June 28th, 2024

Linear Algebra

Via Introduction to Linear Algebra | Fourth Edition

Chapter 1: Introduction to Vectors

- At the heart of linear algebra lies two vector operations: Addition $(v+w)$ and linear combinations $(cv+dw,\ c,d \in \mathbb{R})$
  - Example:

$$cv + wd = c\begin{bmatrix}1\\1\end{bmatrix} + d\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}c+2d\\c+3d\end{bmatrix}$$

- The vector v and its scalar multiple cv lie on the same line when w is not on that line (i.e. they are linearly independent), $cv+dw$ span the whole 2D plane.

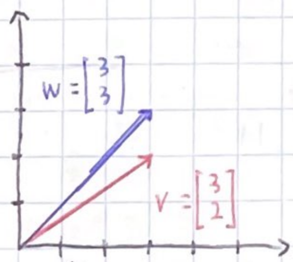


fig 1.1.1

Chapter 1.1: Vectors and Linear Combinations

Introduction

- Vectors in this book are represented as columns:

$$v = \begin{bmatrix}v_1\\v_2\end{bmatrix}, \quad v_1 = \text{First component},\ v_2 = \text{Second component}$$

- Vector Addition:

$$v = \begin{bmatrix}v_1\\v_2\end{bmatrix},\ w = \begin{bmatrix}w_1\\w_2\end{bmatrix},\ v+w = \begin{bmatrix}v_1+w_1\\v_2+w_2\end{bmatrix}$$

- Scalar Multiplication:

$$v = \begin{bmatrix}v_1\\v_2\end{bmatrix},\ 2v = \begin{bmatrix}2v_1\\2v_2\end{bmatrix},\ -v = \begin{bmatrix}-v_1\\-v_2\end{bmatrix}$$

- Zero vector
  - The sum of v and $-v$ is the zero vector $\mathbf{0} = \begin{bmatrix}0\\0\end{bmatrix}$ and it is not the same as 0 scalar.

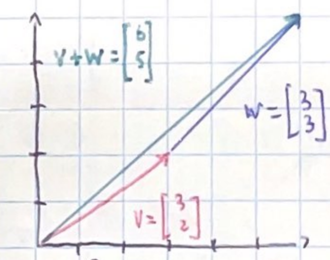


fig 1.1.2

Linear Combinations

- The vector sum $cv + dw$ is a Linear Combination of v and w
- Vectors can be visualized on a cartesian plane, although it should be noted the position of a vector is not part of its identity. Vectors can be "moved" as long as its magnitude and direction remain the same.
  - See Fig 1.1.1 and Fig 1.1.2

Vectors in Three Dimensions

- A vector with 2 components correspond to a point in the Cartesian Plane. (It is the position vector of that point)
- Similarly, a vector in three dimensions represents a point in 3D space.

$$v = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad w = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \quad v + w = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$$; See fig 1.1.3

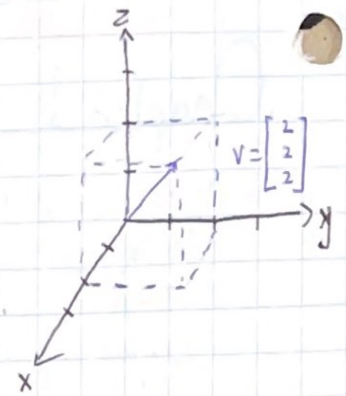


fig 1.1.3

- To save space, a column vector may be written as $v = (v_1, v_2, v_3, \ldots)$. Note this is <u>not</u> the same as a row vector $v = [v_1, v_2, v_3, \ldots]$
- Given 3 non-zero <u>linearly independent</u> vectors u, v, w,
  - The combinations cu fill a line
  - The combinations $cu + dv$ fill a plane
  - The combinations $cu + dv + ew$ fill a three-dimension space
- A set of vectors is <u>linearly independent</u> if no vector in the set can be expressed as a linear combination of the others

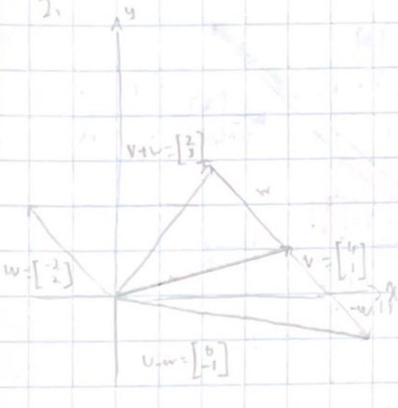
Problem Set 1.1

1.

(a) Since $v = (1,2,3)$ and $w = (3,6,9)$ are linearly dependent, the combinations $cv + dw$ span only a line.

(b) Since $v = (1,0,0)$ and $w = (0,2,3)$ are linearly independent, the combinations $cv + dw$ span a plane

2.



3.

$$\begin{cases} v + w = \begin{bmatrix} 5 \\ 1 \end{bmatrix} & ① \\ v - w = \begin{bmatrix} 1 \\ 5 \end{bmatrix} & ② \end{cases}$$

$① + ②: \ 2v = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$

$v = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

Sub $v = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ into ②

$\begin{bmatrix} 3 \\ 3 \end{bmatrix} - w = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

$w = \begin{bmatrix} 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

$= \begin{bmatrix} 2 \\ -2 \end{bmatrix}$

$(v, w) = \left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}\right)$

5.

$u + v + w = (0,0,0)$

Since $\exists\ (c,d,e) \neq (0,0,0),\ c,d,e \in \mathbb{R}$ s.t. $cu + dv + ew = 0$, the vector set $\{u, v, w\}$ are linearly dependent and span a plane

6.

$cv + dw = (3,3,-6)$

$c(1,-2,1) + d(0,1,-1) = (3,3,-6)$

$(c, -2c+d, c-d) = (3,3,-6)$

$$\begin{cases} c = 3 \\ -2c+d = 3 \\ c-d = -6 \end{cases}$$

$(c,d) = (3,9)$

Chapter 1.2 : Lengths and Dot Products

- The dot product or inner product of vectors $v=(v_1, v_2)$ and $w=(w_1, w_2)$ is defined as:

  angle between the two vectors "tail-to-tail"

  $$v \cdot w = v_1 w_1 + v_2 w_2 = \|v\|\|w\| \cos\theta$$

- If, for any 2 non-zero vectors, their dot product is zero, the two vectors are perpendicular. For example

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

- The dot product is commutative, that is, $v \cdot w = w \cdot v$
- Example Application : Engineering

  Given a seesaw with weights as such:

  The seesaw will remain balanced because the dot product $(4)(-1)+(2)(2)=0$.
  The vector of weights $(w_1, w_2)=(4,2)$ and the vector of distances $(v_1, v_2)=(-1,2)$
  The weight multiplied by distance give the "moments", a quantity proportional to the torque generated. If the moments equal zero, the seesaw remains balanced.

- Example Application : Economics

  We have 3 goods to buy and sell. Their prices are (p_1, p_2, p_3) - The price vector.
  The quantities we buy or sell is (q_1, q_2, q_3) – positive when selling, negative when buying.
  Their dot product in three dimensions is the total income:

  $p_1 q_1 + p_2 q_2 + p_3 q_3 =$ total income $= (p_1, p_2, p_3) \cdot (q_1, q_2, q_3)$

- Dot product (General form):

  Given $v=(v_1, v_2, v_3, \ldots v_n)$, $w=(w_1, w_2, w_3, \ldots w_n)$, $n \in \mathbb{N}$

  $$v \cdot w = \sum_{i=1}^{n} v_i w_i$$

Lengths and unit vectors

- A vector's dot product with itself generates a scalar equal to the <u>magnitude of the vector squared.</u>
- The <u>length of a vector</u> $v=(v_1, v_2, v_3, \ldots, v_n)$ is $\|v\| = \sqrt{\sum_{i=1}^{n} v_i^2}$
- A <u>Unit vector</u> is a vector with magnitude 1, often in the direction of another vector
- For any vector v, $v \neq 0$, we can get a unit vector in its direction by dividing the vector by its length. $\hat{v} = \frac{1}{\|v\|} v$

- The standard unit vectors along the x, y, and z-axis are $\hat{i}, \hat{j}, \hat{k}$
- In the xy plane, a unit vector making an angle "θ" with the positive x-axis is given as $u = (\cos\theta, \sin\theta)$
  - When $\theta = 0°$, the horizontal vector u is $\hat{i}$.
  - When $\theta = 90°$, the vertical vector u is $\hat{j}$.
  - For any θ, $u \cdot u = 1$ because $\cos^2\theta + \sin^2\theta = 1$
- Since $u \cdot v = |u||v|\cos\theta$, the angle between 2 vectors is given as

$$\theta = \cos^{-1}\left(\frac{u \cdot v}{|u||v|}\right)$$

- Starting from $\frac{u \cdot v}{|u||v|}$, which is $-1 \le \frac{u \cdot v}{|u||v|} \le 1$,

$$\frac{|u \cdot v|}{|u||v|} \le 1$$

$$\boxed{|u \cdot v| \le |u||v|}$$

which is the Cauchy - Bunyakovsky - Schwartz Inequality for Euclidean Vectors.
In Euclidean Space, this becomes

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \le \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right)$$

- Taking the dot product of $u = (a, b)$ and $v = (b, a)$ is $2ab$. The lengths of both u and v are $\sqrt{a^2 + b^2}$. The Cauchy-Swartz Inequality says $2ab \le a^2 + b^2$

$$2ab \le a^2 + b^2 \Rightarrow ab \le \frac{a^2+b^2}{2} \Rightarrow \text{let } x = a^2, y = b^2, \boxed{\sqrt{xy} \le \frac{x+y}{2}}$$

which is the famous AM-GM Inequality.

- Triangle Inequality $\boxed{\|u + v\| \le \|u\| + \|v\|}$

Problem Set 1.2

1.

$u \cdot v = (-0.6)(3) + (0.8)(4) = \frac{7}{5}$

2.

$\|u\| = \sqrt{(-0.6)^2 + (0.8)^2}$
$= 1$

$\|v\| = \sqrt{(3)^2 + (4)^2}$
$= 5$

$\|w\| = \sqrt{(8)^2 + (6)^2}$
$= 10$

3.

$\hat{u} = \frac{1}{\|u\|} u$

$= \frac{1}{\sqrt{(0.6)^2 + (0.8)^2}} u$

$= u$

4.

a.

$v \cdot (-v)$
$= \begin{bmatrix} 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ -4 \end{bmatrix}$
$= (3)(-3) + (4)(-4)$
$= -25$

b.

$(v + w) \cdot (v - w)$
$= v \cdot v + w \cdot w$
$= \begin{bmatrix} 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 8 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 6 \end{bmatrix}$
$= (3)(3) + (4)(4) + (8)(8) + (6)(6)$
$= 125$

5.

$u_1 = \frac{1}{\|v\|} v$

$= \frac{1}{\sqrt{(3)^2 + (4)^2}} v$

$= \frac{1}{5} \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$= \left(\frac{3}{5}, \frac{4}{5}\right)$

$u_2 = \left(-\frac{4}{5}, \frac{3}{5}\right)$

7.

b.

$v = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \quad w = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$

$\theta = \cos^{-1}\left(\frac{v \cdot w}{\|v\|\|w\|}\right)$

$= \cos^{-1}\left(\frac{(2)(2) + (2)(-1) + (-1)(2)}{\sqrt{(2)^2 + (2)^2 + (-1)^2}\sqrt{(2)^2 + (-1)^2 + (2)^2}}\right) = 90°$

12.

$v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; w = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

Set $v \cdot (w - cv) = 0$

$v \cdot w - c\, v \cdot v = 0$

$(1,1) \cdot (1,5) - c\,(1,1) \cdot (1,1) = 0$

$(1)(1) + (1)(5) - c(1)(1) - c(1)(1) = 0$

$6 - 2c = 0$

$c = 3$

Let $v = (v_1, v_2)$, $w = (w_1, w_2)$

Set $v \cdot (w - cv) = 0$

$v \cdot w - c\, v \cdot v = 0$

$c\, v \cdot v = v \cdot w$

$c = \dfrac{v \cdot w}{v \cdot v}$

16.

$v = (\underbrace{1, 1, 1, \ldots, 1}_{9 \text{ times}})$

$\|v\| = \sqrt{(1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2}$

$= \sqrt{9}$

$= 3$

$u = \dfrac{1}{\|v\|} v$

$= \frac{1}{3} v$

$= (\underbrace{\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, \ldots \tfrac{1}{3}}_{9 \text{ times}})$

Let $V = (1, -1, 1, -1, 1, -1, 1, -1, 0)$

$\|V\| = \sqrt{\underbrace{(1)^2 + (-1)^2 \cdots + (-1)^2}_{8 \text{ times}} + (0)^2}$

$= \sqrt{8} = 2\sqrt{2}$

$U = \dfrac{1}{\|V\|} V$

$= \frac{\sqrt{2}}{4} V$

$= (\underbrace{\tfrac{\sqrt{2}}{4}, -\tfrac{\sqrt{2}}{4}, \tfrac{\sqrt{2}}{4}, -\tfrac{\sqrt{2}}{4}, \ldots -\tfrac{\sqrt{2}}{4}}_{8 \text{ times}}, 0)$

17.

$\cos\alpha = \dfrac{v \cdot i}{\|v\|\|i\|}$

$= \dfrac{(1,0,-1) \cdot (1,0,0)}{\sqrt{(1)^2 + (0)^2 + (-1)^2}\sqrt{(1)^2 + (0)^2 + (0)^2}}$

$= \dfrac{1}{\sqrt{2}}$

18.

$\|v\|^2 + \|w\|^2 = \|v + w\|^2$

$\|(4,2)\|^2 + \|(-1,2)\|^2 = \|(3,4)\|^2$

$(4)^2 + (2)^2 + (-1)^2 + (2)^2 = (3)^2 + (4)^2$

$25 = 25$

21

$\|u + v\| \leq \|u\| + \|v\|$

$\|u + v\|^2 \leq (\|u\| + \|v\|)^2$

$(u + v) \cdot (u + v) \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2$

$u \cdot u + 2u \cdot v + v \cdot v \leq u \cdot u + 2\|u\|\|v\| + v \cdot v$

$2u \cdot v \leq 2\|u\|\|v\|$

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

which is the Cauchy-Schwarz Inequality

22.

$|v \cdot w| \leq |v||w|$

$|v \cdot w|^2 \leq |v|^2 |w|^2$

$(v_1 w_1 + v_2 w_2)^2 \leq (v_1^2 + v_2^2)(w_1^2 + w_2^2)$

$v_1^2 w_1^2 + 2 v_1 w_1 v_2 w_2 + v_2^2 w_2^2 \leq v_1^2 w_1^2 + v_1^2 w_2^2 + v_2^2 w_1^2 + v_2^2 w_2^2$

$2 v_1 w_1 v_2 w_2 \leq v_1^2 w_2^2 + v_2^2 w_1^2$

$0 \leq v_1^2 w_2^2 + v_2^2 w_1^2 - 2 v_1 w_1 v_2 w_2$

$0 \leq (v_1 w_2 - v_2 w_1)^2$

27.

$|v - w|^2 = (v - w) \cdot (v - w)$

$= v \cdot v - 2v \cdot w + w \cdot w$

$= |v|^2 - 2v \cdot w + |w|^2$ Set $\theta = 90$ ✓

$= |v|^2 + |w|^2 - 2|v||w|\cos\theta$

$= 5^2 + 3^2$

$= 34$

Chapter 1.3: Matrices

- Example 1:

Given $u = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $v = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, $w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Their linear combinations in 3D space are:

$$c\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + d\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + e\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c \\ d-c \\ e-d \end{bmatrix}$$

The above combination can be written as a <u>matrix multiplied by a vector</u>

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} c \\ d \\ e \end{bmatrix}}_{x} = \begin{bmatrix} c \\ d-c \\ e-d \end{bmatrix}$$

$$Ax = \begin{bmatrix} u & v & w \end{bmatrix}\begin{bmatrix} c \\ d \\ e \end{bmatrix} = cu + dv + ew$$

This is more than a simple rewrite. Instead of viewing it <u>as the scalars c, d, and e multiplying the vectors</u>, now <u>the matrix is multiplying the scalars.</u>

<u>The matrix A acts on the vector x.</u>

The result of Ax is a combination b of the columns of A.

Rewriting with new variables:

$$Ax = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = b$$

$$\boxed{A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad \text{"difference matrix"}}$$

The <u>input vector is x</u> and the <u>output vector is b</u>.

This specific matrix is called a <u>"difference matrix"</u> because its output consists of the differences of the input vector x.

Computing $Ax = b$ with the squares

$$Ax = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 1-0 \\ 4-1 \\ 9-4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

- When multiplying a matrix and column vector, the output is another column vector, where the ith entry is the dot product of the ith row of the matrix and the input matrix.

Linear Equations

- Originally, we viewed the equation $Ax=b$ with b as the unknown.
- Now, we view b as known and x as unknown. That is, we want to solve for x.
  - Instead of: compute the linear combination $x_1u + x_2v + x_3w$ to find b.
  - Which combination of u, v, and w produces a particular vector b.
- From the previous example:

$$\begin{matrix} x_1 & = b_1 \\ x_2 - x_1 & = b_2 \\ x_3 - x_2 & = b_3 \end{matrix} \Rightarrow \begin{matrix} x_1 = b_1 \\ x_2 = b_1 + b_2 \\ x_3 = b_1 + b_2 + b_3 \end{matrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

- Most linear systems are not easy to solve. The above example was in a particular arrangement (it is lower triangular).
- We can say matrix A is invertible: From b, we can recover x.

The Inverse Matrix

- The "inverse matrix" of the aforementioned difference matrix is the sum matrix

$$\boxed{S = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}}. \quad \text{That is, } Ax=b \Rightarrow x = Sb \Rightarrow A^{-1} = S,\ S = A^{-1}$$

- For example, given $x=(1,2,3)$

$$Ax = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{b} \Rightarrow Sb = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{x}$$

S and A are inverse matrices

Cyclic Differences

- Modifying the difference matrix slightly, we get the cyclic difference matrix C.

$$\boxed{C = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}} \Rightarrow Cx = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = b$$

- This matrix is not triangular, so it is not easy to solve for x when given b.
  - There are either 0 or infinitely many solutions
  - Case $Cx = 0$

$$\begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ is solved by } x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c \\ c \\ c \end{bmatrix}$$

  - Case $Cx = (1,3,5)$

$$\begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$ has no solutions. Geometrically, this means no combination of $(1,-1,0)$, $(0,1,-1)$, and $(-1,0,1)$ will create $(1,3,5)$; they are linearly dependent and $(1,3,5)$ does not lie on the spanned plane.

Linear Independence and Dependence

- A set of vectors is said to be <u>linearly independent</u> is there exists no nontrivial linear combination of the vectors that equals the zero vector. If there exists nontrivial solutions, the vectors are said to be <u>linearly dependent</u>
- Linear independence implies that no vector in the set can be expressed as a linear combination of the others.
- A set of n linearly independent vectors in $\mathbb{R}^n$ will span $\mathbb{R}^n$.
  - For example, three linearly independent vectors in $\mathbb{R}^3$ will span $\mathbb{R}^3$.
  - If one vector can be expressed as a linear combination of the two others, the set can effectively "reduced" to two and other a plane in $\mathbb{R}^3$ can be spanned.
- For the <u>cyclic difference matrix</u>, since the vectors are linearly dependent $(w = -u - v)$, the vectors lie on one plane (coplanar).
  - This explains why a solution is not guaranteed for $Cx = b$
- Iff the columns of an n by n square matrix are <u>independent</u> (forms a spanning set of $\mathbb{R}^n$):
  - $Ax = 0$ has 1 solution (trivial). <u>A is invertible.</u> $Ax = b$ has 1 solution
- Iff the columns are <u>dependent</u>
  - $Ax = 0$ has multiple solutions. <u>A is singular.</u> $Ax = b$ may have solution

Triangular Matrices

- A triangular matrix is a special case of a square matrix.
  - A lower triangular matrix has all entries above the main diagonal as zero
  - An upper triangular matrix has all entries below the main diagonal as zero

$$L = \begin{bmatrix} l_{1,1} & & & 0 \\ l_{2,1} & l_{2,2} & & \\ \vdots & \vdots & \ddots & \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n} \end{bmatrix}, \quad U = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ & u_{2,2} & & u_{2,n} \\ & & \ddots & \vdots \\ & & & u_{n,n} \end{bmatrix}$$

Problem Set 1.3

1.

