be $n \times n$. Then the trace of A is the **sum** of eigenvalues:*

$$\mathrm{Tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Example 11. Let $A = \begin{bmatrix} \lambda_1 & b \\ 0 & \lambda_2 \end{bmatrix}$. What are the eigenvalues and what is $\mathrm{Tr}(A)$?

Solution. The eigenvalues are λ_1, λ_2 and $\mathrm{Tr}(A) = \lambda_1 + \lambda_2$.

5.3 The Characteristic Polynomial for 2×2

2×2 matrices are easy.

Theorem 4. *Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then the characteristic polynomial is*

$$p(\lambda) = \lambda^2 - (a+d)\lambda + ad - bc = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

Example 12. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. What are the eigenvalues and what is the characteristic polynomial?

Solution. $\mathrm{Tr}(A) = 6$, $\det(A) = 8$, so $p(\lambda) = \lambda^2 - 6\lambda + 8$. Also in terms of eigenvalues $\mathrm{Tr}(A) = \lambda_1 + \lambda_2$ and $\det(A) = \lambda_1\lambda_2$. So $\lambda_1 = 2, \lambda_2 = 4$

6 Practice problems

Example 13. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$.

Example 14. What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ -1 & 1 & 3 & 0 \\ 0 & 1 & 2 & 4 \end{bmatrix}$. No calculations!

Math 415 - Lecture 31

Markov matrices and Google

Monday November 9th 2015

Textbook reading: Chapter 5.3

Suggested practice exercises: Chapter 5.3: 8, 9, 12, 14, 10.

Khan Academy video: Finding Eigenvectors and Eigenspaces example

Strang lecture: Lecture 21: Eigenvalues and eigenvectors Lecture 24: Markov Matrices and Fourier Series.

1 Review

1.1 Properties of eigenvectors and eigenvalues

- If $A\mathbf{x} = \lambda\mathbf{x}$ then $\mathbf{x}$ is an **eigenvector** of A with **eigenvalue** λ.All eigenvectors (plus $\mathbf{0}$) with eigenvalue λ form **eigenspace** of λ.
- λ is an eigenvalue of $A \iff \underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0$.Why? Because $A\mathbf{x} = \lambda\mathbf{x} \iff (A - \lambda I)\mathbf{x} = \mathbf{0}$.By the way: this means that the eigenspace of λ is just $\mathrm{Nul}(A - \lambda I)$.
- E.g. if $A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 6 & 10 \\ 0 & 0 & 2 \end{bmatrix}$ then $\det(A - \lambda I) = (3 - \lambda)(6 - \lambda)(2 - \lambda)$.

If $A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 6 & 10 \\ 0 & 0 & 2 \end{bmatrix}$ then the eigenvalues are $2, 3, 6$ with corresponding eigenvectors

$$\begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2/3 \\ 1 \\ 0 \end{bmatrix}.$$

These three vectors are independent. By the next result, this is always so.

1.2 Independent eigenvectors

Theorem 1. *If $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are eigenvectors of A corresponding to different eigenvalues, then they are independent.*

Proof. Suppose, for contradiction, that $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are dependent. By kicking out some vectors, we may assume that there is (up to multiples) only one linear relation: $c_1\mathbf{x}_1 + \ldots + c_m\mathbf{x}_m = \mathbf{0}$. In other words, the matrix with columns $\mathbf{x}_1, \ldots, \mathbf{x}_m$ has one-dimensional null space. Now multiply this relation with A:

$$A(c_1\mathbf{x}_1 + \ldots + c_m\mathbf{x}_m) = c_1\lambda_1\mathbf{x}_1 + \ldots + c_m\lambda_m\mathbf{x}_m = \mathbf{0}$$

This is a second independent relation! In other words, two independent vectors living in a one-dimensional vector space. Contradiction. □

2 Relations between eigenvalues

2.1 Product of Eigenvalues

If A is $n \times n$ get in principle n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. How are these eigenvalues related?

Theorem 2. *The product of eigenvalues $\lambda_1\lambda_2\ldots\lambda_n$ is equal to the determinant of A.*

Proof. The characteristic polynomial $p(\lambda) = \det(A - \lambda I)$ has constant term $\det(A)$. On the other hand $p(\lambda)$ factors, because the roots are the eigenvalues we get $p(\lambda) = (\pm 1)(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n)$, which has constant term $\lambda_1\lambda_2\ldots\lambda_n$. □

Example 1. Let $A = \begin{bmatrix} \lambda_1 & b \\ 0 & \lambda_2 \end{bmatrix}$. Then the eigenvalues are λ_1, λ_2 and $\det(A) = \lambda_1\lambda_2$.

2.2 Sum of Eigenvalues

What other relations are there between the eigenvalues?

Definition 2. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$ be $n \times n$. Then the **TRACE** of A is the sum of the diagonal entries: $\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$.

Theorem 3. *Let A be $n \times n$. Then the trace of A is the **sum** of eigenvalues:*

$$\mathrm{Tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Example 3. Let $A = \begin{bmatrix} \lambda_1 & b \\ 0 & \lambda_2 \end{bmatrix}$. What are the eigenvalues and what is $\mathrm{Tr}(A)$?

Solution. The eigenvalues are λ_1, λ_2 and $\mathrm{Tr}(A) = \lambda_1 + \lambda_2$.

2.3 The Characteristic Polynomial for 2×2

2×2 matrices are easy.

Theorem 4. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *Then the characteristic polynomial is*

$$p(\lambda) = \lambda^2 - (a+d)\lambda + ad - bc = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

Example 4. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. What are the eigenvalues and what is the characteristic polynomial?

Solution. $\mathrm{Tr}(A) = 6$, $\det(A) = 8$, so $p(\lambda) = \lambda^2 - 6\lambda + 8$. Also in terms of eigenvalues $\mathrm{Tr}(A) = \lambda_1 + \lambda_2$ and $\det(A) = \lambda_1\lambda_2$. So $\lambda_1 = 2, \lambda_2 = 4$

3 Practice problems

Example 5. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$.

Example 6. What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ -1 & 1 & 3 & 0 \\ 0 & 1 & 2 & 4 \end{bmatrix}$. No calculations!

Example 7. Find the eigenvalues of A as well as a basis for the corresponding eigenspaces, where

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

Solution.

- The characteristic polynomial is:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 2-\lambda & 0 & 0 \\ -1 & 3-\lambda & 1 \\ -1 & 1 & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)\begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)[(3-\lambda)^2 - 1] \\ &= (2-\lambda)(\lambda-2)(\lambda-4) \end{aligned}$$

- A has eigenvalues $2, 2, 4$ $\left(A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix}\right)$ Since $\lambda = 2$ is a double root, it has **(algebraic) multiplicity** 2.

- $\lambda_1 = 2$:
$$(A - \lambda_1 I)\mathbf{x} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$$
Two independent solutions: $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ In other words, the eigenspace for $\lambda_1 = 2$ is Span $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}$.

- $\lambda_2 = 4$: $\left(A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix} \right)$
$$(A - \lambda_2 I)\mathbf{x} = \begin{bmatrix} -2 & 0 & 0 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

- In summary, A has eigenvalues 2 and 4:
  - eigenspace for $\lambda_1 = 2$ has basis $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$,
  - eigenspace for $\lambda_2 = 4$ has basis $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

4 Markov matrices

Definition 8. An $n \times n$ matrices A is **Markov matrix** if has non negative entries, and the entries in each column add to 1.

Theorem 5. *Let A be a Markov matrix. Then*

(i) 1 is an eigenvalue of A and any other eigenvalue λ satisfies $|\lambda| \leq 1$.

(ii) If A has only positive entries, then any other eigenvalue satisfies $|\lambda| < 1$.

Example 9. Let A be
$$\begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix}.$$
Is A a Markov matrix?

Theorem 6. *Let A be an $n \times n$-Markov matrix with only positive entries and let $\boldsymbol{v} \in \mathbb{R}^n$. Then*
$$\boldsymbol{v}_\infty := \lim_{k \to \infty} A^k \boldsymbol{v} \text{ exists,}$$
and $A\boldsymbol{v}_\infty = \boldsymbol{v}_\infty$. In this case $\boldsymbol{v}_\infty$ is often called the ***steady state****.*

Proof. If $\mathbf{x}$ is any vector and $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}$ is an eigenbasis for a Markov matrix ($A\mathbf{v_1} = 1\mathbf{v_1}$):

$$\mathbf{x} = c_1\mathbf{v_1} + \cdots + c_n\mathbf{v_n},$$

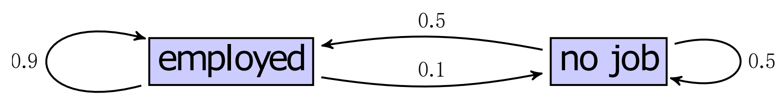
then

$$A^k\mathbf{x} = c_1\lambda_1^k\mathbf{v_1} + \cdots + c_n\lambda_n^k\mathbf{v_n} \to c_1\mathbf{v_1},$$

if the eigenspace of $\lambda = 1$ is 1-dimensional.

□

Example 10. Consider a fixed population of people with or without job. Suppose that each year, 50% of those unemployed find a job while 10% of those employed lose their job.What is the unemployment rate in the long term equilibrium?



Solution. x_t: proportion of population employed at time t (in years) y_t: proportion of population unemployed at time t

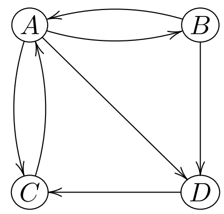
$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 0.9x_t + 0.5y_t \\ 0.1x_t + 0.5y_t \end{bmatrix} = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

The matrix $\begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}$ is a **Markov matrix**. Its columns add to 1 and it has no negative entries. $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$ is an equilibrium if $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix} = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$.In other words, $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$ is an eigenvector with eigenvalue 1.Eigenspace of $\lambda = 1$: $\text{Nul}\left(\begin{bmatrix} -0.1 & 0.5 \\ 0.1 & -0.5 \end{bmatrix}\right) = \text{Span}\left\{\begin{bmatrix} 5 \\ 1 \end{bmatrix}\right\}$Since $x_\infty + y_\infty = 1$, we conclude that $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix} = \begin{bmatrix} 5/6 \\ 1/6 \end{bmatrix}$. Hence, the unemployment rate in the long term equilibrium is 1/6

5 Page rank (or: the 25000000000 $ eigenvector)

Google's success is based on an algorithm to rank webpages, the **Page rank**, named after Google founder Larry Page. The idea is to determine how likely it is that a web user randomly gets to a given webpage. The webpages are ranked by these probabilities.
Suppose the internet consisted of the only four webpages A, B, C, D linked as

in the following graph.



Imagine a surfer following these links at random. For the probability $PR_n(A)$ that she is at A (after n steps), we need to think about how she could have reached A. We add:

- the probability that she was at B (at exactly one step before), and left for A,(that's $PR_{n-1}(B) \cdot \frac{1}{2}$)
- the probability that she was at C, and left for A,
- the probability that she was at D, and left for A.

Hence: $PR_n(A) = PR_{n-1}(B) \cdot \frac{1}{2} + PR_{n-1}(C) \cdot \frac{1}{1} + PR_{n-1}(D) \cdot \frac{0}{1}$.

Encode the probabilties at step n in a state vector with four entries. $\begin{bmatrix} PR_n(A) \\ PR_n(B) \\ PR_n(C) \\ PR_n(D) \end{bmatrix} =$
$\begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} PR_{n-1}(A) \\ PR_{n-1}(B) \\ PR_{n-1}(C) \\ PR_{n-1}(D) \end{bmatrix}$

Definition 11. The **PageRank vector** is the long-term equilibrium. It is an eigenvector of the Markov matrix with eigenvalue 1.

Let's call the Markov matrix with the probabilities T:

- $T - 1I = \begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$

  $\implies$ eigenspace of $\lambda = 1$ is spanned by $\begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix}$.

- Now we need to make the entries add up to 1.

$$\begin{bmatrix} PR(A) \\ PR(B) \\ PR(C) \\ PR(D) \end{bmatrix} = \frac{3}{16} \begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}.$$

  This is the PageRank vector.

- The corresponding ranking of the webpages is A, C, D, B.

Remark. In practical situations the system might be too large for finding the eigenvalues by row operations.

- Google reports having met 60 trillion webpages. Google's search index is over $100,000,000$ gigabytes. Number of Google's servers is secret: about $2,500,000$ More than 1,000,000,000 websites (i.e. hostnames; about 75% not active)
- Thus we have a gigantic but very sparse matrix.

An alternative to row operations is the **power method** (see Theorem 2):

Power method
If T is an (acyclic and irreducible) Markov matrix, then for any $\mathbf{v}_0$ the vectors $T^n\mathbf{v}_0$ converge to an eigenvector with eigenvalue 1

Here:

$$T = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Start with an arbitrary state vector, hit it with powers of T.

$$\left(\begin{bmatrix} PR(A) \\ PR(B) \\ PR(C) \\ PR(D) \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}\right), T\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.083 \\ 0.333 \\ 0.208 \end{bmatrix}$$

. Note that the ranking of the webpages is already A, C, D, B if we stop here.

$$T\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.083 \\ 0.333 \\ 0.208 \end{bmatrix}, \qquad T^2\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.333 \\ 0.167 \end{bmatrix}, \qquad T^3\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.396 \\ 0.125 \\ 0.292 \\ 0.188 \end{bmatrix}$$

Remark.

- If all entries of T are positive (no zero entries!), then the power method is guaranteed to work.
- In the context of PageRank, we can make sure that this is the case by replacing T with

$$(1-p)\cdot\begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} + p\cdot\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Just to make sure: still a Markov matrix, now with positive entries Google used to use $p = 0.15$.

6 Practice problems

Problem 12. *Can you see why* 1 *is an eigenvalue for every Markov matrix?*

Problem 13 (just for fun)**.** *The real web contains pages which have no outgoing links. In that case, our random surfer would get "stuck" (the transition matrix is not a Markov matrix). Do you have an idea how to deal with this issue?*

Math 415 - Lecture 32

Complex numbers and eigenvectors

Wednesday November 11th 2015

Textbook reading: first part of Chapter 5.5

Suggested practice exercises: 5.5 1, 2, 3

Khan Academy video: Complex Numbers (part 1)

Strang lecture: Lecture 21: Eigenvalues and eigenvectors

SOME DATES.

* Friday November 13th: No class.
* Next week there will be discussion sections, but no quiz. Prepare for the midterm!
* Thursday November 19th, 7-8:15PM: Midterm 3.
* Friday November 20th: No class.
* November 23-27th Thanksgiving break: no class.
* Wednesday December 9th: last day of class
* Thursday December 17th: Final Exam.

1 Review

1.1 Properties of eigenvectors and eigenvalues

- If $A\mathbf{x} = \lambda\mathbf{x}$ then $\mathbf{x}$ is an **eigenvector** of A with **eigenvalue** λ.
- λ is an eigenvalue of $A \iff \underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0$.

Definition 1. An $n \times n$ matrix A is a **Markov matrix** if has non negative entries, and the entries in each column add to 1.

Theorem 1. *Let A be a Markov matrix. Then*

(i) 1 is an eigenvalue of A and any other eigenvalue λ satisfies $|\lambda| \leq 1$.

(ii) If A has only positive entries, then any other eigenvalue satisfies $|\lambda| < 1$.

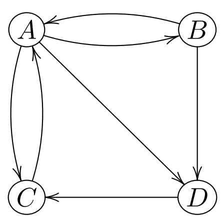
Theorem 2. *Let A be an $n \times n$-Markov matrix with only positive entries and let $\boldsymbol{v} \in \mathbb{R}^n$. Then*

$$\boldsymbol{v}_\infty := \lim_{k\to\infty} A^k \boldsymbol{v} \text{ exists,}$$

and $A\boldsymbol{v}_\infty = \boldsymbol{v}_\infty$. In this case $\boldsymbol{v}_\infty$ is often called the ***steady state****.*

2 Page rank (or: the 25000000000 \$ eigenvector)

Google's success is based on an algorithm to rank webpages, the **Page rank**, named after Google founder Larry Page. The idea is to determine how likely it is that a web user randomly gets to a given webpage. The webpages are ranked by these probabilities.
Suppose the internet consisted of the only four webpages A, B, C, D linked as in the following graph.



Imagine a surfer following these links at random. For the probability $PR_n(A)$ that she is at A (after n steps), we need to think about how she could have reached A. We add:

- the probability that she was at B (at exactly one step before), and left for A,(that's $PR_{n-1}(B) \cdot \frac{1}{2}$)
- the probability that she was at C, and left for A,
- the probability that she was at D, and left for A.

Hence: $PR_n(A) = PR_{n-1}(B) \cdot \frac{1}{2} + PR_{n-1}(C) \cdot \frac{1}{1} + PR_{n-1}(D) \cdot \frac{0}{1}$.

Encode the probabilties at step n in a state vector with four entries. $\begin{bmatrix} PR_n(A) \\ PR_n(B) \\ PR_n(C) \\ PR_n(D) \end{bmatrix} =$

$$\begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} PR_{n-1}(A) \\ PR_{n-1}(B) \\ PR_{n-1}(C) \\ PR_{n-1}(D) \end{bmatrix}$$

Definition 2. The **PageRank vector** is the long-term equilibrium. It is an eigenvector of the Markov matrix with eigenvalue 1.

Let's call the Markov matrix with the probabilities T:

- $T - 1I = \begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$

  $\implies$ eigenspace of $\lambda = 1$ is spanned by $\begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix}$.

- Now we need to make the entries add up to 1.

$$\begin{bmatrix} PR(A) \\ PR(B) \\ PR(C) \\ PR(D) \end{bmatrix} = \frac{3}{16} \begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}.$$

  This is the PageRank vector.

- The corresponding ranking of the webpages is A, C, D, B.

Remark. In practical situations the system might be too large for finding the eigenvalues by row operations.

- Google reports having met 60 trillion webpages. Google's search index is over $100,000,000$ gigabytes. Number of Google's servers is secret: about $2,500,000$ More than 1,000,000,000 websites (i.e. hostnames; about 75% not active)
- Thus we have a gigantic but very sparse matrix.

An alternative to row operations is the **power method** (see Theorem 2):

Power method
If T is an (acyclic and irreducible) Markov matrix, then for any $\mathbf{v}_0$ the vectors $T^n\mathbf{v}_0$ converge to an eigenvector with eigenvalue 1

Here:

$$T = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Start with an arbitrary state vector, hit it with powers of T.

$$\left(\begin{bmatrix} PR(A) \\ PR(B) \\ PR(C) \\ PR(D) \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix} \right), T \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.083 \\ 0.333 \\ 0.208 \end{bmatrix}$$

. Note that the ranking of the webpages is already A, C, D, B if we stop here.

$$T \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.083 \\ 0.333 \\ 0.208 \end{bmatrix}, \quad T^2 \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.333 \\ 0.167 \end{bmatrix}, \quad T^3 \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.396 \\ 0.125 \\ 0.292 \\ 0.188 \end{bmatrix}$$

Remark.

- If all entries of T are positive (no zero entries!), then the power method is guaranteed to work.
- In the context of PageRank, we can make sure that this is the case by replacing T with

$$(1-p) \cdot \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} + p \cdot \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

  Just to make sure: still a Markov matrix, now with positive entries Google used to use $p = 0.15$.

3 Eigenbasis?

3.1 Number of (independent) eigenvectors

An $n \times n$ matrix A has up to n different eigenvalues. Namely, the roots of degree n characteristic polynomial $\det(A - \lambda I)$.

- For each eigenvalue λ, A has at least one eigenvector. That is because $\mathrm{Nul}(A - \lambda I)$ has dimension at least one.
- If λ has multiplicity m, then A has up to m (independent) eigenvectors for λ.

Ideally, we would like to find a total of n (independent) eigenvectors for A. This would give an **EIGENBASIS**. Why can there be no more than n independent eigenvectors?!

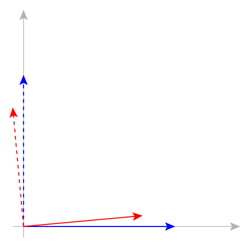
Two sources of trouble: eigenvalues can be

- complex numbers (that is, not enough real roots), or
- repeated roots of the characteristic polynomial.

3.2 Trouble I: complex eigenvalues

Example 3. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Geometrically, what is the trouble?

Solution.



$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

i.e. multiplication by A is a rotation by $90°$ (counter-clockwise).

Which vector is parallel after rotation by $90°$? Trouble.

4 Complex numbers review

Definition. $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$

- $i = \sqrt{-1}$, or $i^2 = -1$.
- Any point in $\mathbb{R}^2$ can be viewed as a complex number: $\binom{x}{y} \leftrightarrow x + iy$

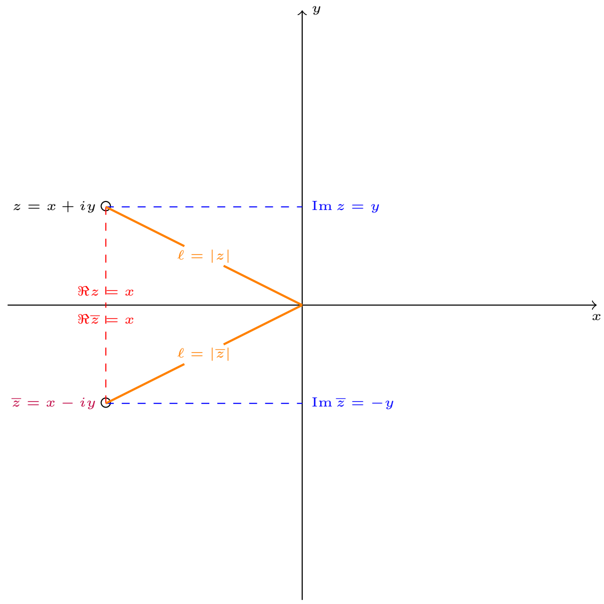
Definition. Let $z = x + iy$ be a complex number

Real part The real part of z, denoted $\Re(z)$ is defined by $\Re(z) = x$.

Imaginary part The imaginary part of z, denoted $\mathrm{Im}(z)$ is defined by $\mathrm{Im}(z) = y$.

Complex conjugate The complex conjugate of z, denoted $\overline{z}$, is defined by $\overline{z} = x - iy$.

Absolute value The absolute value, or magnitude of z, denoted $|z|$ or $\|z\|$, is given by $|z| = \sqrt{x^2 + y^2}$.



Adding complex numbers

Definition. Given $z = x + iy$, $w = u + iv$, we define

$$z + w = (x + u) + i(y + v)$$

Remark. This corresponds exactly to addition of vectors in $\mathbb{R}^2$.

Multiplying complex numbers

Definition. Given $z = x + iy$, $w = u + iv$, we define

$$\begin{aligned} zw &= (x + iy)(u + iv) \\ &= xu + x(iv) + (iy)u + (iy)(iv) \\ &= (xu - yv) + i(xv + yu) \end{aligned}$$

Absolute value and complex conjugate

Remark.

- $\overline{\overline{z}} = z$
- $|z|^2 = z\overline{z}$
- $|z| = |\overline{z}|$

Proof.

$$\begin{aligned} z\overline{z} &= (x + iy)(x - iy) \\ &= x^2 - x(iy) + (iy)x - (iy)(iy) \\ &= x^2 + y^2 \end{aligned}$$

□

4.1 Complex Linear Algebra

Until now we took as our scalars the real numbers. In particular we used the vector space $\mathbb{R}^n$ of column vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

If c is a real number (a scalar) we defined

$$c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}.$$

Now we want to use **COMPLEX** scalars. We need a new context to make sense of this.