$2s_1 + 3s_2 + 4s_3 = b$

$b = 2(1,1,1) + 3(0,1,1) + 4(0,0,1)$
$= (2,2,2) + (0,3,3) + (0,0,4)$
$= (2,5,9)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} (1,0,0)\cdot(2,3,4) \\ (1,1,0)\cdot(2,3,4) \\ (1,1,1)\cdot(2,3,4) \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix}$$

2.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_1 + y_2 \\ y_1 + y_2 + y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ y_1 + y_2 \\ y_1 + y_2 + y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}$$

$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

S is invertible

3.

$$\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}B_1\\B_2\\B_3\end{bmatrix}$$

$$\begin{bmatrix}y_1\\y_1+y_2\\y_1+y_2+y_3\end{bmatrix}=\begin{bmatrix}B_1\\B_2\\B_3\end{bmatrix}$$

$$\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}B_1\\B_2-y_1\\B_3-y_1-y_2\end{bmatrix}$$

$$\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}B_1\\B_2-B_1\\B_3-B_1-B_2+B_1\end{bmatrix}=\begin{bmatrix}B_1\\B_2-B_1\\B_3-B_2\end{bmatrix}$$

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

The matrix is invertible $\Leftrightarrow$ The column vectors are linearly independent

4.

$$\begin{bmatrix}1&4&7\\2&5&8\\3&6&9\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$\begin{bmatrix}x_1+4x_2+7x_3\\2x_1+5x_2+8x_3\\3x_1+6x_2+9x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$x=\begin{bmatrix}0\\0\\0\end{bmatrix}\text{ or }x=\begin{bmatrix}1\\-2\\1\end{bmatrix}$$

$w_1-2w_2+w_3=0$

These column vectors are linearly dependent and coplanar

$(a,b)\Rightarrow k(c,d)$

$(a,b)\Rightarrow(kc,kd)$

$(b,d)=(kd,d)$
$=d(k,1)$

$(a,c)=(kc,c)$
$=c(k,1)$

7. Each row is a vector in 3D which is perpendicular to x because their dot product is 0.

8.

$$Ax=\begin{bmatrix}1&0&0&0\\-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}$$

$$\begin{bmatrix}x_1\\x_2-x_1\\x_3-x_2\\x_4-x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}\Rightarrow\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}b_1\\b_2+b_1\\b_3+b_2+b_1\\b_4+b_3+b_2+b_1\end{bmatrix}=\underbrace{\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&1&1&0\\1&1&1&1\end{bmatrix}}_{S}\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}$$

9.

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

$$Cx=\begin{bmatrix}1&0&0&-1\\-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}x_1-x_4\\x_2-x_1\\x_3-x_2\\x_4-x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

$$x=\begin{bmatrix}c\\c\\c\\c\end{bmatrix}$$

10.

$$\Delta z=\begin{bmatrix}-1&1&0\\0&-1&1\\0&0&-1\end{bmatrix}\begin{bmatrix}z_1\\z_2\\z_3\end{bmatrix}$$

$$=\begin{bmatrix}z_2-z_1\\z_3-z_2\\-z_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

$$z=\begin{bmatrix}-b_3-b_2-b_1\\-b_3-b_2\\-b_3\end{bmatrix}=\begin{bmatrix}-b_1-b_2-b_3\\-b_2-b_3\\-b_3\end{bmatrix}$$

$$\underbrace{\begin{bmatrix}-1&-1&-1\\0&-1&-1\\0&0&-1\end{bmatrix}}_{\Delta^{-1}}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

12

$$Cx=\begin{bmatrix}0&1&0&0\\-1&0&1&0\\0&-1&0&1\\0&0&-1&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}$$

$$\begin{bmatrix}x_2\\x_3-x_1\\x_4-x_2\\-x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}\Rightarrow\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}-b_4-b_2\\b_1\\-b_4\\b_3+b_1\end{bmatrix}=\begin{bmatrix}&-b_2&&-b_4\\b_1&&&\\&&&-b_4\\b_1&&b_3&\end{bmatrix}$$

$$C^{-1}=\begin{bmatrix}0&-1&0&-1\\1&0&0&0\\0&0&0&-1\\1&0&1&0\end{bmatrix}B_1$$

Chapter 2: Solving Linear Equations

Chapter 2.1: Vectors and Linear Equations

- The central problem of linear algebra is to solve a system of linear equations.
- A simple example:

  $\begin{cases} x - 2y = 1 : l_1 \\ 3x + 2y = 11 : l_2 \end{cases}$

  $(x, y) = (3, 1)$,

  that is $(3,1) \in l_1$

  and $(3,1) \in l_2$

  - See fig 2.1.1

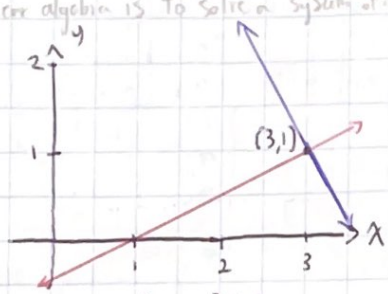
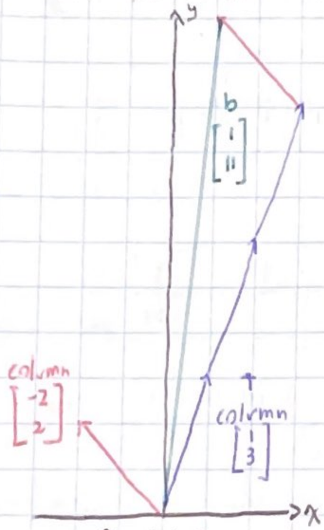


fig 2.1.1

- However, as a matrix, we see 2 different viewpoints
  - As rows: we see 2 simultaneous equations whose solution is one which satisfies both equations.
  - As columns: we instead see a vector equation (see fig 2.1.2)

$$\underbrace{x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}}_{\text{"Find the linear combination"}} \Rightarrow \underbrace{\begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}}_{\text{"A"}}\underbrace{\begin{bmatrix}x\\y\end{bmatrix}}_{\text{"x"}} = \underbrace{\begin{bmatrix}1\\11\end{bmatrix}}_{\text{"b"}} \Rightarrow 3\begin{bmatrix}1\\3\end{bmatrix} + 1\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$$



Sig 2.1.2

- We view the 2 equations as a linear combination problem. Find the coefficients x and y which generate the vector $(1, 11)$ from the column vectors.
- The matrix generated by combining the column vectors is called the coefficient matrix.

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

- Finally we can view this system as a matrix equation:

  $Ax = b \Rightarrow \begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$

  $\begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}\begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$

- This chapter addresses the use of matrices to solve n equations for n unknowns with concepts such as matrix multiplication and matrix inversion.
- Four steps needed to understand elimination using matrices.
  1. Elimination goes from A to triangular U by a sequence of matrix steps E_{ij}
  2. The inverse matrices E_{ij}^{-1} in reverse order bring U back to the original A
  3. In matrix language that reverse language is $A = LU$ = (lower triangle)(upper triangle)
  4. Elimination succeeds if A is invertible (It may need row exchanges)
- The most used algorithm (LU decomposition) uses these steps

Three Equations in Three Unknowns

- Given the equations

$$\begin{cases} x + 2y + 3z = 6 \\ 2x + 5y + 2z = 4 \\ 6x - 3y + z = 2 \end{cases}$$

- As a row, we see 3 planes which, in this case, intersect at a point.
- As a column we see the following vector equation:

$$x\begin{bmatrix}1\\2\\6\end{bmatrix} + y\begin{bmatrix}2\\5\\-3\end{bmatrix} + z\begin{bmatrix}3\\2\\1\end{bmatrix} = \begin{bmatrix}6\\4\\2\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}6\\4\\2\end{bmatrix}$$

- Since our column vectors are linearly independent, any vector b in $\mathbb{R}^3$ can be generated.
  - In this case $(x,y,z) = (0,0,2)$ produces $(6,4,2)$

The Matrix Form of the Equations

- With our above example, our coefficient matrix is:

$$A = \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix} \Rightarrow Ax = \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$$

- We can perform the multiplication in 2 ways
  - Rows:

$$Ax = \begin{bmatrix}(\text{row } 1)\cdot x\\(\text{row } 2)\cdot x\\(\text{row } 3)\cdot x\end{bmatrix}$$

  - Columns

$$Ax = x(\text{column } 1) + y(\text{column } 2) + z(\text{column } 3)$$

- Example with 2 matrices: A and I

$$Ax = \begin{bmatrix}1 & 0 & 0\\1 & 0 & 0\\1 & 0 & 0\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}4\\4\\4\end{bmatrix}, \quad Ix = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}4\\5\\6\end{bmatrix}$$

  - I is the identity matrix, a matrix with 1s on the "main diagonal"; multiplying a matrix by it leaves the matrix unchanged
    - $Ix = x$

Matrix Notation

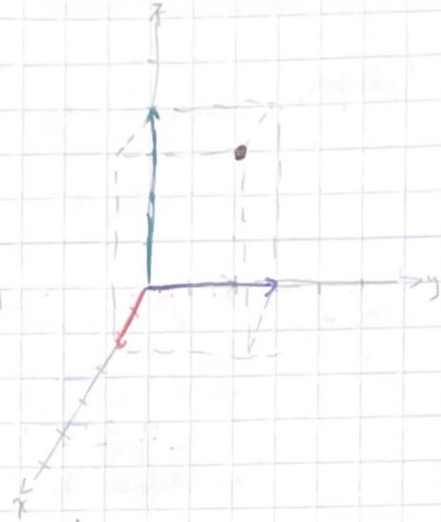
- We use the notation:

$$A = \underbrace{\begin{bmatrix}a_{11} & a_{12} & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{bmatrix}}_{a_{\text{row column}}}$$

- For an $m \times n$ matrix,
  - i (row) goes from 1 to m (inclusive)
  - j (column) goes from 1 to n (inclusive)

Problem set 2.1

1.



3.

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} \Rightarrow \begin{bmatrix}1&0&0\\1&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}2\\5\\4\end{bmatrix}$$

- All but the solutions have changed.

4.

$$\begin{cases} x+y+3z = 6 & ① \\ x-y+z = 4 & ② \end{cases}$$

let $t=z$

①+②: $2x+4t=10$

$2x = 10-4t$

$x = 5-2t$

sub $x=5-2t$, $t=z$ into ②

$5-2t-y+t=4$

$5-t-y=4$

$y=1-t$

$(x,y,z) = (5-2t, 1-t, t)$

sub $z=t=2$

$(x,y,z) = (1,-1,2)$

sub $z=t=0$

$(x,y,z) = (5,1,0)$

sub $z=t=1$

$(x,y,z) = (3,0,1)$

7.

$$\begin{bmatrix}1&1&1\\1&2&1\\2&3&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}2\\3\\5\end{bmatrix} \Rightarrow \begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\1\\1\end{bmatrix}$$

$$\begin{bmatrix}1&1&1\\1&2&1\\2&3&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}4\\6\\c\end{bmatrix} \Rightarrow c=10$$

8

$$\begin{bmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{bmatrix}\begin{bmatrix}x\\y\\z\\t\end{bmatrix} = \begin{bmatrix}3\\3\\3\\2\end{bmatrix}$$

$\Downarrow$

$$\begin{cases} x+y+z+t=3 \\ y+z+t=3 \\ z+t=3 \\ t=2 \end{cases} \Rightarrow \begin{cases} x=0 \\ y=0 \\ z=1 \\ t=2 \end{cases}$$

9.

a) $\begin{bmatrix}1&2&4\\-2&3&1\\-4&1&2\end{bmatrix}\begin{bmatrix}2\\2\\3\end{bmatrix} = \begin{bmatrix}20\\5\\0\end{bmatrix}$ b) $\begin{bmatrix}2&1&0&0\\1&2&1&0\\0&1&2&1\\0&0&1&2\end{bmatrix}\begin{bmatrix}1\\1\\1\\2\end{bmatrix} = \begin{bmatrix}3\\3\\3\\5\end{bmatrix}$

10

$$Ax = 2\begin{bmatrix}1\\-2\\-4\end{bmatrix} + 2\begin{bmatrix}2\\3\\1\end{bmatrix} + 3\begin{bmatrix}4\\1\\2\end{bmatrix} = \begin{bmatrix}20\\5\\0\end{bmatrix}$$

11.

a) $\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 14 \\ 22 \end{bmatrix}$ b) $\begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$

12.

a) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} z \\ y \\ x \end{bmatrix}$ b) $\begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ c) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 6 \end{bmatrix}$

13

a) An m×n matrix multiplies a vector with n components to produce a vector with m components

b) The planes from the m equations $Ax=b$ are in n-dimensional space.
The combination of columns of A is in m dimensional space

14.

$\begin{bmatrix} 2 & 3 & 1 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 8 \end{bmatrix}$

15.

a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = I$

b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix} = P$

16

a) $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix} = R$

b) $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ -y \end{bmatrix} = R^2$

17.

$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{P}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$

18

$E\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \Rightarrow \underbrace{\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}}_{E}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

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

19.

$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}}_{E}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z+x \end{bmatrix}, \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}}_{E^{-1}}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z-x \end{bmatrix}$

$E\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 8 \end{bmatrix} \Rightarrow E^{-1}\begin{bmatrix} 3 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$

21.

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

and

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

$R = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$

22.

$\begin{bmatrix} 1 & 4 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$

plane $\perp$ vector $\Rightarrow (1, 4, 5)$

1D

20

$P_1 x = b \Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$

$P_2 x = b \Rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$

$P_1\begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$

$P_2\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Chapter 2.2: The Idea of Elimination

- Elimination is a systematic way to solve linear equations!
- For example:

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \Rightarrow \begin{cases} x - 2y = 1 \\ \quad\ 8y = 8 \end{cases}$$ gives us $y = 1 \Rightarrow x = 3$

- Elimination produces an <u>upper triangular system</u>, which is the goal.
  - That system is quickly solved via substitution from the bottom up (<u>back substitution</u>)
  - If instead we form a lower triangular system, substitution would start from the top down (<u>forward substitution</u>)
- In the example above, to eliminate x, we subtracted a multiple of equation 1 from equation 2.
  - How was the multiplier (3) found?
    The first pivot was 1 (the coefficient of x) and the second equation had a coefficient of 3 for the x. Therefore the multiplier is 3.
- If instead we had the equations:

$$\begin{cases} 4x - 8y = 4 \\ 3x + 2y = 11 \end{cases}$$

*the nth pivot is the entry of a_{nn} after performing elimination $n-1$ times

our pivot would be 4 and the multiplier $\ell = \frac{3}{4}$

- Pivot*: First <u>nonzero entry</u> in the row performing the elimination. $\ell_{ij} = \dfrac{\text{entry to eliminate in row } i}{\text{pivot in row } j}$
- Multiplier: (entry to eliminate) / (pivot)
- To solve n equations, we want n pivots

Breakdown of Elimination

- Sometimes, when performing elimination, a pivot may not exist (all zeroes)
  - If we end with $0x = c$, $c \neq 0$, there are no solutions
  - On the other hand, $0x = 0$ implies infinite solutions
- From our viewpoints (<u>No Solutions</u>)
  - Rows

$$\begin{cases} x - 2y = 1 \\ 3x - 6y = 11 \end{cases} \Rightarrow$$

"Parallel lines"

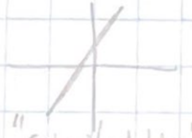
  - Columns

$$x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\-6\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$$

Collinear columns cannot form certain vectors

- From our viewpoints (<u>Infinite Solutions</u>)
  - Rows

$$\begin{cases} x - 2y = 1 \\ 3x - 6y = 3 \end{cases}$$



"coincident lines"

  - Columns

$$x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\-6\end{bmatrix} = \begin{bmatrix}1\\3\end{bmatrix}$$

(1,3) lies in the span of (1,3) and (-2,-6)

- Case 3: Temporary failure
  - Given

$$\begin{cases} 0x + 2y = 4 \\ 3x - 2y = 5 \end{cases} \Rightarrow$$ we can simply exchange rows to get an upper triangular matrix $$\Rightarrow \begin{cases} 3x - 2y = 5 \\ \quad\quad 2y = 4 \end{cases}$$

Three Equations in Three Unknowns

- Gaussian Elimination
- Combine a coefficient matrix and right side to form augmented matrix

Problem Set 2.2

12.

$$\begin{array}{l} 2x+3y+z=8 \\ 4x+7y+5z=20 \\ -2y+2z=0 \end{array} \xrightarrow[\text{into } R_2]{R_2-2R_1} \begin{array}{l} 2x+3y+z=8 \\ y+3z=4 \\ -2y+2z=0 \end{array} \xrightarrow[\text{into } R_3]{R_3+2R_2} \begin{array}{l} 2x+3y+z=8 \\ y+3z=4 \\ 8z=8 \end{array} \rightarrow \begin{array}{l} 2x+3y+z=8 \\ y+3z=4 \\ z=1 \end{array} \rightarrow \begin{array}{l} x=2 \\ y=1 \\ z=1 \end{array}$$

21.

$$\begin{array}{l} 2x+y=0 \\ x+2y+z=0 \\ y+2z+t=0 \\ z+2t=5 \end{array} \rightarrow \begin{array}{l} 2x+2y+2z+t=0 \\ x+2y+2z+2t=5 \\ y+2z+t=0 \\ z+2t=5 \end{array} \xrightarrow[\text{into } R_2]{R_2-\frac{1}{2}R_1} \begin{array}{l} 2x+2y+2z+t=0 \\ y+z+\frac{3}{2}t=5 \\ y+2z+t=0 \\ z+2t=5 \end{array}$$

$$\xrightarrow[\text{into } R_3]{R_2-R_3} \begin{array}{l} 2x+2y+2z+t=0 \\ y+z+\frac{3}{2}t=5 \\ -z+\frac{1}{2}t=5 \\ z+2t=5 \end{array} \xrightarrow[\text{into } R_4]{R_4+R_3} \begin{array}{l} 2x+2y+2z+t=0 \\ y+z+\frac{3}{2}t=5 \\ -z+\frac{1}{2}t=5 \\ \frac{5}{2}t=10 \end{array} \rightarrow \begin{array}{l} x=-1 \\ y=2 \\ z=-3 \\ t=4 \end{array}$$

$$\begin{array}{l} 2x-y=0 \\ -x+2y-z=0 \\ -y+2z-t=0 \\ -z+2t=5 \end{array} \xrightarrow[\text{into } R_1]{R_1+R_3} \begin{array}{l} 2x-2y+2z+t=0 \\ -x+2y-z=0 \\ -y+2z-t=0 \\ -2+2t=5 \end{array} \xrightarrow[\text{into } R_2]{R_2+R_4} \begin{array}{l} 2x-2y+2z+t=0 \\ -x+2y-2z+2t=5 \\ -y+2z-t=0 \\ -z+2t=5 \end{array} \quad \begin{array}{l} x=-\frac{25}{7} \\ y=\frac{10}{7} \\ z=\frac{5}{7} \\ t=\frac{20}{7} \end{array}$$

$$\xrightarrow[\text{into } R_2]{R_2+\frac{1}{2}R_1} \begin{array}{l} 2x-2y+2z+t=0 \\ y-z+\frac{5}{2}t=5 \\ -y+2z-t=0 \\ -z+2t=5 \end{array} \xrightarrow[\text{into } R_3]{R_2+R_3} \begin{array}{l} 2x-2y+2z+t=0 \\ y-z+\frac{5}{2}t=5 \\ z+\frac{3}{2}t=5 \\ -z+2t=5 \end{array} \xrightarrow[\text{into } R_4]{R_3+R_4} \begin{array}{l} 2x-2y+2z+t=0 \\ y-z+\frac{5}{2}t=5 \\ z+\frac{3}{2}t=5 \\ \frac{7}{2}t=10 \end{array}$$

Chapter 2.3 Elimination Using Matrices

- Example System:

"$Ax=b$"

$$\begin{cases} 2x_1+4x_2-2x_3=2 \\ 4x_1+9x_2-3x_3=8 \\ -2x_1-3x_2+7x_3=10 \end{cases} \text{ is equivalent to } \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}$$

  - An $n \times n$ square matrix multiplied by a vector in n-dimensional space
- Ax is a combination of the columns of A
- The ith component of Ax is the dot product of the ith row of A and x
  - The ith component of $Ax = \sum_{j=1}^{n} a_{ij} x_j$
- Example:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 4 \cdot 1 \\ 5 \cdot 2 + 6 \cdot 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 16 \end{bmatrix}$$

The Matrix Form of One Elimination Step

- We want a way to represent an elimination step as a matrix
- From our example:

$$\begin{bmatrix} 2 & 4 & 2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}$$

- From our algorithm, we want subtract twice the first row from the second row.

$$b = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \Rightarrow b = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

- We can use an elimation matrix to represent this

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow E \begin{bmatrix} 2 & 4 & 2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 1 & -7 \\ -2 & -3 & 7 \end{bmatrix}$$

  - To generate an elimination/elementary matrix, start with the identity matrix and change one of its zeroes to $-\ell$.
- The purpose of E_{ij} is to produce a zero in the (i,j)th position
- Start with A. Apply E's to produce zeros to get a upper triangular U.
- The vector x remains the same - The solution does not change.
- $Ax = b \Rightarrow$ Apply E on both sides $\Rightarrow \underbrace{EA}_{\text{new matrix}}x = Eb$

Matrix Multiplication

- Matrix multiplication is:
  - Associative: $A(BC) = (AB)C$
  - Non-commutative: $AB \neq BA \ \forall A, B$
- We can keep our viewpoint, that for EA, E acts on the columns of A in the same way it would act on a column vector
  - $AB = A[b_1\ b_2\ b_3] = [Ab_1, Ab_2, Ab_3]$
- The (i,j)th entry of the product is the dot product of the ith row of the first matrix and the jth column of the second matrix
- The number of columns of A must equal the number of rows of B and the output has the number of rows of A and the number of columns of B
- Example

red • blue

$$\begin{bmatrix} \boxed{1 \ \ 2 \ \ 4} \\ 7 \ \ -2 \ \ 5 \\ 3 \ \ 3 \ \ -1 \end{bmatrix} \begin{bmatrix} 2 & -4 & -7 \\ 6 & -2 & 5 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 18 & -12 & -1 \\ 7 & -29 & -64 \\ 23 & -17 & -8 \end{bmatrix}$$

The Matrix P_{ij} for a Row Exchange

- To subtract row j from row i, we use E_{ij}
- To exchange/permute rows, we use another matrix, P_{ij}
- To get P_{ij}, we simply exchange the row of the identity matrix

$$P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow P_{23}\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ 2 & 2 & 2 \end{bmatrix}$$

- Multiplying by P_{ij} swaps the ith and jth components of a column vector, so it works on matrices as well.

The Augmented Matrix

- Elimination performs the same row operations to A and to b ($EAx = Eb$), so we simply include b as another column of the coefficient matrix
- Starting from

$$\begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix} \Rightarrow E_{21}\begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ -2 & -3 & 7 & 10 \end{bmatrix}$$

$-\ell = -\frac{4}{2} = -2$

  - $E_{32}E_{31}E_{21}A$ is a triangular matrix
- We can now express elimination as a sequence of prefix multiplications on A. We can also merge operations in a prefix manner.
  - Example

$$P_{32}E_{21} = \text{eliminate entry at (2,1) and exchange row 2 \& 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -\ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -\ell & 1 & 0 \end{bmatrix}$$

Problem Set 2.3

1.

a) $\begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

2.

$$E_{32}E_{21}b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ -35 \end{bmatrix}$$

$$E_{21}E_{32}b = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix}$$

Row 3 feels no effect from row 2

3.

$$A=\begin{bmatrix}1&1&0\\4&6&1\\-2&2&0\end{bmatrix} \Rightarrow E_{21}A=\begin{bmatrix}1&0&0\\-4&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&1&0\\4&6&1\\-2&2&6\end{bmatrix}=\begin{bmatrix}1&1&0\\0&2&1\\-2&2&0\end{bmatrix}\overset{S}{\begin{bmatrix}-&-&-\\-&-&-\\-&-&17\end{bmatrix}}$$

$$E_{31}(E_{21}A)=\begin{bmatrix}1&0&0\\0&1&0\\2&0&1\end{bmatrix}\begin{bmatrix}1&1&0\\0&2&1\\-2&2&0\end{bmatrix}=\begin{bmatrix}1&1&0\\0&2&1\\0&4&0\end{bmatrix}$$

$$E_{32}(E_{31}E_{21}A)=\begin{bmatrix}1&0&0\\0&1&0\\0&-2&1\end{bmatrix}\begin{bmatrix}1&1&0\\0&2&1\\0&4&0\end{bmatrix}=\begin{bmatrix}1&1&0\\0&2&1\\0&0&-2\end{bmatrix}$$

$$E_{32}E_{31}E_{21}=\begin{bmatrix}1&0&0\\0&1&0\\0&-2&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\2&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\-4&1&0\\0&0&1\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&-2&1\end{bmatrix}\begin{bmatrix}1&0&0\\-4&1&0\\2&0&1\end{bmatrix}=\begin{bmatrix}1&0&0\\-4&1&0\\10&-2&1\end{bmatrix}$$

4.

$$E_{32}E_{31}E_{21}b=\begin{bmatrix}1&0&0\\-4&1&0\\10&-2&1\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}1\\-4\\10\end{bmatrix}\Rightarrow\begin{bmatrix}1&1&0\\0&2&1\\0&0&-2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}1\\-4\\10\end{bmatrix}\Rightarrow\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\-5\end{bmatrix}$$

6.

$$\begin{bmatrix}a&b&c\\a&b&c\\a&b&c\end{bmatrix}=A\Rightarrow E_{21}A=\begin{bmatrix}1&0&0\\-1&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}a&b&c\\a&b&c\\a&b&c\end{bmatrix}=\begin{bmatrix}a&b&c\\0&0&0\\a&b&c\end{bmatrix}\Rightarrow E_{31}E_{21}A=\begin{bmatrix}1&0&0\\0&1&0\\-1&0&1\end{bmatrix}\begin{bmatrix}a&b&c\\0&0&0\\a&b&c\end{bmatrix}$$

7.

$$E=\begin{bmatrix}1&0&0\\0&1&0\\-7&0&1\end{bmatrix}\Rightarrow\begin{bmatrix}1&0&0\\0&1&0\\7&0&1\end{bmatrix}=E^{-1}\quad EE^{-1}=\begin{bmatrix}1&0&0\\0&1&0\\-7&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\7&0&1\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

8.

$$M=\begin{bmatrix}a&b\\c&d\end{bmatrix}\quad M^*=\begin{bmatrix}1&0\\-l&1\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}a&b\\-al+c&-bl+d\end{bmatrix}$$

$$\det(M^*)=a(-bl+d)-b(-al+c)$$
$$=-abl+ad+abl-bc$$
$$=ad-bc$$

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

10.

$$\begin{cases}X=2Y\\X+Y=33\end{cases}\Rightarrow\begin{matrix}X-2Y=0\\X+Y=33\end{matrix}$$

11.

$$\begin{bmatrix}1&2&20\\1&1&1\\1&0&1\end{bmatrix}\quad\left[\begin{array}{cc|c}1&-2&0\\1&1&33\end{array}\right]$$

$$\begin{bmatrix}1&0\\-1&1\end{bmatrix}\left[\begin{array}{cc|c}1&-2&0\\1&1&33\end{array}\right]=\left[\begin{array}{cc|c}1&-2&0\\0&3&33\end{array}\right]\Rightarrow\begin{bmatrix}X\\Y\end{bmatrix}=\begin{bmatrix}22\\11\end{bmatrix}$$

17.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 14 \end{array}\right] = A \Rightarrow E_{12}A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 14 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 4 \\ 1 & 3 & 9 & 14 \end{array}\right]$$

$$E_{31}(E_{12}A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 4 \\ 1 & 3 & 9 & 14 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 4 \\ 0 & 2 & 8 & 10 \end{array}\right]$$

$$E_{32}(E_{31}E_{12}A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 4 \\ 0 & 2 & 8 & 10 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 2 & 2 \end{array}\right] \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

19.

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

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

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

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

26.

$$\begin{bmatrix} 1 & 4 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Apply $E_{21} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ to both sides

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$\begin{bmatrix} x + 4y \\ -y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -7 \\ 2 \end{bmatrix}$$

chapter 2.4: Rules for Matrix Operations

- Scaling a Matrix
  - Works the exact same as a vector; "distribute" the constant to all entries
- Adding Matrices
  - Added entry by entry, much like vectors
- Multiplying matrices
  - Refer to chapter 2.3: Matrix Multiplication

The Laws for Matrix Operations

- Addition
  - $A + B = B + A$ (commutative law)
  - $c(A+B) = cA + cB$ (distributive law)
  - $A + (B+C) + (A+B) + C = (A+C) + B$ (associative law)
- Multiplication
  - $C(A+B) = (A+CB$ (distributive prefix)
  - $(A+B)C = AC + BC$ (distributive postfix)
  - $A(BC) = (AB)C$ (associative law)

- $AI = IA$; All square matrices commute with I and cI.
- When $A = B = C =$ square matrix, then its exponents follow the same rules as we would expect

$$A^p = \underbrace{AAA\ldots A}_{p \text{ factors}} \quad (A^p)(A^q) = A^{p+q} \quad (A^p)^q = A^{pq}$$

  - These rules still hold when p or q are zero or negative.

    A^{-1} is the inverse matrix, and $A^{-1}A = A^0 = I$. A will not always be invertible

Block Matrices and Block Multiplication

- Matrices can be cut into blocks (smaller matrices). This often happens naturally
- Example

$$\left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} I & I & I \\ I & I & I \end{bmatrix}$$

- If we have B with the same dimensions and the block sizes match, we can add A+B block at a time.
- For multiplication, block multiplication (blocks x blocks) is allowed when their shapes permit
  - If all submatrix products are defined, block multiplication is allowed.
  - Generally:

    Given an $(m \times p)$ matrix with q row partitions and s column partitions and a $(p \times n)$ matrix with s row partitions and r column partitions that are compatible with the partitions of A, $C = AB$ can be formed blockwise, as an $(m \times n)$ matrix with q row partitions and r column partitions

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix}$$

- Example (special case):
  - Let the blocks of A be its n columns. Let the blocks of B be its n rows

$$A = \begin{bmatrix} - & - & - & \ldots \\ - & - & - & \ldots \\ - & - & - & \ldots \end{bmatrix} \quad AB = \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_n \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1b_1 + \ldots a_nb_n \end{bmatrix}$$

$a_1 \quad a_2 \quad a_3 \ldots$

$$\begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 5 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 2 \\ 8 & 2 \end{bmatrix}$$

- Example (Elimination by Blocks):
  - Suppose the first column of A contains 1, 3, 4:

$$A = \begin{bmatrix} 1 & \cdots \\ 3 & \cdots \\ 4 & \cdots \end{bmatrix}, \text{ then our elimination matrices would be } E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

  - The block idea is to combine the matrices into one, which clear (2,1) and (3,1)

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

  - Suppose A has 4 blocks A, B, C, D we can perform elimination in blocks, using inverse matrices. The $D - CA^{-1}B$ we get as the last block is called Schur complement

$$\left[\begin{array}{c|c} I & 0 \\ \hline -CA^{-1} & I \end{array}\right] \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} A & B \\ \hline 0 & D - CA^{-1}B \end{array}\right]$$

Problem set 2.4

1.

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

a) $BA = \begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 \end{bmatrix}$ b) $AB = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{bmatrix}$ c) $ABD = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}$

6.

$$(A+B)^2 = \begin{bmatrix} 2 & 2 \\ 3 & 0 \end{bmatrix}^2 = \begin{bmatrix} 2 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 4 \\ 6 & 6 \end{bmatrix}$$

$$A^2 + B^2 = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}$$

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

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

$$(A+B)(A+B) = (A+B)A + (A+B)B$$
$$= AA + BA + AB + B^2$$

$$BA + AB = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 2 \\ 3 & 6 \end{bmatrix}$$

11.

c) $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} =$

12.

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

$$\begin{bmatrix} a & 0 \\ c & 0 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \Rightarrow b = c = 0$$

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

$$\begin{bmatrix} 0 & a \\ 0 & c \end{bmatrix} = \begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix} \quad d = a$$

17.

a) $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$ b) $A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$ c) $A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 2 & 1 & \frac{2}{3} \\ 3 & \frac{3}{2} & 1 \end{bmatrix}$

18

a) $A = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix}$ b) $A = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ c) $A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ d) $A = \begin{bmatrix} a & b & c \\ a & b & c \\ a & b & c \end{bmatrix}$

20.

$A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $A^2 = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

$A^3 = (A^2)A = \begin{bmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

25.

$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = a_1\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + a_2\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + a_3\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} a_1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & a_2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & a_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$

26.

$AB = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 3 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 3 & 3 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 4 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 0 \\ 6 & 6 & 0 \\ 6 & 6 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 0 \\ 10 & 14 & 4 \\ 7 & 8 & 1 \end{bmatrix}$

27.

$AB = \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix}\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} x & x & x \end{bmatrix} + \begin{bmatrix} x \\ x \\ 0 \end{bmatrix}\begin{bmatrix} 0 & x & x \end{bmatrix} + \begin{bmatrix} x \\ x \\ x \end{bmatrix}\begin{bmatrix} 0 & 0 & x \end{bmatrix}$

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

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

28.

(1)

$$\begin{bmatrix} & A & \end{bmatrix}\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & Ab_3 & Ab_4 \end{bmatrix}$$

(2)

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\begin{bmatrix} B \end{bmatrix} = a_1B + a_2B$$

32.

$$Ax_1 + Ax_2 + Ax_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$Ax = A\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$Ax_1 \quad Ax_2 \quad Ax_3$

29.

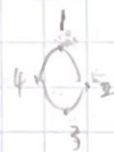
$$E = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \Rightarrow EA = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ -2 & 0 & 1 \\ 8 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix}, \; c = \begin{bmatrix} -2 \\ 8 \end{bmatrix}, \; b = [1, 0]$$

$$D - cb/a = \begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix} - \begin{bmatrix} -2 \\ 8 \end{bmatrix}[1 \; 0]\frac{1}{2} = \begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -2 & 0 \\ 8 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

35.

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



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

Graphs

- Given the adjacency matrix of a graph (above), A, A^k is another matrix whose (i,j)th entry is the count of k long paths from i to j
- Additionally, with words, the (i,j)th entry is the number of words of length $k+1$ starting with the ith letter and ending with the jth.

Chapter 2.5: Inverse matrices

- Suppose A is a square matrix, we look for an inverse matrix A^{-1} such that $AA^{-1} = I$. A^{-1} undoes A on a matrix: $A^{-1}Ax = x$:
  $Ax = b \Rightarrow A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$
- Invertibility
  - An $n \times n$ (square) matrix A is called invertible if there exists an $n \times n$ matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$, where I_n is the $n \times n$ identity matrix
  - A^{-1} does not always exist and A is called singular if it does not exist

- Rules for A^{-1}
  1. The inverse (A^{-1}) exists if and only if elimination produces n pivots (row exchanges allowed)
  2. The matrix A cannot have multiple different inverses. Suppose $BA=I$ and $AC=I$, then $B(AC)=(BA)C \Rightarrow BI=IC \Rightarrow B=C$. A left-inverse (prefix multiplication) and right-inverse (postfix multiplication) must be the same matrix
  3. If A is invertible, the one and only solution to $Ax=b$ is $x=A^{-1}b$
  4. Suppose there is a nonzero vector x such that $Ax=0$, then A cannot have an inverse. No matrix can bring 0 back to x. This implies the columns of A must be linearly independent.

     A is invertible $\Leftrightarrow$ $Ax=0$ can only have the zero solution $A^{-1}0=0$

  5. A matrix is invertible if and only if the determinant is non-zero.
     For a 2x2 matrix, $\det A = ad-bc$. For a 2x2 matrix:

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

  6. A diagonal matrix has an inverse provided no diagonal entries are zero

$$A=\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \Rightarrow A^{-1}=\begin{bmatrix} \frac{1}{d_1} & & \\ & \ddots & \\ & & \frac{1}{d_n} \end{bmatrix}$$

The Inverse of a Product AB

- The sum of a product $A+B$ does not have an inverse, much in the same way that an "inverse" of $a+b=0$ does not make sense.
- The product AB has an inverse if and only if the two factors A and B are themselves invertible (and the same size)
  - Note that A^{-1} and B^{-1} are in the reverse order

$$(AB)^{-1}=B^{-1}A^{-1} \qquad ABB^{-1}A^{-1}=AIA^{-1}=IAA^{-1}=II=I$$

  - To see why the order is reversed, try multiplying $B^{-1}A^{-1}$ and AB (above)
- This rule generalizes to multiple multiplications

$$(ABC)^{-1}=C^{-1}B^{-1}A^{-1}$$

- If we view matrices as transformations upon another, to undo them, we need to apply the inverse transformations in the inverse order.

• For an elimination matrix

eliminates the "-5" from E

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow EE^{-1} = E^{-1}E = I$$

• Example: Suppose:

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \Rightarrow F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}, \text{ then } FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 20 & -4 & 1 \end{bmatrix}, \; E^{-1}F^{-1} = (FE)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

FE contains 20 but its inverse $E^{-1}F^{-1}$ does not.

E subtracts 5 times row 1 from row 2, then F subtracts 4 times the new row 2.

In FE, row 3 is affected by row 1. In the reverse order $E^{-1}F^{-1}$, row 3 feels no effect from row 1.