Definition. $\mathbb{C}^n$ is the (complex) vector space of *complex* column vectors $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$, where $z_1, z_2, \ldots . z_n$ are complex numbers.

- Now multiplication by a complex scalar makes sense.
- We can define subspaces, Span, independence, basis, dimension for $\mathbb{C}^n$ in the usual way.
- We can multiply complex vectors by complex matrices. Column space and Null space still make sense.
- The only difference is the dot product, you need to use the complex conjugate to get a good notion of length. (Later more.)

5 Back to eigenvectors

Example 4. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Now, we can use complex numbers!

Solution (continued).

- $\det(A - \lambda I) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1$ So the eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$.

- $\lambda_1 = i$: $\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$ Let us check $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ i \end{bmatrix} = i \begin{bmatrix} i \\ 1 \end{bmatrix}$

- $\lambda_2 = -i$: $\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix}$

Summary: We had $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

- Eigenvalues: $i, -i$ These are conjugates!
- Eigenvectors: $\mathbf{x}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix}$ These are also conjugates!

Theorem 3. *If A is a matrix with real entries and λ is a* ***complex eigenvalue****, then $\bar{\lambda}$ is also a complex eigenvalue. Furthermore, if $\mathbf{x}$ is an eigenvector with eigenvalue λ, then $\bar{\mathbf{x}}$ is an eigenvector with eigenvalue $\bar{\lambda}$.*

Remark. Note that we are using vectors in $\mathbb{C}^2$, instead of vectors in $\mathbb{R}^2$. Works pretty much the same!

5.1 Trouble II: generalized eigenvectors

Example 5. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. What is the trouble?

Solution.

- $\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$ So: $\lambda = 1$ is the only eigenvalue (it has multiplicity 2).
- $\lambda = 1$: $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ So the eigenspace is Span $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. Only dimension 1!
- Trouble: We can not find an **Eigenbasis** for this matrix. This kind of problem cannot really be fixed. We have to lower our expectations and look for generalized eigenvectors. These are solutions to $(A - \lambda I)^2 \mathbf{x} = \mathbf{0}, (A - \lambda I)^3 \mathbf{x} = \mathbf{0}, \ldots$

6 Practice problems

Example 6. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -5 & 0 \\ 1 & 8 & 1 \end{bmatrix}$.

Math 415 - Lecture 33

Diagonalization

Monday November 16th 2015

Textbook reading: Chapter 5.2

Suggested practice exercises: Chapter 5.2: 1, 2, 3, 4, 5, 6, 7, 15, 16, 17, 18, 19, 20, 25, 26, 29, 30, 31, 32, 33

Strang lecture: Lecture 22: Diagonalization and powers of A

1 Review

- **Eigenvector** equation: $A\mathbf{x} = \lambda\mathbf{x} \iff (A - \lambda I)\mathbf{x} = \mathbf{0}$ λ is an **eigenvalue** of $A \iff \underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0.$
- An $n \times n$ matrix A has up to n eigenvectors for λ.
  - The **eigenspace** of λ is $\text{Nul}(A - \lambda I)$.
  - If λ has **multiplicity** m, then A has up to m (independent) eigenvectors for λ. At least one eigenvector is guaranteed (because $\det(A - \lambda I) = 0$).
  - An **Eigenbasis** for an $n \times n$ matrix A is a basis $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}$ of $\mathbb{R}^n$ so that each $\mathbf{v_i}$ is also an eigenvector: $A\mathbf{v_i} = \lambda_i \mathbf{v_i}$.
- Test yourself! What are the eigenvectors and eigenvalues? Is there an eigen basis?
  - $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\lambda = 1, 1$ (i.e. multiplicity 2), eigenspace is $\mathbb{R}^2$. Any basis is eigen basis.
  - $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\lambda = 0, 0$, eigenspace is $\mathbb{R}^2$. Again any basis is an eigenbasis.

  These are trivial cases. Is there always an eigenbasis?

Example 1. To solve $A\mathbf{x} = \mathbf{b}$ we use row operations. If we want to find eigenvectors, $A\mathbf{x} = \lambda\mathbf{x}$, can we also use row operations? Try $A = \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}$.

- What is the echelon form U of A?
- What are the characteristic polynomials $\det(A - \lambda I)$ and $\det(U - \lambda I)$? Roots?
- Do A and U have the same eigenvalues? Eigenvectors?

Solution.

- If $A = \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}$ then $U = \begin{bmatrix} 1 & 1 \\ 0 & -4 \end{bmatrix}$.
- Then $\det(A - \lambda I) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A) = = \lambda^2 - 0\lambda + (-4) = \lambda^2 - 4 = (\lambda - 2)(\lambda + 2)$, and $\det(U - \lambda I) = \lambda^2 - \mathrm{Tr}(U)\lambda + \det(U) = = \lambda^2 - 3\lambda + (-4) = (\lambda - 1)(\lambda + 4)$.
- So the eigenvalues of A and U are **DIFFERENT!**. Can check that eigenvectors are also different.

Upshot: **Don't use row operations to deal with eigenvalues and eigenvectors!** (Can use row operations to calculate determinants, though.)

Example 2. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. What is the trouble?

Solution.

- $\det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2$ So: $\lambda = 1$ is the only eigenvalue (it has multiplicity 2).
- $\lambda = 1 : \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ So the eigenspace is Span $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. Only dimension 1!
- Trouble: We can not find an **Eigenbasis** for this matrix. This kind of problem cannot really be fixed. We have to lower our expectations and look for generalized eigenvectors. These are solutions to $(A - \lambda I)^2 \mathbf{x} = \mathbf{0}, (A - \lambda I)^3 \mathbf{x} = \mathbf{0}, \ldots$

2 Diagonalization

2.1 Powers of diagonal matrices

Diagonal matrices are very easy to work with.

Example 3. Let $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$. What is A^2? What is A^{100}?

Solution. $A^2 = \begin{bmatrix} 2^2 & 0 & 0 \\ 0 & 3^2 & 0 \\ 0 & 0 & 4^2 \end{bmatrix}$ and $A^{100} = \begin{bmatrix} 2^{100} & 0 & 0 \\ 0 & 3^{100} & 0 \\ 0 & 0 & 4^{100} \end{bmatrix}$.

2.2 Powers of generic matrices

Example 4. If $A = \begin{bmatrix} 6 & -1 \\ 2 & 3 \end{bmatrix}$, then $A^{100} =?$

Solution.

- characteristic polynomial: $\begin{vmatrix} 6-\lambda & -1 \\ 2 & 3-\lambda \end{vmatrix} = \cdots = (\lambda - 4)(\lambda - 5)$
  - $\lambda_1 = 4 : \begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix} \mathbf{v} = \mathbf{0} \implies$ eigenvector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
  - $\lambda_2 = 5 : \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \mathbf{v} = \mathbf{0} \implies$ eigenvector $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- Key observation: $A^{100}\mathbf{v}_1 = \lambda_1^{100}\mathbf{v}_1$ and $A^{100}\mathbf{v}_2 = \lambda_2^{100}\mathbf{v}_2$. For A^{100}, we need $A^{100}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $A^{100}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = -\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies$

$$A^{100}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = A^{100}\left(-\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = -4^{100}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2 \cdot 5^{100}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\implies A^{100} = \begin{bmatrix} 2 \cdot 5^{100} - 4^{100} & * \\ 2 \cdot 5^{100} - 2 \cdot 4^{100} & * \end{bmatrix}$$

- We find the second column of A^{100} likewise. Left as exercise!

The key idea of previous example is to work with respect to an *Eigenbasis*, a basis given by eigenvectors.

- Put the eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as columns into a matrix P.

$$A\mathbf{x}_i = \lambda\mathbf{x}_i \implies A\begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\mathbf{x}_1 & \cdots & \lambda_n\mathbf{x}_n \\ | & & | \end{bmatrix}$$
$$= \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- In summary $AP = PD$. Such a diagonalization is possible if and only if A has enough eigenvectors.

So we are going to use eigenvalues and eigenvectors for A to factor A and A^{100} in a useful way. This is called *diagonalization*.

Definition. A square matrix A is said to be **diagonalizable** if there is a invertible matrix P and a diagonal matrix D such that

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

Theorem 1. *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.*

We can express the relation between A and D in terms of change of base matrices.

$$\begin{tikzcd}[row sep=huge, column sep=huge]
\substack{\text{coords for } \mathbf{x} \\ \text{in standard basis}} \arrow[r, "A"] \arrow[d, shift right, "P^{-1}"'] & \substack{\text{coords for } A\mathbf{x} \\ \text{in standard basis}} \arrow[d, shift right, "P^{-1}"'] \\
\substack{\text{coords for } \mathbf{x} \\ \text{in eigen- basis}} \arrow[r, "D"] \arrow[u, shift right, "P"'] & \substack{\text{coords for } A\mathbf{x} \\ \text{in eigen- basis}} \arrow[u, shift right, "P"']
\end{tikzcd}$$

$D = P^{-1}AP$, $A = PDP^{-1}$

P changes from eigenbasis coordinates to standard coordinates, and P^{-1} goes the other way! Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^n$ and $\mathcal{B}$ the basis of eigenvectors of A, then

$$P = I_{\mathcal{E},\mathcal{B}} \text{ and } P^{-1} = I_{\mathcal{B},\mathcal{E}}.$$

3 Application: Large powers

If A has an eigenbasis, then we can raise it to large powers easily!

Theorem 5. *If $A = PDP^{-1}$, where D is a diagonal matrix, then for any m,*

$$A^m = PD^mP^{-1}$$

Proof.

$$\begin{aligned}
A &= PDP^{-1} \\
A^m &= (PDP^{-1})^m \\
&= \underbrace{(PDP^{-1}) \cdot (PDP^{-1}) \cdots (PDP^{-1})}_{m \text{ times}} \\
&= (PD)(P^{-1} \cdot P)(DP^{-1}) \cdots (PDP^{-1}) \\
&= PD \cdot DP^{-1} \cdots PDP^{-1} \\
&= PD \cdot D \cdots D \cdot P^{-1} \\
&= PD^mP^{-1}
\end{aligned}$$

Only the outside P and P^{-1} remain! □

Finding D^m is easy!

$$D^m = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}^m = \begin{bmatrix} (\lambda_1)^m & & \\ & \ddots & \\ & & (\lambda_n)^m \end{bmatrix}$$

Why?

Example 6. Let $A = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 6 \\ 0 & 0 & 2 \end{bmatrix}$. A has eigenvectors and eigenvalues

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{with} \quad \lambda_1 = \frac{1}{2}$$

$$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{with} \quad \lambda_2 = 1$$

$$\mathbf{x}_3 = \begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix} \quad \text{with} \quad \lambda_3 = 2$$

Find A^{100}. Hint: Write $A = PDP^{-1}$.

Solution. Eigenvectors of A form an Eigenbasis! So we can write $A = PDP^{-1}$:

Matrix of eigenvectors $P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix}$

Find P^{-1}

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 6 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R2 \rightsquigarrow R2 - R1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 6 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R2 \rightsquigarrow R2 - 6R3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & -6 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix}$$

Eigenvectors of A form a Eigenbasis! So we can write $A = PDP^{-1}$:

Matrix of eigenvalues: $D = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Finally, write $A = PDP^{-1}$**:**

$$\underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 6 \\ 0 & 0 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix}}_{P^{-1}}$$

Take power

$$
\begin{aligned}
\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 6 \\ 0 & 0 & 2 \end{bmatrix}^{100} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}^{100} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2^{100}} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^{100} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2^{100}} & 0 & 0 \\ (\frac{1}{2^{100}} - 1) & 1 & (6 \cdot 2^{100} - 6) \\ 0 & 0 & 2^{100} \end{bmatrix}
\end{aligned}
$$

Math 415 - Lecture 34

Discrete dynamical systems, Spectral Theorem

Wednesday November 18th 2015

Textbook reading: Chapter 5.3, Chapter 5.6 p. 297-298

Suggested practice exercises: Chapter 5.3, 2, 3, 4, 7, 8, 9, 10, 12, 14

Strang lecture: Lecture 25: Symmetric Matrices and Positive Definiteness

1 Review

Diagonalization
Suppose that A is an $n \times n$ and has independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then A can be **diagonalized** as $A = PDP^{-1}$.