Calculating A^{-1} by Gauss-Jordan Elimination

- A^{-1} might not be explicitly needed.
  - Elimination goes directly to x
- Elimination is also the way to calculate A^{-1}
- To do this, we use Gauss-Jordan Elimination

1. We start with an augmented matrix $[A \; I]$

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

2. The goal is to get K to reduced row echelon form (get A to I) using elementary row operations E and P.

$$[K \; I] = \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \quad \left(\text{Apply } E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ to K and I}\right)$$

$$= \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \quad \left(\text{Apply } E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \text{ to K and I}\right)$$

We are at row echelon form, which is where Gaussian would stop for back substitution

$$= \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \quad \left(\text{Apply } E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{3}{4} \\ 0 & 0 & 1 \end{bmatrix} \text{ to K and I}\right)$$

$$= \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \quad \left(\text{Apply } E_{12} = \begin{bmatrix} 1 & \frac{2}{3} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ to K and I}\right)$$

$$= \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \xrightarrow{\text{divide each row by pivot}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{3}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right]$$

$$\underbrace{\begin{bmatrix} \frac{3}{4} & \frac{1}{2} & \frac{3}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}}_{K^{-1}}$$

- If A cannot be reduced to I, then A is singular

- Through elementary row operations, we get A to I,
  $\underbrace{E_n \dots E_3 E_2 E_1}_{A^{-1}} A = I$, $\underbrace{E_n \dots E_3 E_2 E_1}_{A^{-1}} I = A^{-1} I = A^{-1}$
  - Through the augmented identity matrix, we keep track of the product of the row operation, which gets us the inverse matrix.
- Terminology
  - Symmetric: A symmetric matrix is a square matrix where $a_{ij} = a_{ji}$; it is symmetric across its main diagonal
  - Band Matrix: A matrix whose non-zero entries are confined to a diagonal band comprising the main diagonals and its surrounding diagonals
  - Tridiagonal: A band matrix whose band with is 3 (the diagonal consist of the main diagonal and a surrounding diagonal of each side)
  - Dense Matrix: A matrix with no zero entries. Generally, the inverse of a band matrix is a dense matrix.
- The product of the pivots of K is $(2)(\frac{3}{2})(\frac{4}{3}) = 4$, which is the determinant of K. The computation of K^{-1} involves division by the determinant, which is why it must be non-zero
- Gauss-Jordan is relatively expensive. We must solve n equations for its n columns
  - To solve $Ax=b$ without A^{-1}, we deal with one column b to find column x.
  - Instead we could simply perform elementary row operations on [A b] and perform back substitution upon reaching U.
  - The cost to solve for the n columns of A^{-1} is only multiplied by 3 compared to Gaussian since we have 3 augmented columns (I) instead of one (b).
  - A^{-1} takes n^3 steps whereas solving for x directly is $\frac{n^3}{3}$

Singular vs Invertible

- A^{-1} exists if and only if A has a full set of n pivots (pivot test).
  - Elimination solves all equations $Ax_i = I_i$. The columns x_i go into A^{-1}. Then $AA^{-1} = I$ and A^{-1} is at least a right-inverse
  - Elimination is really a sequence of multiplications by E, P, and D^{-1}. D^{-1} divides by the pivots.
    $(D^{-1} \dots E \dots P)A = I \Rightarrow (D^{-1} \dots E \dots P) = A^{-1}$. (Left-inverse)
  - The right-inverse equals the left inverse.
- Reasoning about the pivot test (by contradiction)
  1. If A doesn't have n pivots, elimination leads to a zero row.
  2. Those elimination steps are taken by an invertible matrix M. Then a row of MA is 0.
  3. If $AC = I$ is possible, then $MAC = M$. The zero row of MA times C gives a zero row in M.
  4. An invertible matrix M can't have a zero row. Therefore A must have n pivots.

- If L is lower triangular with 1's on the diagonal, so is L^{-1}
- A triangular matrix is invertible if and only if no diagonal entries are zero

$$[L\ I] = \left[\begin{array}{ccc|ccc} 1&0&0&1&0&0\\ 3&1&0&0&1&0\\ 4&5&1&0&0&1\end{array}\right] \Rightarrow E_{21}[L\ I] = \left[\begin{array}{ccc|ccc} 1&0&0&1&0&0\\ 0&1&0&-3&1&0\\ 4&5&1&0&0&1\end{array}\right]$$

$$\Rightarrow E_{31}E_{21}[L\ I] = \left[\begin{array}{ccc|ccc} 1&0&0&1&0&0\\ 0&1&0&-3&1&0\\ 0&5&1&-4&0&1\end{array}\right] \Rightarrow E_{32}E_{31}E_{21}[L\ I] = \begin{bmatrix} 1&0&0\\ 0&1&0\\ 0&-5&1\end{bmatrix}\left[\begin{array}{ccc|ccc} 1&0&0&1&0&0\\ 0&1&0&-3&1&0\\ 0&5&1&-4&0&1\end{array}\right]$$

$$E_{32}E_{31}E_{21}[L\ I] = \left[\begin{array}{ccc|ccc} 1&0&0&1&0&0\\ 0&1&0&-3&1&0\\ 0&0&1&11&-5&1\end{array}\right] = [I\ L^{-1}] \Rightarrow L^{-1} = \begin{bmatrix} 1&0&0\\ -3&1&0\\ 11&-5&1\end{bmatrix}$$

Problem Set 2.5

1.

$$[A\ I] = \left[\begin{array}{cc|cc} 0&3&1&0\\ 4&0&0&1\end{array}\right] \Rightarrow P_{12}[A\ I] = \left[\begin{array}{cc|cc} 4&0&0&1\\ 0&3&1&0\end{array}\right] \Rightarrow D^{-1}P_{12}[A\ I] = \left[\begin{array}{cc|cc} 1&0&0&\frac{1}{4}\\ 0&1&\frac{1}{3}&0\end{array}\right]$$

$$D^{-1}P_{12}[A\ I] = A^{-1}[A\ I] = [A^{-1}A\ \ A^{-1}I] = [I\ A^{-1}] = \left[\begin{array}{cc|cc} 1&0&0&\frac{1}{4}\\ 0&1&\frac{1}{3}&0\end{array}\right] \Rightarrow A^{-1} = \begin{bmatrix} 0&\frac{1}{4}\\ \frac{1}{3}&0\end{bmatrix}$$

$$B = \begin{bmatrix} 2&0\\ 4&2\end{bmatrix} \Rightarrow B^{-1} = \frac{1}{\det B}\begin{bmatrix} 2&0\\ -4&2\end{bmatrix} = \frac{1}{4}\begin{bmatrix} 2&0\\ -4&2\end{bmatrix} = \begin{bmatrix} \frac{1}{2}&0\\ -1&\frac{1}{2}\end{bmatrix}$$

$$[C\ I] = \left[\begin{array}{cc|cc} 3&4&1&0\\ 5&7&0&1\end{array}\right] \Rightarrow E_{21}[C\ I] = \begin{bmatrix} 1&0\\ -\frac{5}{3}&1\end{bmatrix}\left[\begin{array}{cc|cc} 3&4&1&0\\ 5&7&0&1\end{array}\right] = \left[\begin{array}{cc|cc} 3&4&1&0\\ 0&\frac{1}{3}&-\frac{5}{3}&1\end{array}\right]$$

$$E_{12}E_{21}[C\ I] = \begin{bmatrix} 1&-12\\ 0&1\end{bmatrix}\left[\begin{array}{cc|cc} 3&4&1&0\\ 0&\frac{1}{3}&-\frac{5}{3}&1\end{array}\right] = \left[\begin{array}{cc|cc} 3&0&21&-12\\ 0&\frac{1}{3}&-\frac{5}{3}&1\end{array}\right] \Rightarrow D^{-1}E_{12}E_{21}[C\ I] = \begin{bmatrix} \frac{1}{3}&0\\ 0&3\end{bmatrix}\left[\begin{array}{cc|cc} 3&0&21&-12\\ 0&\frac{1}{3}&-\frac{5}{3}&1\end{array}\right]$$

$$D^{-1}E_{12}E_{21}[C\ I] = C^{-1}[C\ I] = [C^{-1}C\ \ C^{-1}I] = [I\ C^{-1}] = \left[\begin{array}{cc|cc} 1&0&7&-4\\ 0&1&-5&3\end{array}\right] \Rightarrow C^{-1} = \begin{bmatrix} 7&-4\\ -5&3\end{bmatrix}$$

2.

$$P = \begin{bmatrix} 0&0&1\\ 0&1&0\\ 1&0&0\end{bmatrix} \Rightarrow P^{-1}P = I \Rightarrow \overbrace{\begin{bmatrix} 0&0&1\\ 0&1&0\\ 1&0&0\end{bmatrix}}^{P^{-1}}\begin{bmatrix} 0&0&1\\ 0&1&0\\ 1&0&0\end{bmatrix} = \begin{bmatrix} 1&0&0\\ 0&1&0\\ 0&0&1\end{bmatrix}$$

3.

$$\begin{bmatrix} 10&20\\ 20&50\end{bmatrix}\begin{bmatrix} x\\ y\end{bmatrix} = \begin{bmatrix} 1\\ 0\end{bmatrix} \Rightarrow \left[\begin{array}{cc|c} 10&20&1\\ 20&50&0\end{array}\right] \Rightarrow E_{21}\left[\begin{array}{cc|c} 10&20&1\\ 20&50&0\end{array}\right] = \begin{bmatrix} 1&0\\ -2&1\end{bmatrix}\left[\begin{array}{cc|c} 10&20&1\\ 20&50&0\end{array}\right] = \left[\begin{array}{cc|c} 10&20&1\\ 0&10&-2\end{array}\right] \Rightarrow \begin{bmatrix} x\\ y\end{bmatrix} = \begin{bmatrix} \frac{1}{2}\\ -\frac{1}{5}\end{bmatrix}$$

4.

$$\left[\begin{array}{cc|cc} 1&2&1&0\\ 3&6&0&1\end{array}\right] = [A\ I] \Rightarrow E_{21}[A\ I] = \begin{bmatrix} 1&0\\ -3&1\end{bmatrix}\left[\begin{array}{cc|cc} 1&2&1&0\\ 3&6&0&1\end{array}\right] = \left[\begin{array}{cc|cc} 1&2&1&0\\ 0&0&-3&1\end{array}\right]$$

6.

A is invertible $\Rightarrow$ for $AB = AC \Rightarrow A^{-1}AB = A^{-1}AC \Rightarrow IB = IC \Rightarrow B = C$

7.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11}+a_{21} & a_{12}+a_{22} & a_{13}+a_{23} \end{bmatrix} \Rightarrow A \text{ does not produce n pivots because elimination needs to a zero row}$$

$\Rightarrow A$ is Singular

a)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11}+a_{21} & a_{12}+a_{22} & a_{13}+a_{23} \end{bmatrix} \Rightarrow Ax = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow E_{31}Ax = E_{31}b$$

b)

$b = (a, b, a+b)$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11}+a_{21} & a_{12}+a_{22} & a_{13}+a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \Rightarrow 0 = -1$$

8.

$$\begin{bmatrix} a_1 & b_1 & a_1+b_1 \\ a_2 & b_2 & a_2+b_2 \\ a_3 & b_3 & a_3+b_3 \end{bmatrix} = A \Rightarrow \text{set } Ax = 0 \Rightarrow \begin{bmatrix} a_1 & b_1 & a_1+b_1 \\ a_2 & b_2 & a_2+b_2 \\ a_3 & b_3 & a_3+b_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x_1 \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + x_2 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + x_3 \begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ a_3+b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\Rightarrow x = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ is a non-zero solution $\Rightarrow A$ is not invertible

9.

$$A \text{ is invertible} \Rightarrow \overbrace{P_{12}A}^{\text{both invertible}} = B \Rightarrow B^{-1} = (P_{12}A)^{-1} \Rightarrow B^{-1} = A^{-1}P_{12}^{-1}$$

10.

$$[B\ I] = \left[\begin{array}{cccc|cccc} 3 & 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 6 & 5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 7 & 6 & 0 & 0 & 0 & 1 \end{array}\right] \Rightarrow E_{21}[B\ I] = \left[\begin{array}{cccc|cccc} 3 & 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & -\frac{4}{3} & 1 & 0 & 0 \\ 0 & 0 & 6 & 5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 7 & 6 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\Rightarrow E_{43}E_{21}[B\ I] = \left[\begin{array}{cccc|cccc} 3 & 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & -\frac{4}{3} & 1 & 0 & 0 \\ 0 & 0 & 6 & 5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & -\frac{7}{6} & 1 \end{array}\right] \Rightarrow E_{12}E_{43}E_{21}[B\ I] = \left[\begin{array}{cccc|cccc} 3 & 0 & 0 & 0 & 9 & -6 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & -\frac{4}{3} & 1 & 0 & 0 \\ 0 & 0 & 6 & 5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & -\frac{7}{6} & 1 \end{array}\right]$$

$$\Rightarrow E_{34}E_{12}E_{43}E_{21}[B\ I] = \left[\begin{array}{cccc|cccc} 3 & 0 & 0 & 0 & 9 & -6 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & -\frac{4}{3} & 1 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 & 30 & -30 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & -\frac{7}{6} & 1 \end{array}\right] \Rightarrow D^{-1}E_{34}E_{12}E_{43}E_{21}[B\ I] = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 3 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 & -4 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 6 & -5 \\ 0 & 0 & 0 & 1 & 0 & 0 & -7 & 6 \end{array}\right]$$

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

11.

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

12.

$$C=AB \Rightarrow CB^{-1}=ABB^{-1} \Rightarrow CB^{-1}=AI=A \Rightarrow A^{-1}=BC^{-1}$$

13.

$$M=ABC \Rightarrow A^{-1}M=A^{-1}ABC \Rightarrow A^{-1}M=BC \Rightarrow A^{-1}MC^{-1}=BCC^{-1} \Rightarrow A^{-1}MC^{-1}=B \Rightarrow B^{-1}=CM^{-1}A$$

14.

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

15

$$\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}d&-b\\-c&a\end{bmatrix}=\begin{bmatrix}ad-bc&0\\0&-bc+ad\end{bmatrix}$$

23.

$$[A\ I]=\left[\begin{array}{ccc|ccc}2&1&0&1&0&0\\1&2&1&0&1&0\\0&1&2&0&0&1\end{array}\right] \Rightarrow E_{21}[A\ I]=\left[\begin{array}{ccc|ccc}2&1&0&1&0&0\\0&\frac{3}{2}&1&-\frac{1}{2}&1&0\\0&1&2&0&0&1\end{array}\right]$$

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

$$\Rightarrow E_{32}E_{21}[A\ I]=\left[\begin{array}{ccc|ccc}2&1&0&1&0&0\\0&\frac{3}{2}&1&-\frac{1}{2}&1&0\\0&0&\frac{4}{3}&\frac{1}{3}&-\frac{2}{3}&1\end{array}\right] \Rightarrow E_{23}E_{32}E_{21}[A\ I]=\left[\begin{array}{ccc|ccc}2&1&0&1&0&0\\0&\frac{3}{2}&0&-\frac{3}{4}&\frac{3}{2}&-\frac{3}{4}\\0&0&\frac{4}{3}&\frac{1}{3}&-\frac{2}{3}&1\end{array}\right]$$

$$\Rightarrow E_{12}E_{23}E_{32}E_{21}[A\ I]=\left[\begin{array}{ccc|ccc}2&0&0&\frac{3}{2}&-1&\frac{1}{2}\\0&\frac{3}{2}&0&-\frac{3}{4}&\frac{3}{2}&-\frac{3}{4}\\0&0&\frac{4}{3}&\frac{1}{3}&-\frac{2}{3}&1\end{array}\right] \Rightarrow D^{-1}E_{12}E_{23}E_{32}E_{21}[A\ I]=\left[\begin{array}{ccc|ccc}1&0&0&\frac{3}{4}&-\frac{1}{2}&\frac{1}{4}\\6&1&0&-\frac{1}{2}&1&-\frac{1}{2}\\0&0&1&\frac{1}{4}&-\frac{1}{2}&\frac{3}{4}\end{array}\right]$$

24.

a) True b) False $\begin{bmatrix}1&1\\1&1\end{bmatrix}$ c) True

30.

$$A=\begin{bmatrix}2&c&c\\c&c&c\\8&7&c\end{bmatrix} \Rightarrow \det A = 2c^2+8c^2+7c^2-8c^2-c^3-14c$$
$$=-c^3+9c^2-14c$$
$$=-c(c^2-9c+14c)$$
$$=-c(c-2)(c-7) \Rightarrow c\neq 0,2,7$$

Chapter 2.6: Elimination = Factorization: $A=LU$

- Lower-upper (LU) factorization/decomposition factors a matrix as the product of a lower triangular and upper triangular matrix.
- Performing Gaussian elimination takes A to U. If we take the inverse of the product of the elementary row operations used in elimination, we get a lower triangular matrix L. Hence $A=LU$
- For example:

$$E_{21}A=\begin{bmatrix}1&0\\-3&1\end{bmatrix}\begin{bmatrix}2&1\\6&8\end{bmatrix}=\begin{bmatrix}2&1\\0&5\end{bmatrix}=U \text{ (from A to U)}$$

$$\underbrace{E_{21}^{-1}}_{L}\overbrace{E_{21}A}^{U}=E_{21}^{-1}U=\begin{bmatrix}1&0\\3&1\end{bmatrix}\begin{bmatrix}2&1\\0&5\end{bmatrix}=\begin{bmatrix}2&1\\6&8\end{bmatrix}=A \text{ (from U to A)}$$

Hence:

$$A=LU \Rightarrow \begin{bmatrix}2&1\\6&8\end{bmatrix}=\begin{bmatrix}1&0\\3&1\end{bmatrix}\begin{bmatrix}2&1\\0&5\end{bmatrix}$$

- L will be the inverse of all elementary row operations during elimination

$$E_n \dots E_3 E_2 E_1 A = U \Rightarrow A = \underbrace{E_1^{-1} E_2^{-1} E_3^{-1} \cdots E_n^{-1}}_{L} U = \underbrace{(E_n \dots E_3 E_2 E_1)^{-1}}_{L} U$$

Explanations and Examples

- Every inverse matrix E^{-1} is lower triangular. Its off-diagonal entry is l_{ij} to offset the subtraction produced by $-l_{ij}$. The main diagonals of E and E^{-1} have 1's.
- Each multiplier l_{ij} goes directly into its ij position unchanged
- Example 1:

$$A = LU \Rightarrow \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

- Example 2

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

- When can we predict zeroes in L and U?
  - When a row of A starts with zeroes, so does that row of L
  - When a column of A starts with zeroes, so does that column of U.
  - If a row starts with zero, an elimination step is not needed. If a column starts with zeroes, they survive into U
- The pivot rows are the same rows as U. When computing the third row of, we subtract multiples of earlier rows of U (not the rows of A)

  Row 3 of U = (Row 3 of A) $- l_{31}$(Row 1 of U) $- l_{32}$(Row 2 of U)

  Rewrite the equation:

  (Row 3 of A) $= l_{31}$(Row 1 of U) $+ l_{32}$(Row 2 of U) + 1(Row 3 of U)

  $$= [\, l_{31} \;\; l_{32} \;\; 1 \,] U$$

  Then (Row 3 of A) = (row 3 of L)U $\Rightarrow A = LU$

- The LU factorization is unsymmetrical because U has the pivots on the main diagonal whereas L has 1's. To fix this, divide U by a diagonal matrix D that contains the pivots

  Split U into $\begin{bmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \\ & & & d_n \end{bmatrix} \begin{bmatrix} 1 & \frac{u_{12}}{d_1} & \frac{u_{13}}{d_1} & \\ & 1 & \frac{u_{23}}{d_2} & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$

  Now we write A as the product of 3 matrices: A = LDU

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 8 \\ 0 & 5 \end{bmatrix} \text{ further splits into } \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$$

One Square System = Two Triangular Systems

- The matrix L contains a record of gaussian elimination. It holds the multipliers for the pivot rows before subtracting them from lower rows
- We need L as soon as there is a right side b. L and U are determined entirely by the left side (the matrix A). On the right side we use U^{-1} and L^{-1}.
  1. Factor into L and U via elimination on A.
  2. Solve via forward substitution on b using L then back substitution for x using U).
     - First we apply forward elimination to the right side using the multipliers stored in L. This changes b to a new right side c. We are really solving $Lc=b \Rightarrow c=L^{-1}b$
     - Then we solve $Ux=c$ via back substitution.
     - The original system $Ax=b$ is factored into 2 triangular systems

Solve $Lc=b$ and then solve $Ux=c$

L^{-1} is the inverse of the elementary row operations, so L^{-1} is applying the same operations to the right side.

- Example:

$u+2v=5$, $4u+9v=21$ "$Ax=b$" becomes $u+2v=5$, $v=1$ "$Ux=c$" $\quad U=\begin{bmatrix}1&2\\0&1\end{bmatrix} \quad L=E_{21}^{-1}=\begin{bmatrix}1&0\\4&1\end{bmatrix}$ $\quad L^{-1}=E_{21}$

$Lc=b \Rightarrow$ the lower triangular system $\begin{bmatrix}1&0\\4&1\end{bmatrix}\begin{bmatrix}c\end{bmatrix}=\begin{bmatrix}b\end{bmatrix}=\begin{bmatrix}5\\21\end{bmatrix} \Rightarrow c=L^{-1}B=\begin{bmatrix}1&0\\-4&1\end{bmatrix}\begin{bmatrix}5\\21\end{bmatrix}=\begin{bmatrix}5\\1\end{bmatrix}$

$Ux=c \Rightarrow$ the upper triangular system $\begin{bmatrix}1&2\\0&1\end{bmatrix}\begin{bmatrix}x\end{bmatrix}=\begin{bmatrix}c\end{bmatrix}=\begin{bmatrix}5\\1\end{bmatrix} \Rightarrow x=U^{-1}c=\begin{bmatrix}1&-2\\0&1\end{bmatrix}\begin{bmatrix}5\\1\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}$

The cost of Elimination

- Elimination on A requires about $\frac{1}{3}n^3$ multiplications and $\frac{1}{3}n^3$ subtractions
- To solve the right side (b to c to x), it takes n^2 multiplications and n^2 subtractions
- For a band matrix of width w, the cost of factoring (elimination) takes nw^2 and solving takes $2nw$ operations.

Problem Set 2.6

1.

$L=E_{21}^{-1}=\begin{bmatrix}1&0\\-1&1\end{bmatrix}^{-1}=\begin{bmatrix}1&0\\1&1\end{bmatrix} \Rightarrow \begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}x=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} \Rightarrow \begin{bmatrix}1&1\\1&2\end{bmatrix}x=\begin{bmatrix}5\\7\end{bmatrix}$

3.

$U=\begin{bmatrix}1&1&1\\0&1&2\\0&0&1\end{bmatrix} \Rightarrow L=\begin{bmatrix}1&0&0\\1&1&0\\1&2&1\end{bmatrix} \Rightarrow Lc=b \Rightarrow \begin{bmatrix}1&0&0\\1&1&0\\1&2&1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}5\\7\\11\end{bmatrix} \Rightarrow \begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}5\\2\\2\end{bmatrix}$

$\Rightarrow Ux=c \Rightarrow \begin{bmatrix}1&1&1\\&1&2\\&&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}5\\2\\2\end{bmatrix} \Rightarrow \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}5\\-1\\1\end{bmatrix}$

5.

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix} \Rightarrow E_{31}A = \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}}^{E} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix} = \underbrace{\overbrace{\begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 0 & 0 & 5 \end{bmatrix}}^{REF}}_{U} \quad L = E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

$$A = LU \Rightarrow \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

6.

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \Rightarrow L = E_{21}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

7.

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix} \Rightarrow E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 4 & 5 \end{bmatrix} \Rightarrow E_{31}E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 4 & 5 \end{bmatrix}$$

$$\Rightarrow E_{31}E_{21}A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 4 & 2 \end{bmatrix} \Rightarrow E_{32}E_{31}E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = U$$

$$L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ 3 & 2 & 1 \end{bmatrix} \Rightarrow A = LU = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix}$$

8

$$E_{32}E_{31}E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -b & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ac-b & -c & 1 \end{bmatrix}$$

11.

$$A = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix} = U \Rightarrow L = I \Rightarrow D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix} \Rightarrow U = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

12.

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} \Rightarrow L = E_{21}^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \Rightarrow U = \begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}$$

$$A = LDU \Rightarrow \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \Rightarrow U \text{ and } L \text{ are symmetrical}$$

13.

$$A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix}$$

15.

$$Lc=b \Rightarrow \begin{bmatrix}1&0\\4&1\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix}=\begin{bmatrix}2\\11\end{bmatrix} \Rightarrow \begin{bmatrix}c_1\\c_2\end{bmatrix}=\begin{bmatrix}2\\3\end{bmatrix} \quad \begin{bmatrix}1&0\\4&1\end{bmatrix}\begin{bmatrix}2&4\\0&1\end{bmatrix}=\begin{bmatrix}2&4\\8&17\end{bmatrix}$$

$$Ux=c \Rightarrow \begin{bmatrix}2&4\\0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}2\\3\end{bmatrix} \Rightarrow \begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-5\\3\end{bmatrix} \quad \begin{bmatrix}2&4\\8&17\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}2\\11\end{bmatrix}=$$

19.

$$A=\begin{bmatrix}1&1&0\\1&2&1\\0&1&2\end{bmatrix} \Rightarrow E_{21}A=\begin{bmatrix}1&0&0\\-1&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&1&0\\1&2&1\\0&1&2\end{bmatrix}=\begin{bmatrix}1&1&0\\0&1&1\\0&1&2\end{bmatrix} \Rightarrow E_{32}E_{21}A=\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}1&1&0\\0&1&1\\0&1&2\end{bmatrix}$$

$$E_{32}E_{21}A=\begin{bmatrix}1&1&0\\0&1&1\\0&0&1\end{bmatrix}=U \Rightarrow L=E_{21}^{-1}E_{32}^{-1}=\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix}=$$

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

Chapter 2.7: Transposes and Permutations

- The transpose of A, denoted by A^T. The columns of A^T are the rows of A.
  - When A is an $m \times n$ matrix, its transpose is $n \times m$.

    If $A=\begin{bmatrix}1&2&3\\0&0&4\end{bmatrix}$ then $A^T=\begin{bmatrix}1&0\\2&0\\3&4\end{bmatrix}$

  - The matrix "flips over" its main diagonal. The entry in row i, column j of A^T comes from row j, column i of the original A

    $(A^T)_{ij} = A_{ji}$

- The transpose of a lower triangular matrix is upper triangular.
- The transpose of A^T is A. That is $(A^T)^T = A$.
- Rules
  - Sum: $(A+B)^T = A^T + B^T$
  - Product: $(AB)^T = B^T A^T$
  - Inverse: $(A^{-1})^T = (A^T)^{-1}$
- Proof of the product identity

  Start with $(Ax)^T = x^T A^T$, where x is a column vector

  Ax combines the columns of A while $x^T A^T$ combines the rows of A^T. This combination is the same for both cases

$$\begin{bmatrix}a_1&a_2&a_3\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_1a_1+x_2a_2+x_3a_3=\begin{bmatrix}x_1&x_2&x_3\end{bmatrix}\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}=x_1a_1+x_2a_2+x_3a_3$$

  Generalizing to AB:

  $(AB)^T=\begin{bmatrix}Ab_1&Ab_2&\cdots\end{bmatrix}^T$ gives $\begin{bmatrix}b_1^TA^T\\ \vdots\end{bmatrix}$ which is B^TA^T.

- The reverse order rule extends to multiple factors

If $A = LDU$ then $A^T = U^T D^T L^T$

- Then apply this product rule to $A^{-1}A = I$,
  $(A^{-1}A)^T = I^T \Rightarrow A^T (A^{-1})^T = I \Rightarrow$ <u>A^T and $(A^{-1})^T$ are inverses</u>
  - We can invert the transpose or transpose the inverse.
  - <u>A^T is invertible if and only if A is invertible.</u>

The meaning of Inner Products

- Instead of using dot notation, we can use matrices to notate dot products
- Given 2 vectors of size n (x and y):
  - $x^T y \Rightarrow (1 \times n)(n \times 1) =$ dot/inner product (scalar)
  - $x y^T \Rightarrow (n \times 1)(1 \times n) =$ outer product (matrix)
    - The outer product is a matrix whose (i,j)th entry is the product of the ith entry of the first vector and the jth entry of the second.
    - Given

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$u \otimes v = uv^T = \begin{bmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_m v_1 & u_m v_2 & \cdots & u_m v_n \end{bmatrix}$$

    - The dot product is the <u>trace</u>, or sum of the main diagonal of the outer product.
    - The outer product is <u>not commutative</u>.

- Thus we may also define the transpose as the matrix that makes these two inner products equal for every x and y.
  $(Ax)^T y = x^T (A^T y) \Rightarrow$ Inner product of Ax and y = Inner product of x and $A^T y$

- Changing the difference matrix to a derivative $A = \frac{d}{dt}$, its transpose is $\left(\frac{dx}{dt}, y\right) = \left(x, -\frac{dy}{dt}\right)$. Then the inner product changes to an integral
  $$x^T y = (x,y) = \int_{-\infty}^{\infty} x(t) y(t)\, dt$$
  Transpose rule $(Ax)^T y = x^T (A^T y)$:
  $$\int_{-\infty}^{\infty} \frac{dx}{dt} y(t)\, dt = \int_{-\infty}^{\infty} x(t) \left(-\frac{dy}{dt}\right) dt$$ Shows A^T (integration by parts)
- The "transpose" of the derivative is

Symmetric Matrices

- A symmetric matrix is a matrix equal to its transpose: $A = A^T$ or $a_{ji} = a_{ij}$
- The inverse of a symmetric matrix is also symmetric when it is invertible.
  $(A^{-1})^T = (A^T)^{-1} = A^{-1}$
- We produce symmetric matrices by <u>multiplying any matrix R by R^T</u>

Symmetric Products R^TR and RR^T and LDL^T

- Choose any matrix R, probably rectangular. Then the product R^TR is automatically a square symmetric matrix.
  - The (i,j)th entry of R^TR is the dot product of the ith row of R^T (the ith column of R) with the jth column of R. The (j,i)th entry is the same dot product (jth column $\cdot$ ith column).
  - RR^T is also symmetric, although it is a different matrix.

$$R=\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \Rightarrow R^T=\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \Rightarrow RR^T=\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } R^TR=\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

  - R^TR is n by n, while RR^T is m by m. Even if $m=n$, it is not very likely $R^TR=RR^T$
- Symmetric matrices in elimination:
  - $A=A^T$ makes elimination faster
  - The upper triangular U is likely not symmetrical. The symmetry is in the triple product $A=LDU$
  - The diagonal matrix D divides the pivots in U, so L and U both have 1's on their main diagonal.
  - When A is symmetric, $A=LDU$ becomes $A=LDL^T$
  - If $A=A^T$ is factored into LDU with no row exchanges, then U is exactly L^T.

$$\begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}}_{L}\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}\underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}}_{L^T}$$

  - Note the transpose of $LDL^T \Rightarrow (LDL^T)^T = (L^T)^T D^T L^T = LDL^T$.
  - This reduces the amount of operations from $\frac{n^3}{3}$ to $\frac{n^3}{6}$, and also reduces storage needed.

Permutation Matrices

- The transpose plays a special role for a permutation matrix. The matrix P has a single 1 in every row and column. Then P^T is also a permutation matrix – maybe the same or maybe different.
- Any product P_1P_2 is also a permutation matrix
- A permutation matrix P has the rows of I in any order.
- For 3x3 matrices we have:

$$I=\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \quad P_{21}=\begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \quad P_{31}=\begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \quad P_{32}=\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}$$

$$P_{32}P_{21}=\begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix} \quad P_{21}P_{32}=\begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}$$

- Every individual P is its own inverse.

- There are $n!$ permutation matrices of order n.
- P^{-1} is also a permutation matrix.
- P^{-1} is always the same as P^{T}
  - $PP^{T} = P^{T}P = I$

The $PA = LU$ Factorization with row Exchanges

- $A = LU$ is awesome, but doesn't always work i.e. when we need row exchanges,
  Then $A = (E^{-1} \dots P^{-1} \dots E^{-1} \dots P^{-1} \dots)U$
- Every exchange is carried out by a P_{ij} and inverted by that P_{ij}.
  We compress all those exchanges into a single Permutation matrix P
- Now the question is where to collect the P_{ij}'s.
  - We can either do all the exchanges before or after the elimination.
  - Exchanging before gives $PA = LU$
  - Exchanging after gives a permutation matrix P_1 in between ($A = L_1 P_1 U_1$)
    This is because the pivot rows are in a strange order after elimination, so P_1 puts U_1 in the correct position
- If $P = I$ (no exchanges needed), then $A = LU$

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 7 & 9 \end{bmatrix} \to PA = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 7 & 9 \end{bmatrix} \to \underset{\ell_{31} = 2}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 3 & 7 \end{bmatrix}} \to \underset{\ell_{32} = 3}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}}$$

Hence

$$PA = LU \Rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 7 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

Problem Set 2.7

1.

$$A = \begin{bmatrix} 1 & 0 \\ 9 & 3 \end{bmatrix} \Rightarrow A^{T} = \begin{bmatrix} 1 & 9 \\ 0 & 3 \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} 1 & -3 \\ 0 & \frac{1}{3} \end{bmatrix}$$

2.

$(AB)^{T}$

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}^{T} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$$

$$B^{T}A^{T} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$$

5.
a)

$$x^{T}Ay = \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 5 \end{bmatrix} = 5$$

7.

$A = A^{T}$, $B = B^{T}$

$(AB)^{T} = B^{T}A^{T} = BA$ False

a

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

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

11

$$A=\begin{bmatrix}0&0&6\\1&2&3\\0&4&5\end{bmatrix} \Rightarrow PA=\begin{bmatrix}&1&\\&&1\\1&&\end{bmatrix}\begin{bmatrix}0&0&6\\1&2&3\\0&4&5\end{bmatrix}=\begin{bmatrix}1&2&3\\0&4&5\\0&0&6\end{bmatrix}$$

$$P_1AP_2=\begin{bmatrix}1&&\\&&1\\&1&\end{bmatrix}\begin{bmatrix}0&0&6\\1&2&3\\0&4&5\end{bmatrix}\begin{bmatrix}&&1\\&1&\\1&&\end{bmatrix}=\begin{bmatrix}1&&\\&&1\\&1&\end{bmatrix}\begin{bmatrix}6&0&0\\3&2&1\\5&4&0\end{bmatrix}=\begin{bmatrix}6&0&0\\5&4&0\\3&2&1\end{bmatrix}$$

16.

a) $(A^2-B^2)^T=(A^2)^T-(B^2)^T=.$

b $[(A+B)(A-B)]^T=(A-B)^T(A+B)^T=(A^T-B^T)(A^T+B^T)=(A-B)(A+B)$

17.

a) $\begin{bmatrix}1&1\\1&1\end{bmatrix}$ b) $\begin{bmatrix}0&1\\1&0\end{bmatrix}$

27.

$$A=\begin{bmatrix}0&1&1\\1&0&1\\2&3&4\end{bmatrix} \Rightarrow P=\begin{bmatrix}&1&\\1&&\\&&1\end{bmatrix} \Rightarrow PA=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&1&1\\1&0&1\\2&3&4\end{bmatrix}=\begin{bmatrix}1&0&1\\0&1&1\\2&3&4\end{bmatrix}$$

Get PA to U $\Rightarrow E_{31}PA=\begin{bmatrix}1&0&0\\0&1&0\\-2&0&1\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&1\\2&3&4\end{bmatrix}=\begin{bmatrix}1&0&1\\0&1&1\\0&3&2\end{bmatrix}$

$$E_{32}E_{31}PA=\begin{bmatrix}1&0&1\\0&1&1\\0&-3&2\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&1\\0&3&2\end{bmatrix}=\begin{bmatrix}1&0&1\\0&1&1\\0&0&-1\end{bmatrix}=U$$

$$L=(E_{32}E_{31})^{-1}=E_{31}^{-1}E_{32}^{-1}=\begin{bmatrix}1&0&0\\0&1&0\\2&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&3&1\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\2&3&1\end{bmatrix}$$

$$PA=LU\Rightarrow\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&1&1\\1&0&1\\2&3&4\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\2&3&1\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&1\\0&0&-1\end{bmatrix}$$

$$A=\begin{bmatrix}1&2&0\\2&4&1\\1&1&1\end{bmatrix} \Rightarrow P=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix} \Rightarrow PA=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1&2&0\\2&4&1\\1&1&1\end{bmatrix}=\begin{bmatrix}1&1&1\\1&2&0\\2&4&1\end{bmatrix}$$

"$R_1 \longleftrightarrow R_3$"

Get PA to U! $E_{21}PA=\begin{bmatrix}1&0&0\\-1&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&1&1\\1&2&0\\2&4&1\end{bmatrix}=\begin{bmatrix}1&1&1\\0&1&-1\\2&4&1\end{bmatrix}$

$$E_{31}E_{21}PA=\begin{bmatrix}1&0&0\\0&1&0\\-2&0&1\end{bmatrix}\begin{bmatrix}1&1&1\\0&1&-1\\2&4&1\end{bmatrix}=\begin{bmatrix}1&1&1\\0&1&-1\\0&2&-1\end{bmatrix} \Rightarrow E_{32}E_{31}E_{21}PA=\begin{bmatrix}1&0&0\\0&1&0\\0&-2&1\end{bmatrix}\begin{bmatrix}1&1&1\\0&1&-1\\0&2&-1\end{bmatrix}=\begin{bmatrix}1&1&1\\0&1&-1\\0&0&1\end{bmatrix}$$

$$L=(E_{32}E_{31}E_{21})^{-1}=E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}=\begin{bmatrix}1&0&0\\1&1&0\\2&2&1\end{bmatrix} \Rightarrow$$

$$PA=LU \Rightarrow \begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1&2&0\\2&4&1\\1&1&1\end{bmatrix}=\begin{bmatrix}1&0&0\\1&1&0\\2&2&1\end{bmatrix}\begin{bmatrix}1&1&1\\0&1&-1\\0&0&1\end{bmatrix}$$

24.

$$A=\begin{bmatrix}0&1&2\\0&3&8\\2&1&1\end{bmatrix} \Rightarrow E_{12}A=\begin{bmatrix}1&0&0\\-3&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&1&2\\0&3&8\\2&1&1\end{bmatrix}=\begin{bmatrix}0&1&2\\0&0&2\\2&1&1\end{bmatrix} \quad P=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix} \Rightarrow P^{-1}=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}$$

$$PE_{12}A=\begin{bmatrix}2&1&1\\0&1&2\\0&0&2\end{bmatrix} \Rightarrow L=E_{12}^{-1}=\begin{bmatrix}1&0&0\\3&1&0\\0&0&1\end{bmatrix} \Rightarrow PL^{-1}A=U \Rightarrow A=LPA$$

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

$$\begin{bmatrix}x\\y\\z\end{bmatrix} \Rightarrow \begin{bmatrix} & & 1\\1& & \\ &1& \end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}z\\x\\y\end{bmatrix}$$

$\theta=\cos^{-1}(\;=128$

31

$$\begin{bmatrix}1&50\\40&1000\\2&50\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}x_1+50x_2\\40x_1+1000x_2\\2x_1+50x_2\end{bmatrix}=A$$

Sub $(x_1,x_2)=(1,1)$

$$A=\begin{bmatrix}51\\1040\\52\end{bmatrix} \Rightarrow A^Ty=\begin{bmatrix}51&1040&52\end{bmatrix}\begin{bmatrix}200\\3\\3000\end{bmatrix}$$

$$=194820$$

Chapter 3: Vector Spaces and Subspaces

chapter 3.1: Spaces of Vectors

- A vector space is a set of whose elements, often called vectors, may be added together and scaled. Within this space, the operations of vector addition and scalar multiplication must satisfy certain axioms*
- The vector spaces $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3, \dots \mathbb{R}^n$, $n \in \mathbb{N}$ contain all column vectors with n components. The components of said vectors are real numbers, hence the $\mathbb{R}$. Complex vector spaces are denoted as $\mathbb{C}, \mathbb{C}^2, \mathbb{C}^3, \dots, \mathbb{C}^n$
- The vector space $\mathbb{R}^2$ is represented by an xy plane. Each vector gives the x and y coordinates of a point on the plane $v = (x, y)$
- Similarly $\mathbb{R}^3$ corresponds to (x, y, z) in three dimensional space and $\mathbb{R}^1$ is a line.
- Above $\mathbb{R}^3$, it is hard to visualize spaces geometrically by algebraically everything is the same.
- From the two operations (vector addition and scalar multiplication), we produce linear combinations.
  - Important, all the operations go on "inside the vector space" - The resultant vector stays within the original vector space.
- *The 8 axioms for a vector space V.
  1. Associativity of vector addition: $u + (v + w) = (u + v) + w$
  2. Commutivity of vector addition: $u + v = v + u$
  3. Identity element of vector addition: $\exists 0 \in V$ s.t. $v + 0 = v \ \forall v \in V$
  4. Inverse elements of vector addition: $\forall v \in V, \exists -v \in V$ s.t. $v + (-v) = 0$
  5. Compatability of scalar multiplication with field multiplication: $(ab)\mathbf{v} = a(b\mathbf{v})$
  6. Identity element of scalar multiplication: $1v = v$, where 1 is the multiplicative identity
  7. Distributivity of scalar multiplication with respect to vector addition. $a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
  8. Distributivity of scalar multiplication with respect to field addition. $(a+b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$
- Other vector spaces:
  - M: The vector space of all 2×2 matrices
  - F: The vector space of all real functions $f(x)$
  - Z: The vector space containing only the zero vector.
- The function space F is infinite-dimensional. A smaller function space is P or P_n, containing all polynomials $a_0 + a_1 x + \dots + a_n x^n$ of degree n.
- The space Z is zero-dimensional, containing just the zero vector
- Every space needs and has its own zero vector: $(0, 0)$, $(0, 0, 0)$, the zero function, the zero matrix, etc

Subspaces

- A subspace is simply a subset of the vectors in a vector space and closed under its two operations
- For example, a plane through the origin in 3D is a subspace of $\mathbb{R}^3$.
    - Adding or scaling vectors in the plane, stay in the plane
    - The vectors in the plane are still in $\mathbb{R}^3$ (they have 3 components)
- A subspace of a vector space is a set of vectors (including 0) that is closed under scaler multiplication and vector addition. If v and w are vectors in the subspace and c is any scalar, then:
    - $v+w$ is in the subspace
    - cv is in the subspace.
- A consequence is that all linear combinations remain in the subspace
- All subspaces must contain the zero vectors
    - If we choose $c=0$, then $cv=0$, which means the zero vector must be contained.
- Lines and planes through the origin are subspaces.
- The entire space and the zero vector space are trivial subspaces.
- A subspace containing v and w must contain all linear combinations $cv+dw$

The column space of A.