- the columns of P are eigenvectors
- the diagonal matrix D has the eigenvalues on the diagonal

Such a diagonalization is possible if and only if A has an eigenbasis.

Calculating Powers
If $A = PDP^{-1}$ for some diagonal matrix D, then $A^n = PD^nP^{-1}$ for every n. This is helpful, because calculating powers of diagonal matrices is very easy!

2 Application: Discrete Dynamical Systems

Suppose you want to describe the evolution of some part of the world. Describe the **state** of your part of the world at time $t = 0$ by a vector $\mathbf{x}_{t=0}$, the **state-vector.** For a simple system the state vector $\mathbf{x}_t$ might have 2 components, for a complicated system there might be thousands of components. Then you want to know what the state $\mathbf{x}_t$ at arbitrary time t is. How? Assume

$$\mathbf{x}_{t+1} = A\mathbf{x}_t.$$

In other words, time evolution by one time step is given by matrix multiplication by some matrix A. If we start with $\mathbf{x}_0$, we get $\mathbf{x}_1 = A\mathbf{x}_0$, $\mathbf{x}_2 = A\mathbf{x_1} = A(A\mathbf{x}_0) = A^2\mathbf{x}_0$, and more generally, the state of the system at arbitrary time $t = k$ is

$$\mathbf{x}_k = A^k\mathbf{x}_0.$$

So to solve our system we need to be able to calculate high powers of the matrix A. Use eigenbasis of A for this.

2.1 Golden ratio and Fibonacci numbers

Example 1. 'A certain man put a pair of rabbits in a place surrounded by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair from which the second month on becomes productive?' (Liber abbaci, chapter 12, p. 283-4)

Solution. Idea: use discrete dynamical system to produce the Fibonacci numbers.

- $F_{n+1} = F_n + F_{n-1} \implies \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}$
- Hence $\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \qquad \left(\begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$
- But we know how to compute $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n$ or $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$!

Solution.

- The characteristic polynomial of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is $\lambda^2 - \lambda - 1$
- The eigenvalues are $\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.618$ (the golden mean!) and $\lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.618$
- Corresponding eigenvectors: $\mathbf{v}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$
- Write $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. $(c_1 = \frac{1}{\sqrt{5}}, c_2 = -\frac{1}{\sqrt{5}})$
- $\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1^n c_1 \mathbf{v}_1 + \lambda_2^n c_2 \mathbf{v}_2$
- Hence, $F_n = \lambda_1^n c_1 + \lambda_2^n c_2 = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right]$. That is **Binet's formula**.
- but $|\lambda_2| < 1$, so $F_n \approx \lambda_1^n c_1 = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n$. In fact, $F_n = \text{round}\left(\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n \right)$.

Fibonacci numbers: $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$ Did you notice: $\frac{13}{8} = 1.625$, $\frac{21}{13} = 1.615$, $\frac{34}{21} = 1.619, \ldots$ The **golden ratio** $\varphi = 1.618...$ Where's that from? We just showed that $F_n = \text{round}\left(\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n\right)$. Therefore

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \left(\frac{1+\sqrt{5}}{2}\right).$$

Definition 2. Let A be a $n \times n$-matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The discrete dynamical system $\mathbf{x}_{t+1} = A\mathbf{x}_t$ is

- **stable** if all eigenvalues satisfy $|\lambda_i| < 1$,
- **neutrally stable** if some $|\lambda_i| = 1$ and all the other $|\lambda_i| < 1$,
- **unstable** if at least one eigenvalue has $|\lambda_i| > 1$.

Example 3.

1. The discrete dynamical system used to construct the Fibonacci numbers is unstable.
2. If A is a Markov matrix with positive entries, then $\mathbf{x}_{t+1} = A\mathbf{x}_t$ is neutrally stable.
3. If $A = \begin{bmatrix} 0 & 4 \\ 0 & \frac{1}{2} \end{bmatrix}$, is $\mathbf{x}_{t+1} = A\mathbf{x}_t$ stable?

3 Spectral Theorem

- Not every matrix A has a basis of eigenvectors ☹
- Special case:

Definition. A is symmetric if $A = A^T$

Theorem 1. *If A is symmetric, then it has an orthonormal basis of eigenvectors and all eigenvalues are real!* ☺

If Q is the matrix of eigenvectors, then Q is orthogonal. So, $Q^{-1} = Q^T$. Thus,

$$A = QDQ^{-1} = QDQ^T$$

and

$$D = Q^{-1}AQ = Q^TAQ$$

Remark.

- The converse is also true: If A has an orthogonal basis of eigenvectors, then A is symmetric! Why?
- It is important that if A is symmetric the eigenvalues are always real. No complex eigenvalues!

Example 4. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. Write A as QDQ^T.

Solution. We've seen this matrix before!

Find eigenvalues: We have seen that A has eigenvalues 2 and 4.

Find eigenbasis corresponding to eigenvalues:

$\lambda_1 = 2$**:** We have seen $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Normalized, we get $\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

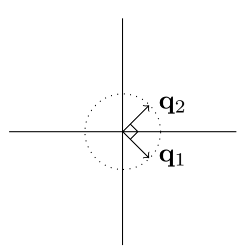
$\lambda_2 = 4$**:** We have seen $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Normalized, we get $\mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

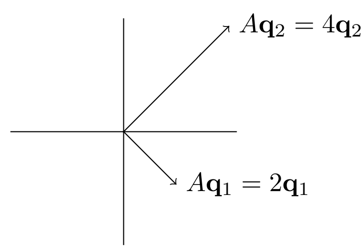
Write D**:** $D = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

Write Q**:** $Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

Get $A = QDQ^T$**:** $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

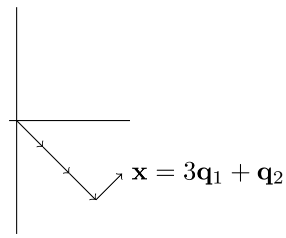
What does A do to the eigenvectors?

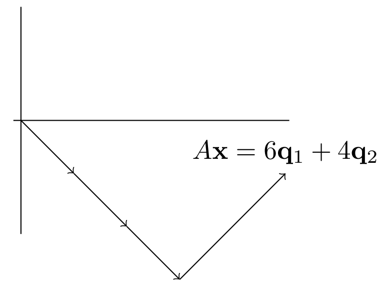




What happens to a vector $\mathbf{x}$?

Suppose $\mathbf{x} = 3\mathbf{q}_1 + \mathbf{q}_2$:





Why are symmetric matrices special? Why does spectral theorem work? If $A = A^T$, and if

$$A\mathbf{x} = \lambda_1 \mathbf{x} \text{ and } A\mathbf{y} = \lambda_2 \mathbf{y}$$

(for $\lambda_1 \neq \lambda_2$), then $\mathbf{x}$ and $\mathbf{y}$ **must** be orthogonal! Why?

Let's show $\mathbf{x} \cdot \mathbf{y} = 0$:

$$\begin{aligned}
\lambda_1(\mathbf{x} \cdot \mathbf{y}) &= (\lambda_1 \mathbf{x}) \cdot \mathbf{y} \\
&= (A\mathbf{x}) \cdot \mathbf{y} \\
&= (A\mathbf{x})^T \mathbf{y} \\
&= \mathbf{x}^T A^T \mathbf{y} \\
&= \mathbf{x}^T A \mathbf{y} \quad \longleftarrow \text{ because } A \text{ is symmetric!} \\
&= \mathbf{x} \cdot (A\mathbf{y}) \\
&= \lambda_2(\mathbf{x} \cdot \mathbf{y})
\end{aligned}$$

Since $\lambda_1 \neq \lambda_2$, must have $\mathbf{x} \cdot \mathbf{y} = 0$! By a similar argument you can show that the eigenvalues of a symmetric matrix **must** be real.

Example 5. Let $A = \begin{bmatrix} 3 & -1 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$. Then

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

Let $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ -2 \end{bmatrix}$. Find $A^3\mathbf{x}$.

Math 415 - Lecture 35

Quadratic forms

Monday November 30th 2015

Textbook reading: Chapter 6.2

Suggested practice exercises: Chapter 6.2, # 1, 2, 4, 5

Strang lecture: Lecture 27: Positive definite matrices and minima

1 Review

Spectral theorem:

- A is a symmetric matrix if $A^T = A$. e.g. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 5 \end{bmatrix}$
- Any $n \times n$ symmetric matrix A has n real eigenvalues and an orthonormal eigenbasis $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$.
- So, we can write
$$A = QDQ^T$$
where
$$D = \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{\text{matrix of eigenvalues}} \text{ and } Q = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix}}_{\text{matrix of eigenvectors}}$$

2 Quadratic forms

2.1 Quadratic forms

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with critical point at $\mathbf{0}$. This means that all partial derivatives at $\mathbf{0}$ vanish. Is $\mathbf{0}$ a max, min, or neither? How to tell?

- Look at the **quadratic** part of f!

Definition. A **quadratic form** $f : \mathbb{R}^n \to \mathbb{R}$ is a polynomial (in n variables) with every term degree two.

e.g., for $n = 2$

$$f(x, y) = 3x^2 + 4xy - 5y^2$$

Example 1. Let

$$f(x, y) = \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Expand $f(x, y)$ as a polynomial in x and y. The dot denotes the dot product!

Solution. Expanding we get

$$\begin{aligned} f(x, y) &= \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 3x + 2y \\ 2x - 5y \end{bmatrix} \\ &= 3x^2 + 4xy - 5y^2 \end{aligned}$$

This is the quadratic function from before!

Theorem 1. *Any quadratic form $f(x_1, x_2, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ can be written*

$$f(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} = \mathbf{x}^T A\mathbf{x}$$

for a symmetric matrix A.

We see symmetric matrices show up "in the wild!"

Example 2. Write $f(x, y, z) = 5x^2 + 7y^2 + 3z^2 + 2xy - 2yz$ as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ where A is symmetric.

Solution. A 3×3 symmetric matrix has the form $A = \begin{bmatrix} a & d & f \\ d & b & e \\ f & e & c \end{bmatrix}$. In general,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} a & d & f \\ d & b & e \\ f & e & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax^2 + by^2 + cz^2 + (2d)xy + (2e)yz + (2f)xz$$

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

2.2 Principal axes for a quadratic form

Intermezzo: From Eigenbasis to Standard Basis and back.

- A symmetric, so $A = QDQ^T$.

- If $x \in \mathbb{R}^n$ and $x_Q = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ is the coordinate vector of x in the Q basis, then $x = c_1 q_1 + c_2 q_2 + \cdots + c_n q_n = Qx_Q$.
- This means that to find the Q coordinate vector for x, multiply by $Q^{-1} = Q^T$:

$$\boxed{x_Q = Q^T x}$$

There is always a "nicest possible" coordinate system for each quadratic form. Just use an eigenbasis of A.

Theorem 2. *Let A be a symmetric matrix, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ an orthonormal basis of eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$. Write*

$$\mathbf{x} = c_1 \mathbf{v}_n + \cdots + c_n \mathbf{v}_n \quad {\scriptstyle How?}$$

Then,

$$q(\mathbf{x}) := \mathbf{x}^T A \mathbf{x} = \lambda_1 (c_1)^2 + \cdots + \lambda_n (c_n)^2$$

Proof. A is symmetric. So write $A = QDQ^T$. Let's find $\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T QDQ^T \mathbf{x}$. We know $Q^T \mathbf{x}$ writes $\mathbf{x}$ in $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ coordinates. So

$$Q^T \mathbf{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

D is the matrix of eigenvalues. So,

$$D \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \lambda_1 c_1 \\ \vdots \\ \lambda_n c_n \end{bmatrix}$$

Since $\mathbf{x}^T Q = (Q^T x)^T$, we have $x^T Q = \begin{bmatrix} c_1 & \ldots & c_n \end{bmatrix}$. Thus,

$$\mathbf{x}^T A \mathbf{x} = \begin{bmatrix} c_1 & \ldots & c_n \end{bmatrix} \begin{bmatrix} \lambda_1 c_1 \\ \ldots \\ \lambda_n c_n \end{bmatrix} = \lambda_1 (c_1)^2 + \cdots + \lambda_n (c_n)^2$$

□

Example 3. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Let $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$.

- Find the eigenvalues λ_1, λ_2 and **orthonormal** eigenbasis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for A.
- Compute $q(\mathbf{x})$ using the formula $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$.
- Compute $q(\mathbf{x})$ using the theorem ($q(\mathbf{x}) = \lambda_1 (c_1)^2 + \lambda_2 (c_2)^2$.)

Are the answers the same? **This is a silly Example. To calculate $q(x)$ you never would go through the eigenvalues.**

Solution. Eigenvalues: Sum $\lambda_1 + \lambda_2 = \text{Tr}(A) = 2$ Product $\lambda_1\lambda_2 = \det(A) = -3$. So, $\lambda_1 = 3$, $\lambda_2 = -1$.

Eigenbasis: $\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$

Compute using formula:

$$\begin{aligned}
q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\
&= \begin{bmatrix}2 & 0\end{bmatrix}\begin{bmatrix}1 & 2\\2 & 1\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix} \\
&= \begin{bmatrix}2 & 0\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix} \\
&= 4
\end{aligned}$$

Using theorem: $\mathbf{x} = \begin{bmatrix}2\\0\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}1\\-1\end{bmatrix} = \sqrt{2}\mathbf{v}_1 + \sqrt{2}\mathbf{v}_2$. So,

$$\begin{aligned}
q(\mathbf{x}) &= \lambda_1(c_1)^2 + \lambda_2(c_2)^2 \\
&= 3(\sqrt{2})^2 + (-1)(\sqrt{2})^2 \\
&= 4
\end{aligned}$$

Get same answer!

We have $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ and

$$q(\mathbf{x}) = \lambda_1(c_1)^2 + \cdots + \lambda_n(c_n)^2$$

- So up to coordinate change, $\mathbf{q}$ is a weighted sum of squares.
- The eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called principal axes

Definition 4. Let A be a symmetric $n \times n$. We say A is **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all non zero $\mathbf{x} \in \mathbb{R}^n$.

Theorem 3. *Let A be a symmetric $n \times n$ and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Then*

1. *If all $\lambda_i > 0$, then A is positive definite,*
2. *If all $\lambda_i < 0$, then $\mathbf{x}^T A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$*
3. *If some $\lambda_i > 0$, some $\lambda_j < 0$, $\mathbf{x}^T A\mathbf{x}$ will have both positive and negative values.*

2.3 Completing the squares

Basic Question. Let A be a symmetric matrix, and $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. Is $q(\mathbf{x})$ always ≥ 0? Or always ≤ 0? How to decide? Write $q(\mathbf{x})$ as a sum of squares!

Example 5. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, so that $q(\mathbf{x}) = x^2 + 4xy + y^2$. Write $q(\mathbf{x})$ as a sum of squares. Is $q(\mathbf{x})$ always positive?

Solution. * $q(\mathbf{x}) = x^2 + 4xy + y^2 = (x + 2y)^2 - 3y^2$.

* Sometimes you get something positive, sometimes something negative.

There are many ways of writing $q(\mathbf{x})$ as a sum of squares. Today we are using eigenvalues to do this.

Math 415 - Lecture 36

Minima, maxima and saddle points, Constrained Optimization

Wednesday December 2nd 2015

Textbook reading: Chapter 6.1

Suggested practice exercises: Chapter 6.1, # 1, 16

Strang lecture: Lecture 27: Positive definite matrices and minima

1 Review

Spectral theorem:

- A is a **symmetric** matrix if $A^T = A$.
- Any $n \times n$ symmetric matrix A has n **real eigenvalues** and an **orthonormal eigenbasis** $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$.
- So, we can write $A = QDQ^T$ where

$$D = \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{\text{matrix of eigenvalues}} \text{ and } Q = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix}}_{\text{matrix of eigenvectors}}$$

- A is called **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- a function of the form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called a **quadratic form**.
- Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Then
  1. If all $\lambda_i > 0$, then A is positive definite,
  2. If all $\lambda_i < 0$, then $\mathbf{x}^T A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$
  3. If some $\lambda_i > 0$, some $\lambda_j < 0$, $\mathbf{x}^T A\mathbf{x}$ will have both positive and negative values.

2 2nd derivative test

2.1 2nd derivative test

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, the **Hessian** matrix of f

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1}(\mathbf{0}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{0}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{0}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{0}) & \frac{\partial^2 f}{\partial^2 x_2}(\mathbf{0}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{0}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{0}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{0}) & \cdots & \frac{\partial^2 f}{\partial^2 x_n}(\mathbf{0}) \end{bmatrix}$$

Idea. If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\mathbf{0}$ is a critical point, then $f(\mathbf{x}) \approx f(\mathbf{0}) + \frac{1}{2}\mathbf{x} \cdot H\mathbf{x}$.

- H is always symmetric
- We're approximating $f(\mathbf{x})$ by $f(\mathbf{0})$ plus a quadratic function, $\frac{1}{2}\mathbf{x} \cdot H\mathbf{x}$!
- We understand $q(\mathbf{x}) = \frac{1}{2}\mathbf{x} \cdot H\mathbf{x} \implies$ we understand if $\mathbf{0}$ is a max, min or neither for f!
- Turns out: $q(\mathbf{x})$ is determined by eigenvectors and eigenvalues of H!

Theorem 1 (2nd derivative test)**.** *If $f : \mathbb{R}^n \to \mathbb{R}$ has a critical point at $\mathbf{0}$, then*

1. *If all eigenvalues of H are positive, then $\mathbf{0}$ is a local min. H is positive-definite, graph is a bowl.*
2. *If all eigenvalues of H are negative, then $\mathbf{0}$ is a local max. H is negative-definite, graph is a dome.*
3. *If one eigenvalue of H is positive and one is negative, then $\mathbf{0}$ is neither a max nor a min. H is indefinite, graph is a saddle*
4. *Otherwise (e.g. all eigenvalues positive or zero), no information!*

Example 2. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ has a critical point at $\mathbf{0}$ and has Hessian $H = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Does f have local max, min or neither at $\mathbf{0}$?

(An example of such a function is $f(x, y) = \frac{1}{2}x^2 + 2xy + \frac{1}{2}y^2$).

Solution. We showed that H has eigenvalues 3 and -1. So f has a saddle point at $\mathbf{0}$.

Example 3. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ has a critical point at $\mathbf{0}$ and has Hessian $H = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$. Does f have local max, min or neither at $\mathbf{0}$?

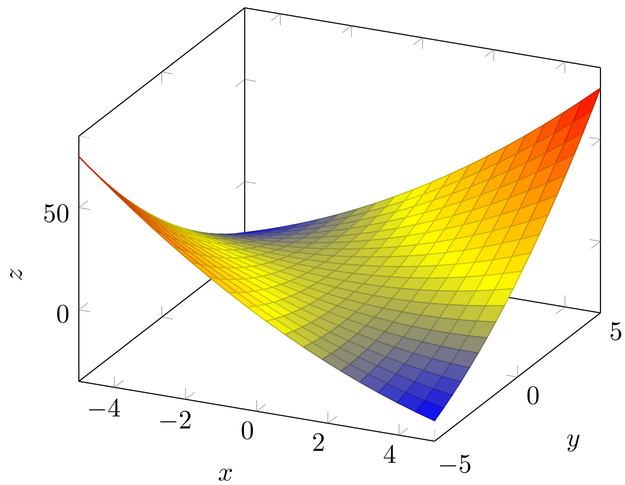


Figure 1: Graph of the function $f(x, y) = \frac{1}{2}x^2 + 2xy + \frac{1}{2}y^2$

Solution. Eigenvalues: Sum $\lambda_1 + \lambda_2 = \mathrm{Tr}(H) = 4$ Product $\lambda_1\lambda_2 = \det(H) = 2$. So, λ_1, λ_2 must be positive! (positive product $\implies$ both positive or both negative. positive sum $\implies$ both positive.)

2nd derivative test says: $f(\mathbf{0})$ is local min.

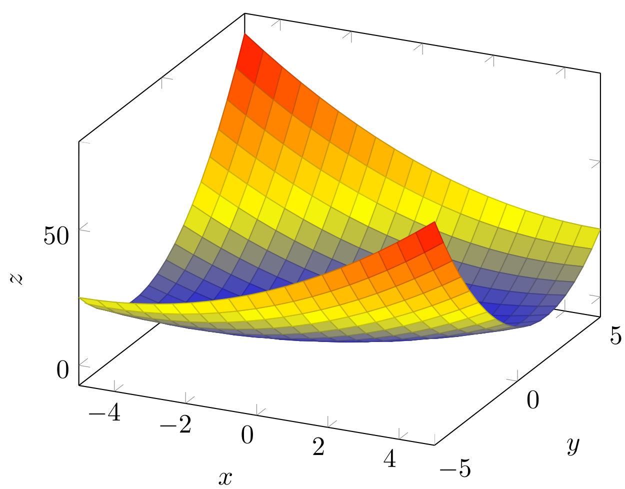


Figure 2: Graph of the function $f(x, y) = \frac{1}{2}x^2 - xy + \frac{3}{2}y^2$

3 Constrained optimization

Problem: Given a quadratic from $q(x)$, find the maximum or minimum value $q(x)$ for $\mathbf{x}$ in some specified set. Typically, the problem can be arranged such that $\mathbf{x}$ varies over the set of vectors with $\mathbf{x}^T\mathbf{x} = 1$.

Example 4. Let $A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Find the maximum and minimum values of $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

Solution. The quadratic form is $q(x_1, x_2, x_3) = 9x_1^2 + 4x_2^2 + 3x_3^2$. We are interested in the maximal value for q when (x_1, x_2, x_3) is such that $x_1^2+x_2^2+x_3^2 = 1$. Now we can give an upper bound for q: we obviously have

$$q(\mathbf{x}) \leq 9x_1^2 + 9x_2^2 + 9x_3^2 = 9$$

Solution. So $q(\mathbf{x})$ can not be bigger than 9, for any $\mathbf{x}$. Can we get $q(\mathbf{x}) = 9$ for some $\mathbf{x}$? Obviously for $\mathbf{x} = (1, 0, 0)$ we achieve the upper bound, so 9 is the maximum value for q (under this constraint.) What is a lower bound? For which $\mathbf{x}$ is the lower bound achieved?

What if A is not diagonal?

Theorem 2. *Let A be a symmetric matrix and let λ_m be the least eigenvalue and λ_M be the greatest eigenvalue of A. Then*

$$\lambda_m = \min\{\mathbf{x}^T A\mathbf{x} \ : \ \mathbf{x}^T\mathbf{x} = 1\},$$

moreover if $\mathbf{u}_m$ is a unit eigenvector corresponding to λ_m, then $\mathbf{u}_m^T A\mathbf{u}_m = \lambda_m$. In addition,

$$\lambda_M = \max\{\mathbf{x}^T A\mathbf{x} \ : \ \mathbf{x}^T\mathbf{x} = 1\},$$

moreover if $\mathbf{u}_M$ is a unit eigenvector corresponding to λ_M, then $\mathbf{u}_M^T A\mathbf{u}_M = \lambda_M$.