- The most important subspaces are tied directly to a matrix A. We are trying to solve $Ax=b$. If A is not invertible, the system is solvable for some b and not solvable for other b. We want to describe the possible b for which $Ax=b$ is solvable. These b's form the column space of A. Also called the image.
- Ax is a linear combination of the columns of A. To get every possible b, we need to use every possible x, ie. every possible linear combinations of the column vectors of A.
- The column space is denoted as $C(A)$.
- The system $Ax=b$ is solvable if and only if b is in the column space of A.
- Suppose A is an m by n matrix. It's columns have m components, so they belong to $\mathbb{R}^m$. The column space of A is a subspace of $\mathbb{R}^m$.
- Instead, we could start with an arbitrary set of vectors S in a vector space V. To get a subspace SS of V, we take all combinations of the vectors in that set. SS will always be the smallest subspace containing S. SS is the span of S, containing all combinations of vectors in S.
- For example:

$$I=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow C(I)=\mathbb{R}^2 \qquad A=\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow C(A)=\mathbb{R}^1$$

Problem Set 3.1

10:

a) $(b_1, b_2, b_3) + (a_1, a_2, a_3)$

$(b_1, b_2, b_3) + ($

$-1) \quad 1 \quad + (1 \; 1 \; 2) = (-1 \; 3, -1)$

a) T

b) F

c)

13.

$P_0: x + y - 2z = 0$

take $v = (1, 1, 1)$,

$w = (2, 2, 2)$

$v + w = (3, 3, 3)$

$(3) + (3) - 2(3) = 0$

$0 = 0$ ✓

14

a) $\mathbb{R}^2$ itself, Z, and lines through the origin

b) D itself, Z, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, ... $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

15.

a) line, plane

b) point, line

a.

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, ad-bc $\neq 0$, If we apply $c = 0$, $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is singular

b)

$\underbrace{\begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}_{\text{singular}} = \begin{bmatrix} 4 & 7 \\ 2 & 3 \end{bmatrix} \rightarrow$ not singular ($\det \neq 0$)

19.

A: line, B: plane, C: line

20.

a) b must be a scalar multiple of $(1, 2, -1)$

b)
$$\begin{matrix} 1 & 2 & -1 & 1 & 2 & -4 \\ 4 & 9 & -4 & 4 & 9 & -4 \end{matrix}$$
$$\begin{matrix} -8 & -4 & 9 \\ -9 & -4 & 8 \\ \hline \end{matrix}$$

$\vec{n} = (1, 0, 1)$

b must be on the plane

$x + z = 0$

21.

A and C have the same column space

23.

unless it is linearly dependent.

It means the columns of A and b are linearly dependent.

Therefore b can be expressed as a linear combination of the columns of A.

28.

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

29. $\mathbb{R}^7$

30.

because S, T are closed

$S + T$:

Addition $(s_1 + t_1) + (s_2 + t_2) = \overbrace{(s_1 + s_2)}^{S} + \overbrace{(t_1 + t_2)}^{T}$

Scalar multiplication: $c(s_1 + t_1) = \underbrace{cs_1}_{S} + \underbrace{ct_1}_{T}$

Chapter 3.2 : The Nullspace of A : Solving $Ax = 0$

- The nullspace of A, denoted as $N(A)$, consists of all solutions of $Ax = 0$, where x is in $\mathbb{R}^n$
  - For invertible matrices, the only solution is $x = 0$.
  - For singular matrices, there exist nonzero solutions to $Ax = 0$
- The nullspace is a subspace of $\mathbb{R}^n$
  - suppose x and y are in the nullspace ($Ax = 0$ and $Ay = 0$)
  - Then $Ax + Ay = A(x + y) = 0 + 0 = 0$
  - Then $A(cx) = c(Ax) = c(0) = 0$
  - Therefore $x + y$ and cx are also in the nullspace, making it a subspace
- Also called the kernel.
- The solution have n components, so the nullspace is a subspace of $\mathbb{R}^n$
- If the right side b is not zero, then the solutions of $Ax = b$ do NOT form a subspace, since $Ax = b$, $b \neq 0$ does not contain $x = 0$ as a solution (i.e. it can't be a subspace)

- Example:

Given $A=\begin{bmatrix}1&2\\3&6\end{bmatrix} \Rightarrow Ax=0 \Rightarrow \begin{bmatrix}1&2\\3&6\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \Rightarrow \begin{cases}x_1+2x_2=0\\3x_1+6x_2=0\end{cases} \Rightarrow \begin{cases}x_1+2x_2=0\\0=0\end{cases}$

$\therefore$ the nullspace of A is all vectors $x=(x_1, x_2)$ such that x_1+2x_2

- we can take a single solution and consider all its multiples to be in the nullspace.

Choose $x=(-2,1) \Rightarrow$ The nullspace of A, $N(A)$, contains all multiples of $s=(-2,1)$

- We can find these "special solutions" by setting a component of x that is not a pivot to zero
  - Example:

$[1\ 2\ 3]\begin{bmatrix}x\\y\\z\end{bmatrix} = x+2y+3z=0 \Rightarrow s_1=\begin{bmatrix}-2\\1\\0\end{bmatrix},\ s_2=\begin{bmatrix}-3\\0\\1\end{bmatrix}$

not free (don't set to 0)

free (set to 0)

- Consider

$A=\begin{bmatrix}1&2\\3&8\end{bmatrix} \quad B=\begin{bmatrix}A\\2A\end{bmatrix}=\begin{bmatrix}1&2\\3&8\\2&4\\6&16\end{bmatrix} \quad C=[A\ \ 2A]=\begin{bmatrix}1&2&2&4\\3&8&6&16\end{bmatrix}$

$N(A)=0$

$N(B)=0$ b/c the first 2 rows lead to the nullspace. Adding additional rows do not widen the nullspace

$N(C)$:

$C=\begin{bmatrix}1&2&2&4\\3&8&6&16\end{bmatrix}$ becomes $U=\begin{bmatrix}1&2&2&4\\0&2&0&4\end{bmatrix} \Rightarrow Ux=0 \Rightarrow \begin{bmatrix}x_1+2x_2+2x_3+4x_4\\2x_2\qquad+4x_4\end{bmatrix}$

pivot columns, free columns

To get special solutions, set the free columns to 1 and 0

$s_1=\begin{bmatrix}-2\\0\\1\\0\end{bmatrix},\ s_2=\begin{bmatrix}0\\-2\\0\\1\end{bmatrix} \Rightarrow N(C)$ is all linear combinations of s_1 and s_2

- The nullspace becomes easiest to see when A reaches reduced row echelon form (its pivot column becomes I), using the matrix C above:

$U=\begin{bmatrix}1&2&2&4\\0&2&0&4\end{bmatrix}$ becomes $R=\begin{bmatrix}1&0&2&0\\0&1&0&2\end{bmatrix} \Rightarrow \begin{cases}x_1+2x_3=0\\x_2+2x_4=0\end{cases}$

- A nullspace of Z implies the columns of A are independent

$N(A)=Z \iff$ the columns of A are linearly independent.

Solving $Ax=0$ by Elimination

- Even when A is rectangular, we still use elimination
- The steps are still the same:
  1. Forward substitution gets A to triangular U or reduced R.
  2. Back substitution in $Ux=0$ or $Rx=0$ produces x.
- Pivots are still nonzero and the column below the pivots are zero but if a column does not produce a pivot, we simply move to the next column

- Given:

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 2 & 8 & 10 \\ 3 & 3 & 10 & 13 \end{bmatrix} \text{ Apply } E_{31}E_{21} \Rightarrow A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \end{bmatrix} \text{ Apply } E_{33} \Rightarrow A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

our second pivot is 4 (in column 3) only 2 pivots
only pivot variables are x_1 and x_3 (column 1 and 3 have pivots)
only free variables are x_2 and x_4 (column 2 and 4 have no pivots)

- Giving the free variables any values then solving for the pivot variables gives us special solutions

$x_1 + x_2 + 2x_3 + 3x_4 = 0$ set $x_2 = 1$ and $x_4 = 0$, then by back substitution $x_3 = 0$ $x_1 = -1$
$4x_3 + 4x_4 = 0$ Then set $x_2 = 0$ and $x_4 = 1$, then by back substitution $x_3 = -1$ and $x_1 = -1$

The nullspace is the linear combination of these 2 solutions

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

- Example: Find the nullspace of U

$$\begin{array}{c} U = \begin{bmatrix} 1 & 5 & 7 \\ 0 & 0 & 9 \end{bmatrix} \\ \uparrow \text{pivot}_1 \quad \uparrow \text{pivot}_2 \end{array}$$

Pivots: x_1, x_3
free: x_2

U as a system $\begin{cases} x_1 + 5x_2 + 7x_3 = 0 \\ 9x_3 = 0 \end{cases}$

set $x_2 = 1$

$\therefore N(U) = c(-5, 1, 0) \longleftarrow \begin{cases} x_1 + 5(1) + 7x_3 = 0 \\ x_3 = 0 \end{cases} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} -5 \\ 1 \\ 0 \end{bmatrix}$

OR convert U to reduced row echelon form (RREF)

$$U = \begin{bmatrix} 1 & 5 & 7 \\ 0 & 0 & 9 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 5 & 7 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{cases} x_1 + 5x_2 = 0 \\ x_3 = 0 \end{cases} \rightarrow \text{Set } x_2 = 1 \Rightarrow x = \begin{bmatrix} -5 \\ 1 \\ 0 \end{bmatrix}$$

$\uparrow$ pivot column $\uparrow$

Echelon Matrices

- Forward elimination goes from A to U. It acts by row operations, including row exchanges. It goes to the next column when no pivot is available in the current column. The m by n "Staircase" U is an echelon matrix.
- For example, here is a 4×7 echelon matrix with three pivots

$$U = \begin{bmatrix} P & x & x & x & x & x & x \\ 0 & P & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & P & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

  - Three pivots x_1, x_2, x_6
  - Four free variables x_3, x_4, x_5, x_7
  - Four vectors as the basis of $N(U)$

  - The column space of U is $C(U) = c(1, 1, 1, 0)$, $c \in \mathbb{R}$
  - The nullspace of this matrix is a subspace of $\mathbb{R}^7$. It is the linear combinations of the four special solutions – one for each free variable.
    - Set one free variable to 1 and the others to zero. Then solve for the pivot variables
    - Doing this for all free variables gets us the four vectors which act as the basis of $N(U)$

- If A has more columns than rows $(n>m)$, there is at least one free variable, ergo there is at least one special solution, ergo $N(A) \neq Z$
  - Suppose A has more unknowns than equations $(n>m)$, then there must be nonzero solutions. There must be free column's.
  - There must be at least $n-m$ free variables, since the number of pivots cannot exceed m (There can only be 1 pivot per row). When there is a free variable, it can be set to one and a nonzero solution to $Ax=0$ is created
- The nullspace is a subspace; Its "dimension" is the number of free variables

The Reduced Row Echelon Matrix R

- From an echelon matrix U we can go one more step"

Given:

$$U=\begin{bmatrix}1&1&2&3\\0&0&4&4\\0&0&0&0\end{bmatrix}\xrightarrow[\text{into } R_2]{R_2\div 4}\begin{bmatrix}1&1&2&3\\0&0&1&1\\0&0&0&0\end{bmatrix}\xrightarrow[\text{into } R_1]{R_1-2R_2}\begin{bmatrix}1&1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}$$

  - The pivots themselves equal one and there are zeros above and below the pivots. (i.e. the pivot rows contain I)

$$R=\text{rref}(U)=\begin{bmatrix}1&1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}$$

- If A is invertible, then its rref is the identity matrix I.
- To find special solutions:

  1. Set $x_2=1$ and $x_4=0$ and solve $Rx=0$ for x_1 and x_3 $(-1, 1, 0, 0)$ ← in column 2 (with plus signs)
  2. Set $x_2=0$ and $x_4=1$ and solve $Rx=0$ for x_1 and x_3 $(-1, 0, -1, 1)$ ← in column 4 (with plus signs)

  - By reversing signs we can directly read off the special solutions
- Example:

Create a 3×4 matrix whose special solutions to $Ax=0$ are s_1 and s_2

$$s_1=\begin{bmatrix}-3\\1\\0\\0\end{bmatrix}\qquad s_2=\begin{bmatrix}-2\\0\\-6\\1\end{bmatrix}$$

Pivots: columns 1 and 3

Free variables: columns 2 and 4

$$R=\begin{bmatrix}1&3&0&2\\0&0&1&6\\0&0&0&0\end{bmatrix}$$

Pivots

The free columns 2 and 4 will be combinations of the pivot columns

- R can be multiplied on the left by any invertible matrix without changing its nullspace

Problem Set 3.2

1.

a) $A = \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 1 & 2 & 3 & 6 & 9 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix} \xrightarrow[\text{into } R_2]{R_2 - R_1} \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix} \xrightarrow[\text{into } R_3]{R_2 - R_3} \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Pivot variables: x_1, x_3

Free variables: x_2, x_4, x_5

$\begin{cases} x_1 + 2x_2 + 2x_3 + 4x_4 + 6x_5 = 0 \\ x_3 + 2x_4 + 3x_5 = 0 \end{cases}$

Set $(x_2, x_4, x_5) = (1, 0, 0)$

$\begin{cases} x_1 + 2(1) + 2x_3 + 4(0) + 6(0) = 0 \\ x_3 + 2(0) + 3(0) = 0 \end{cases}$

$\begin{cases} x_1 + 2 + 2x_3 = 0 \\ x_3 = 0 \end{cases}$

$x_1 + 2 + 2(0) = 0$

$x_1 = -2$

$s_1 = (-2, 0, 0, 0, 0)$

Set $(x_2, x_4, x_5) = (0, 1, 0)$

$\begin{cases} x_1 + 2(0) + 2x_3 + 4(1) + 6(0) = 0 \\ x_3 + 2(1) + 3(0) = 0 \end{cases}$

$\begin{cases} x_1 + 2x_3 + 4 = 0 \\ x_3 + 2 = 0 \end{cases}$

$\begin{cases} x_1 + 2(-2) + 4 = 0 \\ x_3 = -2 \end{cases}$

$s_2 = (0, 0, -2, 1, 0)$

Set $(x_2, x_4, x_5) = (0, 0, 1)$

$\begin{cases} x_1 + 2(0) + 2x_3 + 4(0) + 6(1) = 0 \\ x_3 + 2(0) + 3(1) = 0 \end{cases}$

$\begin{cases} x_1 + 2x_3 + 6 = 0 \\ x_3 + 3 = 0 \end{cases}$

$\begin{cases} x_1 + 2(-3) + 6 = 0 \\ x_3 = -3 \end{cases}$

$s_3 = (0, 0, -3, 0, 1)$

$N(A) = x_1(-2, 0, 0, 0, 0) + x_2(0, 0, -2, 1, 0) + x_3(6, 0, -6, 0, 1)$

$= (-2x_1 + 6x_3, 0, -2x_2 - 6x_3, x_2, x_3), \; x_1, x_2, x_3 \in \mathbb{R}$

b) $B = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4 & 4 \\ 0 & 8 & 8 \end{bmatrix} \xrightarrow[\text{into } R_3]{R_3 - 2R_2} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4 & 4 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[R_2 \div 4]{R_1 \div 2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{into } R_1]{R_1 - 2R_2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Pivot variables: x_1, x_2 Set $x_3 = 1$

Free variables: x_3

$\begin{cases} x_1 - (1) = 0 \\ x_2 + (1) = 0 \end{cases} \Rightarrow s_1 = (1, -1, 1)$

$N(A) = x_1 s_1 = (x_3, -x_3, x_3)$

7.

3. free columns

4.

$U = \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{into } R_1]{R_1 - 2R_2} \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

5. 6. n

a)

$A = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 10 \end{bmatrix} \xrightarrow[\text{into } R_2]{R_2 - 2R_1} \begin{bmatrix} -1 & 3 & 5 \\ 0 & 0 & 0 \end{bmatrix} = U$

$LU = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 3 & 5 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 10 \end{bmatrix}$

pivot: x_1

Free: x_2, x_3

Set $(x_2, x_3) = (1, 0)$

$-x_1 + 3(1) + 5(0) = 0$

$x_1 = 3$

$x = (3, 1, 0)$

Set $(x_2, x_3) = (0, 1)$

$x_1 + 3(0) + 5(1) = 0$

$x_1 = 5$

$x = (5, 0, 1)$

plane

b)

$B = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 7 \end{bmatrix} \xrightarrow[\text{into } R_2]{R_2 - 2R_1} \begin{bmatrix} -1 & 3 & 5 \\ 0 & 0 & -3 \end{bmatrix}$

pivot: x_1, x_3

free: x_2

Set $x_2 = 1$

$\begin{cases} -x_1 + 3x_2 + 5x_3 = 0 \\ -3x_3 = 0 \end{cases}$

$\begin{cases} -x_1 + 3(1) + 5(0) = 0 \\ x_3 = 0 \end{cases} \rightarrow x = (3, 1, 0)$

9.

a) false $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ 1 pivot

b) true.
An invertible matrix has n pivots.
Therefore there are not free variables

c) false

d) true

11.