Proof. We know by the spectral theorem that $A = QDQ^T$, and so we can write $q(\mathbf{x}) = \mathbf{x}^T QDQ^T\mathbf{x} = u^T Du = \lambda_M u_1^2 + \cdots + \lambda_m u_m^2$, where $u = Q^T x$. As before we see that the largest eignvalue λ_M is the upper bound for q, achieved for $u = (1, 0, \ldots, 0)$ or $\mathbf{x} = Qu$, the normalized eigenvector corresponding to λ_M. Same argument for λ_m. □

Example 5. Let $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$. Find the maximum and minimum values of $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

Solution. We first find eigenvectors and eigenvalues for A. Let us ask Wolfram Alpha: $\det(A)$, Eigenvalues and eigenvectors . So $\lambda = 6, 3, 1$, with eigenvectors

$$v_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

Then $q(x)$ has maximum value 6, and $q(v_1) = v_1^T A v_1 = 6\|\mathbf{v_1}\| = 6$. The minimal value is 1 and $q(v_3) = v_3^T A v_3 = 1\|\mathbf{v_3}\| = 1$.

Math 415 - Lecture 37

Singular Value Decomposition

Friday December 4th 2015

Textbook reading: Chapter 6.3

Suggested practice exercises: Chapter 6.3, # 1, 2, 3, 5, 8, 9, 15

Strang lecture: Lecture 29: Singular Value Decomposition

1 Review

- Spectral theorem: If A is an $n \times n$ symmetric matrix, then it has an orthonormal basis of eigenvectors $\mathbf{v}_1 \dots \mathbf{v}_n$, and all eigenvalues are real.
- We can write

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_n \\ | & & | \end{bmatrix}}_{\text{matrix of eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \lambda_n \end{bmatrix}}_{\text{matrix of eigenvalues}} \begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}$$

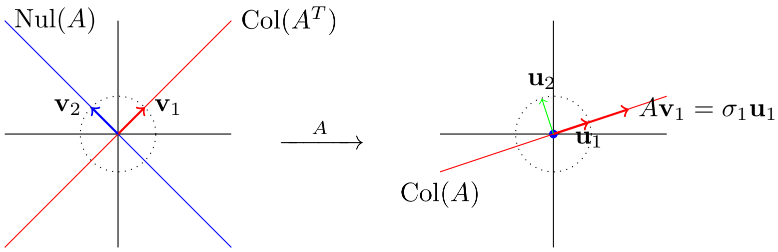
- Today: There is a similar decomposition for any $m \times n$ matrix A.
  - Doesn't even have to be square!
  - The price we pay: different bases on the left and right sides.

2 Singular Value Decomposition

2.1 Goals

Starting with an $m \times n$ matrix A we want to

- Describe the *geometry* of the corresponding map $\mathbb{R}^n \to \mathbb{R}^m$,
- Find a way to *approximate* A by simpler matrices, that are easier/cheaper to calculate with.



How?Remember: for each A we get 4 subspaces

- Input space $\mathbb{R}^n$ contains row space Col(A^T) and Null space Nul(A). Dimensions are r and $n-r$.
- Output space $\mathbb{R}^m$ contains columns space Col(A) and left null space Nul(A^T). Dimensions are r and $m-r$.

2.2 Idea

Idea:
Find an orthonormal basis in each of the input space subspaces, and watch what happens to these basis vectors if we multiply by A.

Choose a basis $\mathbf{v_1}, \ldots, \mathbf{v_r}$ of the row space $Col(A^T)$, and a basis $\mathbf{v_{r+1}}, \ldots, \mathbf{v_n}$ of the null space Nul(A). Then think of what happens when we apply A to each of the basis vectors:

- $A\mathbf{v_{r+1}} = 0 = \cdots = A\mathbf{v_n}$, since each vector belongs to Nul(A).
- The other vectors $A\mathbf{v_i}$, $i = 1, 2, \ldots, r$, will all be nonzero, in fact will be give a basis of Col(A)!
- Rescale these basis vectors to get unit vectors $\mathbf{u_1}$,$\mathbf{u_2}, \ldots, \mathbf{u_r}$. By a miracle they turn out to be orthogonal, if we choose the $\mathbf{v}_1, \mathbf{v}_2, \ldots$ in the right way.
- We get $A\mathbf{v_i} = \sigma_i \mathbf{u_i}$ for $i = 1, 2, \ldots, r$. The stretch factors $\sigma_i > 0$, $i = 1, 2, \ldots, r$ are called the *Singular Values* of A
- Extend the $\mathbf{u_i}$ basis of Col(A) to a basis $\mathbf{u_{r+1}}, \ldots, \mathbf{u_m}$ of the output space.

2.3 What is SVD?

Motto
In Linear Algebra everything is a matrix factorization.

The complicated story with orthonormal basis and singular values for A gives a factorization, called Singular Value Decomposition:

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_m \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

- $A = U\Sigma V^T$. This is just $Av_i = \sigma_i u_i$ rearranged in matrix form.
- U, V are orthogonal. We need to choose the input basis v_i carefully in order for the output basis u_i to be orthonormal.
- Columns of U are an orthonormal basis for $\mathbb{R}^m$. U is $m \times m$.
- Rows of V are an orthonormal basis for $\mathbb{R}^n$. V is $n \times n$.
- Σ is rectangular $m \times n$ and diagonal, the r non zero diagonal entries are called **singular values**, they are positive.

2.4 How to Compute SVD

Here is a recipe for computing SVD:

Compute A^TA. This is a symmetric matrix!! (Why?)

Make V:
- Find orthonormal eigenvectors $\mathbf{v_1}, \dots, \mathbf{v_n}$ of A^TA. (Why can we do this?)
- Magic: The eigenvalues are always positive or zero! $\lambda_1 \geq \dots \geq \lambda_r > 0$, $\lambda_{r+1} = 0 = \dots = \lambda_n$.
- Order $\mathbf{v_1}, \dots, \mathbf{v_n}$ according to the size of their eigenvalues.
- Put $\mathbf{v}_1, \dots, \mathbf{v}_n$ into matrix V,

Make Σ: Set $\sigma_i = \sqrt{\lambda_i}$ for $i = 1 \dots r$ and $\sigma_{r+1} = \dots = \sigma_n = 0$. Put these into diagonal of **rectangular** $m \times n$ matrix Σ.

Make U:
- Set $\mathbf{u_1} = \frac{1}{\sigma_1} A\mathbf{v_1}$, …, $\mathbf{u_r} = \frac{1}{\sigma_r} A\mathbf{v_r}$.
- Magic: The $\mathbf{u}_1, \dots, \mathbf{u}_r$ are orthogonal
- Extend $\mathbf{u_1}, \dots, \mathbf{u_r}$ to an orthonormal basis $\mathbf{u_1}, \dots, \mathbf{u_m}$ for $\mathbb{R}^m$.
- Put $\mathbf{u_1}, \dots, \mathbf{u_m}$ into matrix U.

Now you have $A = U\Sigma V^T$!

Example 1. Compute the SVD of

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

Solution. Compute $A^TA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Make V: Basis of eigenvectors for A^TA: $\mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

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

Make Σ: Eigenvalues are 1 and 1. So, $\sigma_1 = \sigma_2 = \sqrt{1} = 1$.

$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Make U: $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v_1} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

$U = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

We get

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

Note: $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has no real eigenvalues. It's not diagonalizable with real matrices! But, it has an SVD! Wolfram Alpha

Example 2. Compute the SVD of

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

Solution. **Compute** $A^TA = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

Make V: Basis of eigenvectors for A^TA:

$$\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

with eigenvalues $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$.

Make Σ: Eigenvalues were $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$, so $\sigma_1 = \sqrt{3}, \sigma_2 = 1$.

Make U: $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ $\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 =$ $\frac{1}{1} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

Final result:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Notice how A behaves in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$:
$A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1$:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \sqrt{3} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$A\mathbf{v}_2 = \sigma_2 \mathbf{u}_2$:

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

$A\mathbf{v}_3 = 0$.

A matrix might not be diagonalizable:

- If A is rectangular, it does not even have eigenvalues.

But A will always have an SVD! This comes at a cost:

- The SVD is not unique.
- The singular values σ_i are not eigenvalues.

Note the difference: for $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ the eigenvalues are $\lambda = i, -i$ but the singular values are $\sigma = 1, 1$.

2.5 Approximation

* To calculate matrix product AB we can use the **ROW** times **COLUMN** method: the ij component is the product $R_i B_j$, where R_i is row i of A and B_j is the jth column of B.

* What about **COLUMN** times **ROW**?

*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a+3b & 2a+4b \\ c+3d & 2c+4d \end{bmatrix} =$$

$$= \begin{bmatrix} a \\ c \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} b \\ d \end{bmatrix} \begin{bmatrix} 3 & 4 \end{bmatrix}$$

* This works for any matrix multiplication: AB is a sum of **COLUMN** times **ROW** matrices.

It turns out we can write A as a sum:

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_m \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

$$= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \dots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T$$

(Sanity check: An $m \times 1$ column vector times a $1 \times n$ row vector is an $m \times n$ matrix.)

Idea. We can get a good approximation to A by taking the entries of the sum with the largest singular values! We'll see this when we talk about image compression later.

Example 3. If $\mathbf{u}, \mathbf{v}$ are non-zero, then the matrix $\mathbf{u}\mathbf{v}^T$ has rank 1. Why?

Solution. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$. Then

$$\mathbf{u}\mathbf{v}^T = \mathbf{u} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}^T = \mathbf{u} \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = \begin{bmatrix} v_1 \mathbf{u} & \dots v_n \mathbf{u} \end{bmatrix}.$$

Example 4. Use

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

to write $\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ as a sum of rank 1 matrices.

Solution.

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \sqrt{3} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + 1 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

2.6 SVD and the Four Fundamental Subspaces

The SVD of A gives orthonormal bases for all four fundamental subspaces of A.

Given $\{\mathbf{u_1}, \dots, \mathbf{u_m}\}$ and $\{\mathbf{v_1}, \dots, \mathbf{v_n}\}$,

- $Col(A^T) = \text{Span}\{\mathbf{v_1}, \dots, \mathbf{v_r}\}$
- $Nul(A) = \text{Span}\{\mathbf{v_{r+1}}, \dots, \mathbf{v_n}\}$
- $Col(A) = \text{Span}\{\mathbf{u_1}, \dots, \mathbf{u_r}\}$
- $Nul(A^T) = \text{Span}\{\mathbf{u_{r+1}}, \dots, \mathbf{u_m}\}$

2.7 Practice Questions

Example 5. Suppose A is an invertible square matrix. Find a singular value decomposition of A^{-1}.

Example 6. If A is a square matrix, then $|\det(A)|$ is the product of the singular values of A. Why?

Example 7. Find the singular value decomposition of $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$.

Math 415 - Lecture 38

Applications of SVD

Monday December 7th 2015

Textbook reading: Chapter 6.3

Strang lecture: Lecture 29: Singular Value Decomposition

- Final exam is on Thursday 12/17/2015 8am - 11am. We will announce the room assignment this week.
- Conflict final exam is on Tuesday 12/15/2015 8am-11am. If you send your TA an email, you will receive an email with the location this week.
- Lecture on Wednesday 12/9 will be used for review.
- Please check your scores online. If incorrect, contact TA.