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

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

15.
$n-r$, m, m

12.

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

17.
$A = \begin{bmatrix} 1 & -3 & -1 \end{bmatrix}$

17. $(1, 0, 1)$

18.
$(3, 1, 0)$ z

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 12 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

direction vectors

21.
$$\begin{bmatrix} 1 & 0 & -2 & -3 \\ 0 & 1 & -2 & -1 \end{bmatrix}$$

22.
$$\begin{bmatrix} 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

23.
$$\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 1 & 3 & -2 \\ 5 & 1 & -3 \end{bmatrix}$$

24.
$$\begin{matrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{matrix}$$

Chapter 3.3: The Rank and Row Reduced Form

- The numbers m and n give the size of the matrix, but not necessarily the true size of the linear system. An equation like $0=0$ should not count.
  If there are two identical rows in A, they would disappear
- The true size of A is given by its rank.
  - The rank r is the number of pivots.
- This definition is computational.

$$A = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 2 & 6 \end{bmatrix}$$

  - Column 1 & 3 are multiple
  - Column 4 is a combination of column 1 and 2
  - 2 pivots $\to$ rank(A) = rank(U) = 2

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

- The rank is the dimension of the vector space spanned by its columns.
  This corresponds to the maximal number of linearly independent columns of A, which is also equal to the dimension of the vector space spanned by its rows.
  - The column rank (dimension of column span) always equals the row rank (dimension of row span)
- The nullity is, similarly, the dimension of the nullspace of A, which corresponds to the number of free columns of A.
- The rank-nullity theorem states that rank(A) + nullity(A) = n = number of columns.

Rank One

- Matrices of rank one only have one pivot. When elimination produces zero in the first column, it produces zero in all the columns. Every row is a multiple of the pivot row.
- The column space of a rank one matrix is one-dimensional
- For example

$$A=\begin{bmatrix}1&3&10\\2&6&20\\3&9&30\end{bmatrix}\longrightarrow R=\begin{bmatrix}1&3&10\\0&0&0\\0&0&0\end{bmatrix}$$

  - The columns are all multiples of $u=(1,2,3)$
  - $A=\begin{bmatrix}u & 3u & 10u\end{bmatrix}$
  - We can write A as an outer product:

$v^T=[1\ 3\ 10]$

$$uv^T=\begin{bmatrix}1\\2\\3\end{bmatrix}[1\ 3\ 10]=\begin{bmatrix}1&3&10\\2&6&20\\3&9&30\end{bmatrix}$$

- The nullspace is easy to visualize with a rank one matrix
  That equation $u(v^Tx)=0$ leads us to $v^Tx=0$ i.e. the vector v^T is orthogonal (perpendicular) to x. All vectors in the nullspace are on a plane perpendicular to v^T
- Now with numbers:
  pivot variable: x_1
  free variables: x_2, x_3
  Since R is already in rref, we can simply read the solutions off
  $s_1=(-3,1,0)$, $s_2=(-10,0,1)$

$N(A)=c_1s_1+c_2s_2$ (plane) $\xrightarrow[\text{standard}]{\text{convert to}}$

$$\begin{matrix}-3&1&0&-3&1&0\\-10&0&1&-10&0&1\end{matrix}$$

$$\begin{matrix}1&0&0\\0&-3&-10\end{matrix}$$

$\vec{n}=(1,3,10)=v$

$\therefore$ $N(A)$ is the plane $x+3y+10z=0$,
$\therefore$ the nullspace is the plane normal to the row vector $(1,3,10)$

The Pivot columns

- The pivot columns of $R=\text{rref}(A)$ have 1's in their pivots and 0's elsewhere. All together, the r pivot columns contain an $r\times r$ identity matrix. It sits above $m-r$ rows of zeroes
- The numbers of the pivot columns are in the list Pivcol
- The pivot columns of A are not immediately obvious from A itself, but A and R have the same pivot columns

$$A=\begin{bmatrix}1&3&0&2&-1\\0&0&1&4&-3\\1&3&1&6&-4\end{bmatrix}\longrightarrow R=\begin{bmatrix}1&3&0&2&-1\\0&0&1&4&-3\\0&0&0&0&0\end{bmatrix}\quad \text{Pivcol}=(1,3)$$

  - The column space of A and R are different

- The r pivot columns of A are also the first r columns of E^{-1}

$$A = \begin{bmatrix} \boxed{1} & 3 & \boxed{0} & 2 & -1 \\ \boxed{0} & 0 & \boxed{1} & 4 & -3 \\ \boxed{1} & 3 & \boxed{1} & 6 & -4 \end{bmatrix} \quad E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \Rightarrow E^{-1} = \begin{bmatrix} \boxed{1} & \boxed{0} & 0 \\ \boxed{0} & \boxed{1} & 0 \\ \boxed{1} & \boxed{1} & 1 \end{bmatrix}$$

  - When we compute $A = E^{-1}R$, the r by r identity matrix inside R picks out the first r columns of E^{-1} as columns of $A = E^{-1}R$
- The pivot columns are not combinations of earlier (leftward) columns. The free columns are instead combinations of earlier columns. These combinations are the special solution

The Special Solutions

- Each special solution to $Ax=0$ and $Rx=0$ has one free variable equal to 1. The others are all set to zero. These solutions come directly from the RREF R.

$$Rx = \begin{bmatrix} 1 & \boxed{3} & 0 & \boxed{2} & \boxed{-1} \\ 0 & \boxed{0} & 1 & \boxed{4} & \boxed{-3} \\ 0 & \boxed{0} & 0 & \boxed{0} & \boxed{0} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

  - Notice that the first & second entries in each free column corresponds to the values of the pivot variables in special solutions.

1. Set $(x_2, x_4, x_5) = (1,0,0)$
   $s_1 = (-3,1,0,0,0)$
2. Set $(x_2, x_4, x_5) = (0,1,0)$
   $s_2 = (-2,0,-4,1,0)$
3. Set $(x_2, x_4, x_5) = (0,0,1)$
   $s_3 = (1,0,3,0,1)$

- The nullspace matrix N contains the three special solutions as columns. As each column corresponds to a solution x that makes $Ax=0$, it should be no suprise that $AN = 0 =$ zero matrix.

Null space matrix

$n - r = 5 - 2 = 3$ special solutions

by rank-nullity theorem

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

  - Whereas R had the identity matrix (2×2) in its pivot columns, N has the identity matrix (3×3) in its free rows
- The special solutions for R are really easy to find
  - Suppose that the first r columns are the pivot columns. R now looks like

$$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} \begin{matrix} \} \ r \text{ pivot rows} \\ \} \ m-r \text{ zero rows} \end{matrix}$$

r pivot columns $\quad$ $n-r$ free column

- The pivot variables in the $n-r$ special solutions come by changing F to $-F$

$$N = \begin{bmatrix} -F \\ I \end{bmatrix} \begin{matrix} r \text{ pivot variables} \\ n-r \text{ free variables} \end{matrix} \Rightarrow \text{then } RN = 0 \Rightarrow \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -F \\ I \end{bmatrix} = \begin{bmatrix} -IF + FI \\ 0 \end{bmatrix} = \begin{bmatrix} -F + F \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- The columns of N solve $Rx = 0$

$$Rx = 0 \Rightarrow \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{pivot} \\ x_{free} \end{bmatrix} = 0 \Rightarrow I x_{pivot} + F x_{free} = 0 \Rightarrow I x_{pivot} = -F x_{free}$$

- In each special solution, the free variables are a column of I. Then the pivot variables a column of $-F$.
- The idea still holds even when the pivot columns and free columns are mixed together:

Problem Set 3.3

2

a)

$$A = \begin{bmatrix} 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 \end{bmatrix} \rightarrow R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

b)

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \xrightarrow[\text{into } R_2]{R_2 - 2R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -3 \\ 3 & 4 & 5 & 6 \end{bmatrix} \xrightarrow[\text{into } R_3]{R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -3 \\ 0 & -2 & -4 & -6 \end{bmatrix} \xrightarrow[\text{into } R_3]{R_3 - 2R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[\text{into } R_2]{R_2 \times -1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{into } R_1]{R_1 - 2R_2} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R$$

3.

$$B = \begin{bmatrix} A & A \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 3 \\ 2 & 4 & 6 & 2 & 4 & 6 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 2 & 4 & 6 & 2 & 4 & 6 \\ 0 & 0 & 3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[R_2 \div 3]{R_1 \div 2}$$

$$\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 3 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{into } R_1]{R_1 - 3R_2} \begin{bmatrix} 1 & 2 & 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

4

$$R = \begin{bmatrix} F & I \\ 0 & 0 \end{bmatrix} \} \; m-r \text{ zero rows}$$

$$N = \begin{bmatrix} I \\ -F \end{bmatrix}$$

8

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 4 & 8 & 16 \end{bmatrix}$$

9. line, $n-1$, $m \times (n-1)$

16.

10

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

15.

$$AB = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & 4 \\ 3 & 1.5 & 6 \end{bmatrix} = \begin{bmatrix} 8 & 4 & 16 \\ 16 & 8 & 32 \end{bmatrix}$$

rank A = 1 rank B = 1 rank AB = 1

Chapter 3.4: The complete solution to $Ax = b$

- We augment b onto A into the matrix $[A\ b]$
- Given the matrix:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \Longrightarrow [A\ b] = \left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 1 & 3 & 1 & 6 & 7 \end{array}\right]$$

- This ensure all operations on A also act on b.
- In the above example, we can subtract row 1 and row 2 from row 3

$$\left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 1 & 3 & 1 & 6 & 7 \end{array}\right] \xrightarrow[\text{into } R_3]{R_3 - R_2 - R_1} \left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] = [R\ d]$$

  - The third equation has become $0 = 0$. This means equation 3 = equation 1 + equation 2

One Particular Solution

- For the above example, we can get an easy solution by setting the free variables $x_2 = x_4 = 0$. Then the two nonzero equations give $x_1 = 1$ and $x_3 = 6$. our Solution is $(1, 0, 6, 0)$
- For a solution to exist, zero rows in R must be zero in d as well $(0=0)$. Since I is in the pivot rows and columns of R, the pivot variables in our one particular solution come from d once we set all free variables to zero.
- Note that this is one solution out of infinitely many.
- We write the complete solution for $Ax=b$ as

$$x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4\begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

  - The sum of a particular solution and the null space vector
  - $Ax = b$ is solved by one particular solution x_p
  - $Ax = 0$ is solved by the nullspace

$$Ax = \underbrace{Ax_p}_{b} + \underbrace{Ax_n}_{0}$$

$$Ax = b$$

- If A is an invertible square matrix, x_p is the only solution $A^{-1}b$, while the nullspace is 0 because an invertible matrix has n pivots. Then, our complete solution is $x = A^{-1}b + 0 = A^{-1}b$. Invertible square matrices are a special case because their nullspace is 0
- Example:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ -2 & -3 \end{bmatrix} \Rightarrow [A\ b] = \left[\begin{array}{cc|c} 1 & 1 & b_1 \\ 1 & 2 & b_2 \\ -2 & -3 & b_3 \end{array}\right] \xrightarrow[\text{into } R_2]{R_2 - R_1} \left[\begin{array}{cc|c} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ -2 & -3 & b_3 \end{array}\right] \xrightarrow[\text{into } R_3]{R_3 + 2R_1} \left[\begin{array}{cc|c} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & -1 & b_3 + 2b_1 \end{array}\right]$$

$$\xrightarrow[\text{into } R_3]{R_2 + R_3} \left[\begin{array}{cc|c} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 + b_1 + b_2 \end{array}\right] \xrightarrow[\text{into } R_1]{R_1 - R_2} \left[\begin{array}{cc|c} 1 & 0 & 2b_1 - b_2 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 + b_2 + b_1 \end{array}\right] \quad x = x_p + x_n = \begin{bmatrix} 2b_1 - b_2 \\ b_2 - b_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$b_1 + b_2 + b_3$ must $= 0$

- The matrix has full column rank ($r=n$). Every column has a pivot.
  The matrix is reduced to
  $$R = \begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} n \text{ by } n \text{ identity matrix} \\ m-n \text{ rows of zeroes} \end{bmatrix}$$
- Every matrix with full column rank has these properties
  1. All columns of A are pivot columns
  2. There are no free variables or special solutions (nullity $=0$)
  3. The nullspace contains only the zero vector $x=0$
  4. If $Ax=b$ has a solution (it might not), then it only has one

The Complete Solution

- The other extreme case is full row rank. Then $Ax=b$ has one or infinitely many solutions.
- In this case the matrix must have dimensions $m \le n$. If $r=m$ (every row has a pivot), then it has full row rank.
- Example:
  $$\begin{cases} x+y+z=3 \\ x+2y-z=4 \end{cases} \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & -1 & 4 \end{array}\right] = A \quad (\text{rank } A = m = 2)$$
  - These 2 equations are two planes in 3D. They are not parallel so the intersect at a line. Our particular solution x_p will be a point on the line. Adding the nullspace vectors will move us along the line. Then $x = x_p + x_n$ will give us the complete solution (See fig 3.4.1)

$$A \to R \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & -1 & 4 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & -2 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 1 \end{array}\right] = [R \;\; d]$$

Particular Solution:
Set $\underset{\text{free}}{x_3} = 0$,
then $x_1 = 2$, $x_2 = 1$
$x_p = (2, 1, 0)$

nullspace solutions
Set $\underset{\text{free}}{x_3} = 1$ for $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$
then $x_1 = -3$, $x_2 = 2$
$x_n = s_1 = (-3, 2, 1)$

$$x = x_p + x_n = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$$

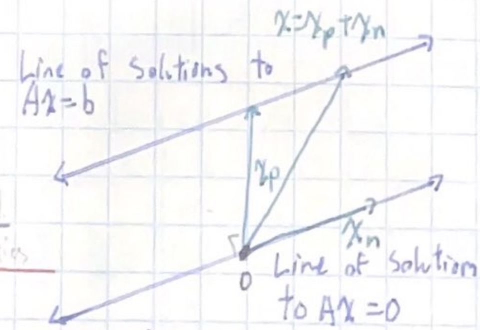


Fig 3.4.1

- If $m<n$, the equation $Ax=b$ is underdetermined
- Every matrix with full row rank has these properties
  1. All rows have pivots, and R has no zero rows
  2. $Ax=b$ has a solution for every right side b
  3. The column space is the whole space $\mathbb{R}^m$
  4. There are $n-r = n-m$ special solutions in the nullspace of A

- The four possibilities for linear equations.
  1. $r=m$ and $r=n$
     - Square and invertible
     - $Ax=b$ has one solution
  2. $r=m$ and $r<n$
     - Short and wide
     - $Ax=b$ has infinite solutions
  3. $r<m$ and $r=n$
     - Tall and thin
     - $Ax=b$ has 0 or 1 solution
  4. $r<m$ and $r<n$
     - Not full rank
     - $Ax=b$ has 0 or infinite solutions

Problem Set 3.4

1.

$$\begin{bmatrix} A & b \end{bmatrix} = \left[\begin{array}{cccc|c} 2 & 4 & 6 & 4 & 4 \\ 2 & 5 & 7 & 6 & 3 \\ 2 & 3 & 5 & 2 & 5 \end{array}\right] \xrightarrow[R_3-R_1]{R_2-R_1} \left[\begin{array}{cccc|c} 2 & 4 & 6 & 4 & 4 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & -1 & -1 & -2 & 1 \end{array}\right] \xrightarrow[\text{into } R_3]{R_2+R_3} \left[\begin{array}{cccc|c} 2 & 4 & 6 & 4 & 4 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Column space: R^2

plane formed by $(2,0,0)$ and $(4,1,0)$

$$\begin{array}{c|cc} 2 & 0 & 0 \\ 4 & 1 & 0 \end{array} \to \begin{matrix} 2 & 0 & 4 \\ 4 & 1 & 0 \end{matrix}$$

$$\begin{matrix} 0 & 0 & 2 \\ 0 & 0 & 0 \end{matrix}$$

$n=(0,0,2)$
$=2(0,0,1)$

$\Pi: z=0$

Nullspace: $\mathbb{R}^2$

s_1: Set $(x_3, x_4)=(1,0)$
$s_1=(-1,-1,1,0)$
s_2: Set $(x_3,x_4)=(0,1)$
$s_2=(2,-2,0,1)$

$x_n=c_1s_1+c_2s_2$

$$=c_1\begin{bmatrix}-1\\-1\\1\\0\end{bmatrix}+c_2\begin{bmatrix}2\\-2\\0\\1\end{bmatrix}$$

Particular Solution:

Set $x_3=x_4=0$
$x_p=(4,-1,0,0)$

$x=x_p+x_n$

$$=\begin{bmatrix}4\\-1\\0\\0\end{bmatrix}+c_1\begin{bmatrix}-1\\-1\\1\\0\end{bmatrix}+c_2\begin{bmatrix}2\\-2\\0\\1\end{bmatrix}$$

$$\begin{bmatrix} U & c \end{bmatrix} = \left[\begin{array}{cccc|c} 2 & 4 & 6 & 4 & 4 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1 \div 2} \left[\begin{array}{cccc|c} 1 & 2 & 3 & 2 & 2 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow[\text{into } R_1]{R_1-2R_2} \left[\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 4 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

2.

$$\begin{bmatrix} A & b \end{bmatrix} = \left[\begin{array}{ccc|c} 2 & 1 & 3 & b_1 \\ 6 & 3 & 9 & b_2 \\ 4 & 2 & 6 & b_3 \end{array}\right] \xrightarrow[\text{into } R_3]{R_3-2R_1} \left[\begin{array}{ccc|c} 2 & 1 & 3 & b_1 \\ 6 & 3 & 9 & b_2 \\ 0 & 0 & 0 & b_3-2b_1 \end{array}\right] \xrightarrow[\text{into } R_2]{R_2-3R_1} \left[\begin{array}{ccc|c} 2 & 1 & 3 & b_1 \\ 0 & 0 & 0 & b_2-3b_1 \\ 0 & 0 & 0 & b_3-2b_1 \end{array}\right]$$

$\begin{cases} b_2-3b_1=0 \\ b_3-2b_1=0 \end{cases}$ these 2 planes put b in the column space, which is a line.