1 Review

Singular Value Decomposition:

$$A = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_m \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

- $A = U\Sigma V^T$, where U $(m \times m)$, V $(n \times n)$ are orthogonal.
- Σ is rectangular $m \times n$ and diagonal, the r non zero diagonal entries are called **singular values**, they are positive.
- Can be rewritten in *column times row form*

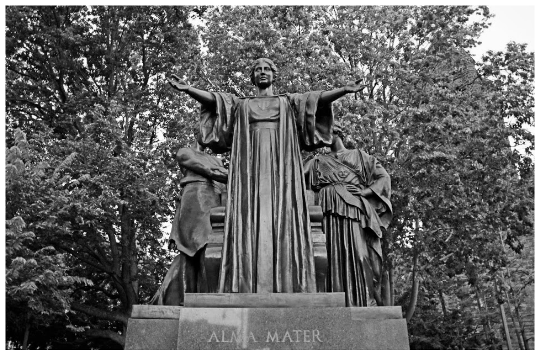
$$A = \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T + \ldots + \mathbf{u}_r\sigma_r\mathbf{v}_r^T$$

- $\mathbf{u}_1, \ldots, \mathbf{u}_m$ orthonormal eigenbasis of AA^T.$\mathbf{v_1}, \ldots, \mathbf{v_n}$ orthonormal eigenbasis for A^TA.

2 Image Compression

2.1 Idea

We have a grayscale picture that is $m \times n$ pixels in size:



Each pixel is a shade of gray from 0 (black) to 255 (white). This gives an $m \times n$ matrix A. Each entry of A is one pixel of the image; that entry is some integer from 0 to 255, giving the brightness of that pixel.

Encoding a large picture takes up a lot of space:

- Say our picture is 625×960 pixels.
- Each pixel has 256 possible values. (This takes up exactly 8 bits, or 1 byte, of memory on a computer).
- The whole picture requires $625 \times 960 = 600000$ bytes, so 600 kB.
- That is not so bad, but color pictures from your personal camera would be about 30MB each without any compression... and that quickly adds up.

Question. Can we do better?

Use singular value decomposition of A:

$$A = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_{625} \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \\ 0 & \sigma_2 & \\ & & \ddots \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_{960}^T & - \end{bmatrix}$$

Recall we can rewrite this as

$$A = \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T + \ldots + \mathbf{u}_r\sigma_r\mathbf{v}_r^T$$

For most pictures $r = 625$, the maximal rank of A.

Idea. Throw away the term $\mathbf{u}_i\sigma_i\mathbf{v}_i^T$ when σ_i is small. If $k \leq r$, define

$$A_k = \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T + \ldots + \mathbf{u}_k\sigma_k\mathbf{v}_k^T$$

The matrix A_k is very close to the matrix A, if $\sigma_{k+1}, \ldots, \sigma_r$ are small.

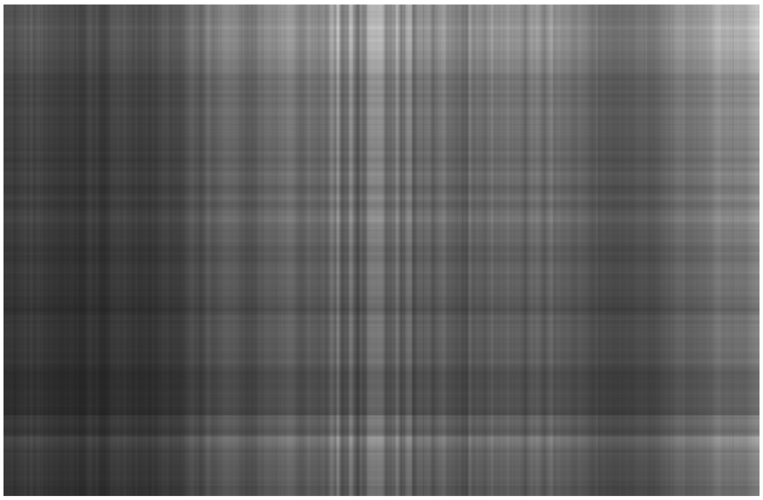
For example, take A_{100}:

A_k is also easier to store:

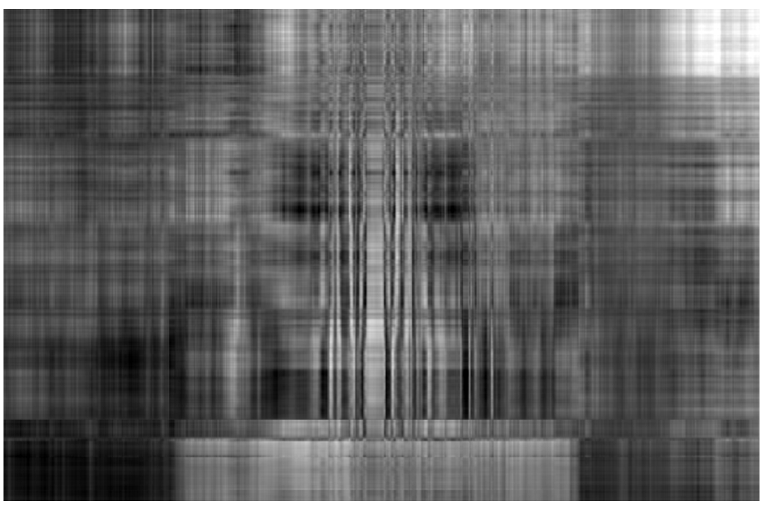
- If $k = 100$, then to store the matrix A_{100} we need the numbers $\sigma_1, \ldots, \sigma_{100}$, the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{100}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_{100}$.
- That's
$$100 + 100(625) + 100(960) = 158600$$
numbers
- Compare to the original matrix which had
$$625 \cdot 960 = 600000$$
numbers.

We reduced the file size by a factor of four!

2.2 Examples



A_1

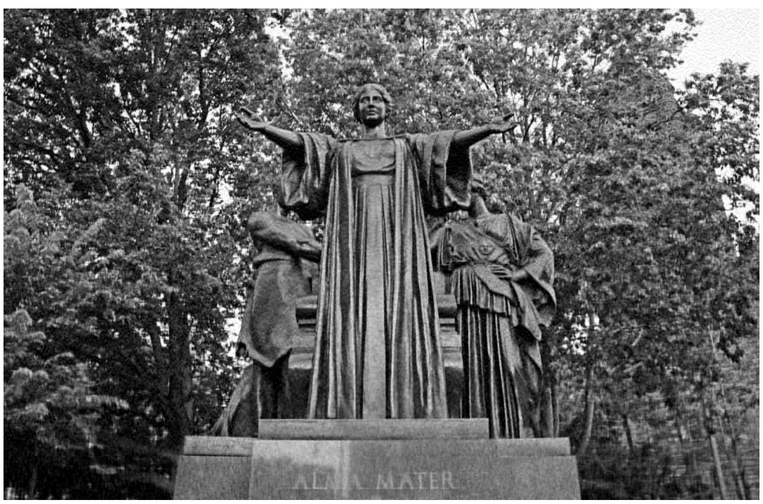


A_5

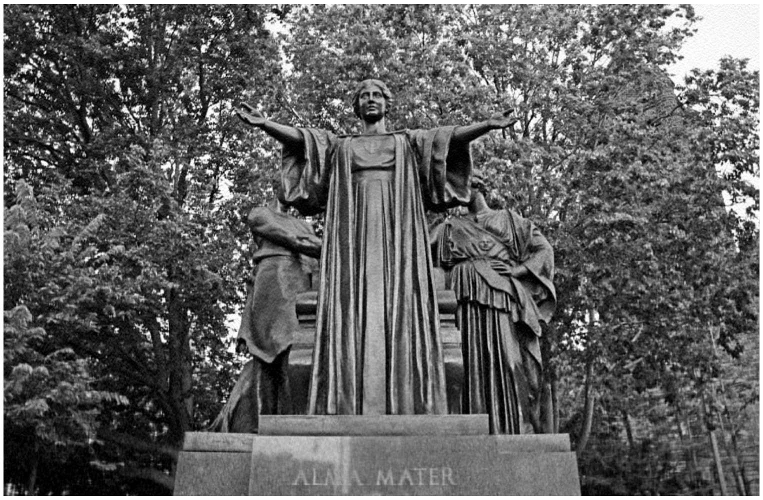
A_{25}

A_{50}

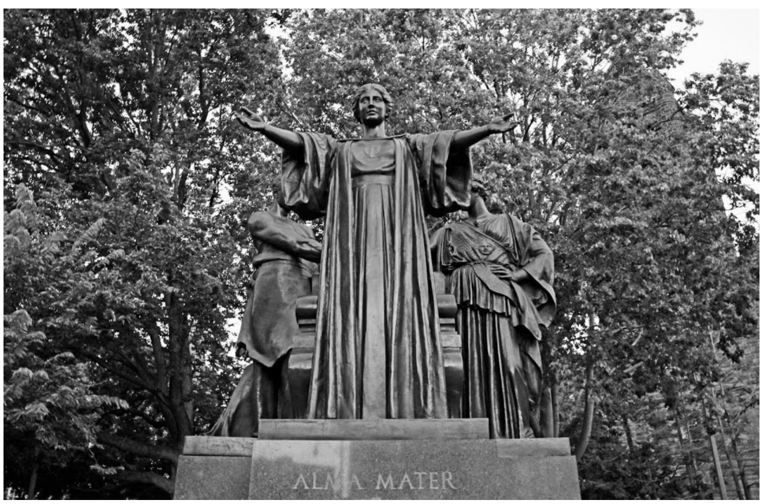
A_{100}



A_{150}



A_{200}



A. Probably rank A is 625, so $A = A_{625}$.

3 Face recognition

3.1 Idea

A set of eigenfaces can be generated by performing a mathematical process called principal component analysis (PCA) on a large set of images depicting different human faces. Informally, eigenfaces can be considered a set of "standardized face ingredients", derived from statistical analysis of many pictures of faces. Any human face can be considered to be a combination of these standard faces. For example, one's face might be composed of the average face plus 10% from eigenface 1, 55% from eigenface 2, and even -3% from eigenface 3.

Remarkably, it does not take many eigenfaces combined together to achieve a fair approximation of most faces. Also, because a person's face is not recorded by a digital photograph, but instead as just a list of values (one value for each eigenface in the database used), much less space is taken for each person's face.

3.2 Algorithm

Part I - Set up

Step 1. Start with a set of d face images - same resolution, say $r \times c$. Each image is treated as one vector by concatenating the columns of pixels in the original image, resulting in vector in $\mathbb{R}^{rc}$. Set $n = rc$. This gives us d vectors $\mathbf{x}_1, \ldots, \mathbf{x}_d$ in $\mathbb{R}^n$. For instance $n = 625 \times 960 = 600000$

Step 2. Calculate average face $\mathbf{a} = \frac{\mathbf{x}_1 + \cdots + \mathbf{x}_d}{d}$. Set $\mathbf{y}_i := \mathbf{x}_i - \mathbf{a}$.Let X be the $n \times d$ matrix whose i-th column is $\mathbf{y}_i$.

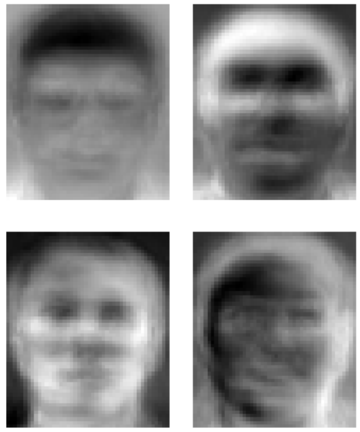
Step 3. Calculate the singular value decomposition of X. Say $X = U\Sigma V^T$. Note that $U = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix}$ is a $n \times n$-matrix, V is $d \times d$-matrix.

Step 4. The vectors $\mathbf{u}_1, \ldots, \mathbf{u_n}$ are called eigenfaces. Forget all eigenfaces, where the corresponding singular value is small. Let's say, we keep $\mathbf{u}_1, \ldots \mathbf{u}_s$ for some $s < n$.

So what are eigenfaces?

We said: 'Eigenfaces can be considered a set of "standardized face ingredients", derived from statistical analysis of many pictures of faces.'

Note that $\mathbf{u}_1, \ldots, \mathbf{u_n}$ are the eigenvalues (principal components/axes) of XX^T. We can think of XX^T as the matrix of all possible combination of the faces. Therefore $\mathbf{u}_1, \ldots, \mathbf{u_n}$ are the principal components of all the possible faces.



Some eigenfaces from AT&T Laboratories Cambridge

Part II - Learning new faces
Suppose we want to add a face $\mathbf{f} \in \mathbb{R}^n$ to the database. Then instead of saving f, we save the following vector in $\mathbb{R}^s$

$$\begin{bmatrix} (\mathbf{f}-\mathbf{a})\cdot\mathbf{u}_1 \\ \vdots \\ (\mathbf{f}-\mathbf{a})\cdot\mathbf{u}_s \end{bmatrix}.$$

We call this vector $\mathbf{w}_f$.

Do you know what the entries represent? The weights/scalars used in the projection of $\mathbf{f}-\mathbf{a}$ onto span$(\mathbf{u}_1,\ldots,\mathbf{u}_s)$! So $\mathbf{f}$ is composed of the average face $\mathbf{a}$ plus scalar $(\mathbf{f}-\mathbf{a})\cdot\mathbf{u}_1$ times eigenface 1 plus scalar $(\mathbf{f}-\mathbf{a})\cdot\mathbf{u}_2$ times eigenface 2, etc.

Part III - Recognizing a face

Suppose we are now given a photo of a person $\mathbf{p} \in \mathbb{R}^n$. First calculate $w_{\mathbf{p}} = \begin{bmatrix} (\mathbf{p}-\mathbf{a})\cdot\mathbf{u}_1 \\ \vdots \end{bmatrix}$. How do we know which person in the database is most likely this person? Easy, simply take the face f in the database such that $\|w_p - w_f\|$ is minimal!

Math 415 - Lecture 39

Review

Wednesday December 6th 2015

Final Information:

- Thursday December 17th, 8:00-11:00AM.
  - 101 Armory: AD3,ADG,ADU,ADW
  - 180 Bevier: ADH,ADP,ADQ,ADX
  - 100 Gregory: ADA,ADB,ADJ,ADK,ADV,ADY
  - 151 Loomis: AD4,AD7,AD8,ADI,ADR
  - 103 Mumford: AD9,ADE,ADF,ADN,ADO
  - 100 MSEB: AD1,AD2,ADS,ADT,ADZ
  - 135 THBH: ADC,ADD,ADL,ADM (THBH is Temple Hoyne Buell Hall)
- Conflict Tuesday, December 15th, 8:00-11:00AM.

Bring university ID, pencils and erasers, there will be a part multiple choice.

1 After Exam 3

After Exam 3

- Diagonalization,
- Discrete Dynamical Systems.
- Spectral Theorem and Quadratic forms: each symmetric matrix A gives a quadratic form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, and conversely. The eigenvalues of A (real!) determine if the quadratic form is always positive.
- Critical points of functions $f\colon \mathbb{R}^n \to \mathbb{R}$ are described by a quadratic form (Hessian) containing the second derivatives of f. Minima, maxima, saddle points. Constrained optimization.

- Singular Value Decomposition of A from spectral theorem for $A^T A$, and AA^T.
- Approximation of a matrix A according to the singular values: image compression.

2 Big Topics

- Solving Systems $A\mathbf{x} = \mathbf{b}$
  - Augmented matrix.
  - Row Operations, Reduced Row echelon form.
  - Pivots, free variables, parametric form of general solution.
  - Inconsistent system, unique solution or infinitely many solutions.

- Vectors and Matrices
  - Linear Combinations
  - Matrix multiplication is linear combination
  - Row/column calculation of matrix multiplication
  - Transpose, symmetric matrices.
  - Elementary row operations and elementary matrices.
  - LU factorization, solving $Ax = b$ by $Lc = b$, $Ux = c$.
  - Inverse of a square matrix, Gauss-Jordan calculation of A^{-1} (Big Augmented Matrix).

- Vector Spaces.
  - Linear combinations.
  - Subspace.
  - Spanning set, independence.
  - Basis and dimension.
  - Coordinates with respect to a basis.

- Linear Transformations
  - Linear transformation determined by basis.

- Coordinate matrix with respect to input/output bases.

- Orthogonality
  - Dot product=inner product.
  - Length of vector.
  - angle between vectors.
  - Orthogonal complement $W^{\perp}$, dimensions add: $\dim(W)+\dim(W^{\perp}) = \dim(\mathbb{R}^n)$.
  - Orthogonal and orthonormal basis.

- Fundamental thm of Linear Algebra.
  - Four fundamental subspaces of A: $\mathrm{Col}(A), \mathrm{Col}(A^T), \mathrm{Nul}(A), \mathrm{Nul}(A^T)$.
  - $\mathrm{Nul}(A)$ and uniquess of solutions of $A\mathbf{x} = \mathbf{b}$.
  - $\mathrm{Col}(A)$ and existence of solutions of $A\mathbf{x} = \mathbf{b}$.
  - 4 subspaces pairwise orthogonal.
  - Dimensions of the subspaces and bases, from echelon form.
  - Networks and fundamental subspaces.

- Projections
  - Projection on a line.
  - Orthogonal basis makes projection easy.
  - Projection matrix.
  - Orthogonal decomposition: x can be written as $x = x_W + x_{W^{\perp}}$ for $x_W \in W$, $x_{W^{\perp}} \in W^{\perp}$.

- Least Squares
  - Approximate solutions of $A\mathbf{x} = \mathbf{b}$: make $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.
  - Least square solution is $\hat{\mathbf{x}}$ satisfying the normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

- The projection of $\mathbf{b}$ on the subspace $\mathrm{Col}(A)$ is $A\hat{\mathbf{x}}$.
- Data Fitting

- Gram-Schmidt
  - From arbitrary basis get orthonormal basis.
  - $A = QR$ factorization.
  - Orthogonal matrix Q: $Q^T Q = I$.

- Determinants
  - Definition through elementary row operations.
  - $\det(AB) = \det(A)\det(B)$, $\det(A^T) = \det(A)$.
  - Cofactor expansion.

- Eigenvalues and eigenvectors: $Ax = \lambda x$
  - Characteristic polynomial.
  - Eigenspace.
  - Eigenbasis and diagonalization.
  - Sum and product of eigenvectors and trace and det of A.
  - Powers of A.
  - Discrete Dynamical systems: state vector $\mathbf{x}_t$ evolves in time by $\mathbf{x}_{t+1} = A\mathbf{x}_t$.

- Symmetric matrices and spectral theorem.
  - if $A = A^T$ then eigenvalues of A are real
  - A has an orthonormal basis of eigenvectors.

3 Random Examples

Example 1. Let

$$a_1 = \begin{bmatrix}1\\2\\1\end{bmatrix}, \quad a_2 = \begin{bmatrix}2\\1\\2\end{bmatrix}, \quad a_3 = \begin{bmatrix}3\\3\\3\end{bmatrix}, \quad b = \begin{bmatrix}3\\4\\4\end{bmatrix}$$

Is b a linear combination of a_1, a_2, a_3? Explain!

Solution 2. *We need to solve a system $Ax = b$, with augmented matrix* $\left[\begin{array}{ccc|c}1 & 2 & 3 & 3\\2 & 1 & 3 & 4\\1 & 2 & 3 & 4\end{array}\right] \simeq$ $\left[\begin{array}{ccc|c}1 & 2 & 3 & 3\\0 & -1 & -3 & -2\\0 & 0 & 0 & 1\end{array}\right]$. *So b is/is not a linear combination?*

Example 3. Let $A = \begin{bmatrix}2 & 1 & 1\\4 & 2 & 3\\6 & 3 & 2\end{bmatrix}$.

- Find the LU factorization.
- Find two descriptions of the column space: $\mathrm{Col}(A)$ is the span of which vectors, and if $b \in \mathrm{Col}(A)$ give equations for b.
- If $b \in \mathrm{Col}(A)$ is the general set of solutions $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ of $Ax = b$ a point, a line, a plane or all of $\mathbb{R}^3$?

Solution 4. $A = \begin{bmatrix}2 & 1 & 1\\4 & 2 & 3\\6 & 3 & 2\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\2 & 1 & 0\\3 & -1 & 1\end{bmatrix}\begin{bmatrix}2 & 1 & 1\\0 & 0 & 1\\0 & 0 & 0\end{bmatrix}$. *So* $\mathrm{Col}(A) = \mathrm{Span}(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\3\\2\end{bmatrix})$. *This is the description by directions. We can also give equations for $b \in \mathrm{Col}(A)$: such b is perpendicular to what space? What is $\dim(\mathrm{Nul}(A^T))$? So we need to find a single equation for b, for instance $5b_1 - b_2 - b_3 = 0$. (If you don't see this immediately, do row operations on* $\left[\begin{array}{ccc|c}2 & 1 & 1 & b_1\\4 & 2 & 3 & b_2\\6 & 3 & 2 & b_3\end{array}\right]$*)If $b \in \mathrm{Col}(A)$ how many solutions of $Ax = b$, how many free variables? Get point, line, plane....?*

Example 5. Let $v_1 = \begin{bmatrix}1\\2\\0\end{bmatrix}$, $v_2 = \begin{bmatrix}3\\1\\0\end{bmatrix}$ and let $V = \mathrm{Span}(v_1, v_2)$. If $x = \begin{bmatrix}1\\2\\3\end{bmatrix}$ we can write $x = x_V + x_{V^\perp}$.

- Explain why $x_V = \frac{\langle x, v_1\rangle}{\langle v_1, v_1\rangle}v_1 + \frac{\langle x, v_2\rangle}{\langle v_2, v_2\rangle}v_2$ is not correct.
- Find an orthonormal basis for V.

- Calculate $x_{V^\perp}$.

Solution 6.

- *The basis is not orthogonal, so we can not use the formula!*
- *Take* $q_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\\0\end{bmatrix}$. *Now* q_2 *must be perpendicular to* q_1 *and belong to* V.

  So $q_2 = \frac{1}{\sqrt{5}}\begin{bmatrix}-2\\1\\0\end{bmatrix}$. *(Gram-Schmidt.)[-.5cm]*

- *Now write* $x_V = \dfrac{\langle \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\2\\0\end{bmatrix}\rangle}{\langle \begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\2\\0\end{bmatrix}\rangle}\begin{bmatrix}1\\2\\0\end{bmatrix} + \dfrac{\langle \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-2\\1\\0\end{bmatrix}\rangle}{\langle \begin{bmatrix}-2\\1\\0\end{bmatrix}, \begin{bmatrix}-2\\1\\0\end{bmatrix}\rangle}\begin{bmatrix}-2\\1\\0\end{bmatrix} = \begin{bmatrix}1\\2\\0\end{bmatrix}$. *Hence*

  $x_{V^\perp} = \begin{bmatrix}0\\0\\3\end{bmatrix}$

Example 7.

- Give an example of a 2×2 matrix A that is not invertible.
- Give an example of a 2×3 matrix A that has rank 0, or explain that that is not possible.
- Give an example of a 2×3 matrix A that has rank 1, but non of the entries are zero, or explain that that is not possible.
- Give an example of a 2×3 matrix A that has rank 2.
- Is the equation $Ax = 0$ always solvable?
- If A is the 2×3 zero matrix, then $\text{Nul}(A) = \{\begin{bmatrix}0\\0\\0\end{bmatrix}\}$. True or false?

Example 8. Let $W = \{\begin{bmatrix}x\\y\\0\end{bmatrix}\}$, subspace of $\mathbb{R}^3$. If possible:

- Find 3 dependent vectors in W.
- Find 1 dependent vector in W.
- Find 2 independent vectors in W.
- Find 3 independent vectors in W.
- Find a spanning set of W containing 3 vectors.
- Find a spanning set of W containing 2 vectors.

- Find a spanning set of W containing 1 vectors.
- Find 2 bases for W.
- Find 2 independent vectors in $W^{\perp}$.