3.

$$[A\ b] = \left[\begin{array}{ccc|c} 1 & 3 & 3 & 1 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 5 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{ccc|c} 1 & 3 & 3 & 1 \\ 0 & 0 & 3 & 3 \\ -1 & -3 & 3 & 5 \end{array}\right] \xrightarrow{R_3+R_1} \left[\begin{array}{ccc|c} 1 & 3 & 3 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 6 & 6 \end{array}\right]$$

$$\xrightarrow{R_3-2R_2} \left[\begin{array}{ccc|c} 1 & 3 & 3 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_2\div 3} \left[\begin{array}{ccc|c} 1 & 3 & 3 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1-3R_2} \left[\begin{array}{ccc|c} 1 & 3 & 0 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] = [R\ d]$$

$$x = x_p + x_n = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + c\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$$

4.

$$[A\ b] = \left[\begin{array}{cccc|c} 1 & 3 & 1 & 2 & 1 \\ 2 & 6 & 4 & 8 & 3 \\ 0 & 0 & 2 & 4 & 1 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{cccc|c} 1 & 3 & 1 & 2 & 1 \\ 0 & 0 & 2 & 4 & 1 \\ 0 & 0 & 2 & 4 & 1 \end{array}\right] \xrightarrow{R_3-R_2} \left[\begin{array}{cccc|c} 1 & 3 & 1 & 2 & 1 \\ 0 & 0 & 2 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_2\div 2} \left[\begin{array}{cccc|c} 1 & 3 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1-R_2} \left[\begin{array}{cccc|c} 1 & 3 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 2 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$x = x_p + x_n = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \end{bmatrix} + c_1\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_2\begin{bmatrix} 0 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

7.

$$[A\ b] = \left[\begin{array}{ccc|c} 1 & 3 & 1 & b_1 \\ 3 & 8 & 2 & b_2 \\ 2 & 4 & 0 & b_3 \end{array}\right] \xrightarrow{R_3-2R_1} \left[\begin{array}{ccc|c} 1 & 3 & 1 & b_1 \\ 3 & 5 & 2 & b_2 \\ 0 & -2 & -2 & b_3-2b_1 \end{array}\right] \xrightarrow{R_2-3R_1} \left[\begin{array}{ccc|c} 1 & 3 & 1 & b_1 \\ 0 & -1 & -1 & b_2-3b_1 \\ 0 & -2 & -2 & b_3-2b_1 \end{array}\right]$$

$$\xrightarrow{R_3-2R_2} \left[\begin{array}{ccc|c} 1 & 3 & 1 & b_1 \\ 0 & -1 & -1 & b_2-3b_1 \\ 0 & 0 & 0 & b_3-4b_2+6b_1 \end{array}\right]$$

8

a)

$$[A\ b] = \left[\begin{array}{ccc|c} 1 & 2 & 1 & b_1 \\ 2 & 6 & 3 & b_2 \\ 0 & 2 & 5 & b_3 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{ccc|c} 1 & 2 & 1 & b_1 \\ 0 & 2 & 1 & b_2-2b_1 \\ 0 & 2 & 5 & b_3 \end{array}\right] \xrightarrow{R_3-R_2} \left[\begin{array}{ccc|c} 1 & 2 & 1 & b_1 \\ 0 & 2 & 1 & b_2-2b_1 \\ 0 & 0 & 4 & b_3-b_2+2b_1 \end{array}\right]$$

$$\xrightarrow{R_3\div 4} \left[\begin{array}{ccc|c} 1 & 2 & 1 & b_1 \\ 0 & 2 & 1 & b_2-2b_1 \\ 0 & 0 & 1 & \frac{1}{4}b_3-\frac{1}{4}b_2+\frac{1}{2}b_1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & b_1-b_2 \\ 0 & 1 & 0 & \frac{5}{8}b_2-\frac{3}{4}b_1-\frac{1}{8}b_3 \\ & & 1 & \end{array}\right]$$

20

$$A = \begin{bmatrix} 3 & 4 & 1 & 0 \\ 6 & 5 & 2 & 1 \end{bmatrix} \longrightarrow \overset{L}{\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}} \begin{bmatrix} 3 & 4 & 1 & 0 \\ 6 & 5 & 2 & 1 \end{bmatrix} = \overset{U}{\begin{bmatrix} 3 & 4 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{bmatrix}}$$

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \xrightarrow{E_{31}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix} \xrightarrow{E_{21}} \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$$

$$\xrightarrow{E_{32}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad A = LU \Rightarrow \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

24

a) $A = \begin{bmatrix} 1 & 5 \\ 2 & 7 \\ 3 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{bmatrix}$ d) $\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$

Chapter 3.5: Independence, Basis, and Dimension

Linear Independence

- The columns of A are linearly independent when the only solution to $Ax = 0$ is $x = 0$
  - That is, no one column vector can be expressed as a linear combination of the others.
  - This also implies that $N(A) = Z$.
  - If there are nonzero (non-trivial) solutions then the vectors are said to be linearly dependent.
  - Examples
    - (1,0) and (0,1) are independent
    - (1,0) and (1, 0.0001) are independent
- Three vectors in $\mathbb{R}^2$ cannot be independent. One way to see this is to write it as a matrix:

  $$\begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

  - There must be a free variable
- Example 1: The columns of A are dependent.

  $$Ax = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = -3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

  - The rank is only 2. Independent columns produce full column rank $r = n = 3$
  - In this matrix, the rows are also dependent. For a square matrix, we will show that dependent columns imply dependent rows.
- Suppose we have a matrix with $m < n$ (short and wide). Then the columns must be dependent. There are at least $n - m$ free variables
- If instead we have a matrix with $n \le m$ (tall and thin). Elimination will reveal the r pivot columns. It is those r pivot columns that are independent.

Vectors that Span a Subspace

- The column space consists of all combinations Ax of the columns of A.
  We now introduce a new term to describe this: Span
- A set of vectors span a space if their linear combinations fill the space
- The columns of a matrix spans its column space.
- Examples:
  - $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ span the entirety of $\mathbb{R}^2$
  - $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $v_3 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$ also span the full space $\mathbb{R}^2$
  - $w_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $w_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ span a line ($\mathbb{R}^1$) in $\mathbb{R}^2$
- The row space is the space spanned by the rows of a matrix.
  The row space of a matrix is the column space of its transpose.
  row space $= C(A^T)$
- The row space of A is a subspace of $\mathbb{R}^n$ (The rows have n components)
- Example:

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

$C(A)$ spans a $\mathbb{R}^2$ plane in $\mathbb{R}^3$
$C(A^T)$ spans a $\mathbb{R}^2$ plane in $\mathbb{R}^2$

A Basis for a Vector Space

- Two vectors can't span all of $\mathbb{R}^3$ even if they are independent. Four vectors can't be independent even if they span $\mathbb{R}^3$.
- We want enough independent vectors to span the space and not more. Hence the "basis"
- A basis for a vector space is a sequence of vectors that both spans the space and are linearly independent.
  - Every vector v in the space is a combination of the basis vectors
  - Moreover, that combination is unique because because the basis vectors are linearly independent.
    - Suppose $v = a_1v_1 + \dots + a_nv_n$ and also $v = b_1v_1 + \dots + b_nv_n$. By subtraction $(a_1 - b_1)v + \dots + (a_n - b_n)v_n = 0$. $a_i - b_i$ must equal zero from the independence of the v's
- The columns of I_n produce the "standard basis" for $\mathbb{R}^n$
- The columns of every n by n invertible matrix give a basis for $\mathbb{R}^n$ — not pivot columns
- The pivot columns of A are a basis for its column space and likewise for the pivot rows and its row space. This also applies to the pivot rows of R
- Given

$$A = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix},$$ Basis for column space: $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ Basis for row space: $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

- Given

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$ Basis for col space: $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ Basis for row space: $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 4 \end{bmatrix}$

- Given 5 vectors in $\mathbb{R}^7$, how do you find a basis for the space they span?
  - Make them rows of A and eliminate to find the non zero rows of R
  - Make them columns of A and eliminate to find the pivot columns of A (not R)
- All bases for a vector space contain the same number of vectors
  - That number is the "dimension" of the space

Dimension of a Vector Space

- We want to prove that all bases for the same vector space have the same number of vectors.
  - If $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$ are both bases for the same vector spaces, then $m = n$
  - Suppose that there are more w's than v's. From $n > m$, we want to reach a contradiction
  - If w_1 equals $a_{11}v_1 + \ldots + a_{m1}v_m$, this is the first column of a matrix multiplication VA. Each w is a combination of the v's

$$W = \begin{bmatrix} w_1 & w_2 & \ldots & w_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdot & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = VA$$

  - We don't know each a_{ij}, but we know the shape of A (m by n). The second vector w_2 is also a combination of the v's. The key is that A has a row for every v and a column for every w. A is a short wide matrix since we assumed $n > m$. So $Ax = 0$ has infinite solutions.
  - $Ax = 0$ gives $VAx = 0$, which is $Wx = 0$. A combination of the w's gives zero, then the w's cannot be a basis.
  - Likewise, assuming $m > n$ and performing the same step leads to the same contradiction.
  - Therefore $m = n$ and the proof is complete
- Dimension: the dimension of a space is the number of vectors in every basis
- The dimension of the column space equals the rank of the matrix

Bases for Matrix Spaces and Function Spaces

- The terms "independence", "bases", and "dimension" are not just restricted to column vectors. It applys to any vector spaces.
  - We can ask if some matrices are independent
  - In differential equation $\frac{d^2y}{dx^2} = y$ has a space of solutions. one basis is $y = e^x$ and $y = e^{-x}$. Counting the basis functions, we get a dimension of 2)
- Matrix spaces
  - The vector space M contains all 2 by 2 matrices. Its dimension is 4. ~~one basis is~~ $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. ~~These matrices are linearly independent~~

- The matrices $\begin{bmatrix}1&0\\0&0\end{bmatrix}$, $\begin{bmatrix}0&1\\0&0\end{bmatrix}$, and $\begin{bmatrix}0&0\\0&1\end{bmatrix}$ are a basis for a subspace, the upper triangular matrices. It has a dimension of three.
- $\begin{bmatrix}1&0\\0&0\end{bmatrix}$, $\begin{bmatrix}1&0\\0&1\end{bmatrix}$ are a basis for diagonal matrices, with dimension 2
- Think about the space of all n by n matrices. One possible basis has a 1 in each position in the matrix, then the dimension is n^2 since there are n^2 positions.
  - The dimension of the whole n by n matrix space is n^2.
  - The dimension of upper triangular matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$
  - The dimension of diagonal matrices is n
  - The dimension of symmetric matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$
- Function spaces
  - $y'' = 0$ is solved by any linear function $y = cx + d$. Basis: $\{x, 1\}$ ← null space
  - $y'' = -y$ is solved by $y = c\sin x + d\cos x$ Basis: $\{\sin x, \cos x\}$
  - $y'' = y$ is solved by $y = ce^x + de^{-x}$, Basis: $\{e^x, e^{-x}\}$
- The dimension is two for the above equations since it is the second derivative.
- The solutions of $y'' = 2$ don't form a subspace. The right side $b = 2$ is not zero
  - A particular solution is $y = x^2$
  - A complete solution is $y = \underbrace{x^2}_{x_p} + \underbrace{cx + d}_{x_n}$
- The space Z contains only the zero vector. The dimension of this set is zero. The basis for Z is the empty set.

Problem Set 3.5

2.

$$V = \begin{bmatrix}1&1&1&0&0&0\\-1&0&0&1&1&0\\0&-1&0&-1&0&1\\0&0&-1&0&-1&-1\end{bmatrix} \xrightarrow[\text{into } R_2]{R_2 + R_1} \begin{bmatrix}1&1&1&0&0&0\\0&1&1&1&1&0\\0&-1&0&-1&0&1\\0&0&-1&0&-1&-1\end{bmatrix} \xrightarrow[\text{into } R_3]{R_3 + R_2} \begin{bmatrix}1&1&1&0&0&0\\0&1&1&1&1&0\\0&0&1&0&0&1\\0&0&-1&0&-1&-1\end{bmatrix}$$

$$\xrightarrow[\text{into } R_4]{R_4 + R_3} \begin{bmatrix}1&1&1&0&0&0\\0&1&1&1&1&0\\0&0&1&0&0&1\\0&0&0&0&-1&0\end{bmatrix} \xrightarrow{R_4 \times -1} \begin{bmatrix}1&1&1&0&0&0\\0&1&1&1&1&0\\0&0&1&0&0&1\\0&0&0&0&1&0\end{bmatrix}$$

Columns 1, 2, 3, 5 are independent and form a basis for $\mathbb{R}^4$

3.

The columns of U are independent if the matrix U has full rank (i.e. there are n pivots). If any of a, d, or f equal zero, there would be less than 3 pivots.

4.

$$Ux = 0 \Rightarrow \begin{bmatrix}a&b&c\\0&d&e\\0&0&f\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Since the matrix has full rank (3 pivots), its nullspace contains only the zero vector.

7.

$a_1v_1 + a_2v_2 + a_3v_3 = 0$

$a_1(u_2 - u_3) + a_2(u_1 - u_3) + a_3(u_1 - u_2) = 0$

$a_1u_2 - a_1u_3 + a_2u_1 - a_2u_3 + a_3u_1 - a_3u_2 = 0$

$(a_2u_1 + a_3u_1) + (a_1u_2 - a_3u_2) + (-a_1u_3 - a_2u_3) = 0$

$(a_2+a_3)u_1 + (a_1-a_3)u_2 - (a_1+a_2)u_3 = 0$

Since u_1, u_2, u_3 are independent, the only solution is

$\begin{cases} a_2 + a_3 = 0 \\ a_1 - a_3 = 0 \\ a_1 + a_2 = 0 \end{cases} \to \begin{cases} a_2 + a_3 = 0 \\ a_1 = a_3 \\ a_1 + a_2 = 0 \end{cases} \to \begin{cases} a_1 + a_2 = 0 \\ a_1 = a_3 \to a_1 + a_2 = 0 \\ a_1 + a_2 = 0 \end{cases}$ $(1, -1, 1)$

$\therefore v_1, v_2, v_3$ are linearly dependent

11. 12. 15. 16 a)

a) $b \in$ $Ax = b$ n, basis, and ... $(1, 1, 1, 1)$

b) $\rho \in$ $A^Ty = c$ 17.   26.

c) $\mathbb{R}^3$ a) ... b_n

d) $\mathbb{R}^3$ b) τm a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

17.

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ $\begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

30.

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

32.

$\frac{dy}{dx} - 2y = 0$

$\frac{dy}{dx} = 2y$

$dy = 2y\,dx$

$\frac{1}{y}dy = 2dx$

$\ln|y| + C_1 = 2x + C_2$

$|y| = e^{2x}e^{C}$

$y = \pm e^{C}e^{2x}$

$= Ae^{2x}$

35.

$p(x) = ax^3 + bx^2 + cx + dx$

$a + b + c + d = 0$

$q(x) = x^3 - x^2$

Chapter 3.6: Dimensions of the Four Subspaces

- The main theorem in this chapter connects rank and dimension.
  The rank of a matrix is the number of pivots. The dimension of a subspace is the number of vectors in a basis.
- The rank of A reveals the dimension of all four fundamental subspaces, which are:
  1. The row space is $C(A^T)$, a subspace of $\mathbb{R}^n$
  2. The column space is $C(A)$, a subspace of $\mathbb{R}^m$
  3. The nullspace is $N(A)$, a subspace of $\mathbb{R}^n$
  4. The left nullspace is $N(A^T)$, a subspace of $\mathbb{R}^m$
- For the left nullspace, we solve $A^T y = 0$, an n by m system.
  The equation can also be written $y^T A = 0^T$.
- The matrices A and A^T are usually different, and so are their column spaces and nullspaces. But these spaces are all connected.
- Part 1 of the fundamental theorem finds the dimensions of the four subspaces
  - One fact stands out: The row space and column space have the same dimension r.
    $\text{Rank}(A) = \dim(C(A)) = \dim(C(A^T))$
  - Another important fact involve the two nullspaces:
    $N(A)$ and $N(A^T)$ have dimensions $n-r$ and $m-r$ respectfully, to make up the full n and m with the row and column spaces.
- Part 2 will describe how the four subspaces fit together.

The Four Subspaces for R

- Suppose A is reduced to its row echelon form R. For that special form, the four subspaces are easy to identify.
- We will find a basis for each subspace and check its dimension. Then we watch how the subspaces change (two of them don't) as we go back to A.
- The main point is that the four dimensions are the same for A and R.
- As a specific example:

$$R = \begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{array}{l} \text{pivot row 1} \\ \text{pivot row 2} \\ \\ \end{array} \Big\} 2$$

Pivot column 1, pivot column 4 $\Big\}$ 2

- $\text{Rank}(A) = \dim(C(A)) = \dim(C(A^T))$

1. Row Space
   - The non-zero rows of R form a basis
2. Column space
   - The pivot columns of R form a basis.
   - Every free column is a combination of the pivot columns. These combinations are the special solutions.
3. Null space
   - The nullspace has dimension $n-r=3$
   - There are 3 free variables
4. Left null space
   - The equation $R^T y = 0$ looks for combinations of the columns of R^T which produce zero. The pivot rows are independent, so to solve the equation, we set the pivot rows coefficient to zero and the zero rows can have any coefficient

- Summary:
  - In $\mathbb{R}^n$ the row space and nullspace have dimension r and $n-r$ (adding to n)
  - In $\mathbb{R}^m$ the column space and left nullspace have dimensions r and $m-r$ (adding to m)
- So far this is proved for echelon matrices R.

The Four Subspaces for A

- The subspace dimensions for A are the same as R. Why?

A reduces to R $\quad A = \begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 3 & 5 & 1 & 9 \end{bmatrix}$ Notice $C(A) \neq C(R)$

- An elimination matrix E takes A to R. The invertible matrix E is the product of the elementary matrix that reduce A to R.

$EA = R$ and $A = E^{-1}R$

- Proof:
  1. Row space
     - Same dimension r and same basis.
     - Every row of A is a combination of the rows of R and vice versa. Elimination changes rows but not row spaces.
     - The basis rows for A are the ones that end up as pivot rows
  2. column space
     - The column space of A has dimension r
     - The number of independent columns equals the number of independent rows
     - The same combinations of the columns are zero for A and R. That is, $Ax=0$ exactly when $Rx=0$, even though their column spaces are different. The r pivot columns of both are independent.
  3. Nullspace
     - A has the same nullspace as R
     - The elimination steps don't change the solutions. The special solutions are a basis for this nullspace. There are $n-r$ free variables, so the dimension is $n-r$.
  4. Left nullspace
     - The left nullspace has dimension $m-r$
     - The column space has dimension r. Since A^T is n by m, the whole space is $\mathbb{R}^m$
- Fundamental Theorem of Linear Algebra, Part One
  - The column space and row space both have dimension r
  - The nullspace and left nullspace have dimension $n-r$ and $m-r$ respectively

- Example 1

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

  - $\dim(C(A)) = r = 1$
  - $\dim(C(A^T)) = r = 1$
  - $\dim(N(A)) = n-r = 3-1 = 2$
  - $\dim(N(A^T)) = m-r = 1-1 = 0$ (Z)

- Example 2

$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$

  - $\dim(C(A)) = r = 1$
  - $\dim(C(A^T)) = r = 1$
  - $\dim(N(A)) = n-r = 3-1 = 2$
  - $\dim(N(A^T)) = m-r = 2-1 = 1$

Matrices of Rank One

- When $r=1$, every row is a multiple of the same row, and likewise with columns
- The row space is a line in $\mathbb{R}^n$ and the column space is a line in $\mathbb{R}^m$
- Every rank one matrix is expressible as an outer product uv^T
- The nullspace is the plane perpendicular to v

  $Ax=0 \Rightarrow (uv^T)x=0 \Rightarrow u(v^Tx)=0 \Rightarrow v^Tx=0 \Rightarrow v$ and x are perpendicular
  - It is this perpendicularity of the subspaces that will be part 2 of the Fundamental Theorem

Problem set 3.6

1.

a) $m=7$

$n=4$

$\dim(C(A))=\dim(C(A^T))=r=5$

$\dim(N(A))=n-r=9-5=4$

$\dim(N(A^T))=m-r=9-7=2$

b) $m=3,\ n=4$

$\dim(C(A))=\dim(C(A^T))=r=3$

$C(A)=\mathbb{R}^3$

$\dim(N(A^T))=m-r=3-3=0=Z$

2.

$$A=\begin{bmatrix}1&2&4\\2&4&8\end{bmatrix}$$

row basis: $\{(1,2,4)\}$

column basis: $\{(1,2)\}$

$\dim(C(A))=\text{rank}(A)=1$

$\dim(C(A^T))=\text{rank}(A)=1$

$\dim(N(A))=n-r=2$

$\dim(N(A^T))=m-r=1$

$$B=\begin{bmatrix}1&2&4\\2&5&8\end{bmatrix}$$

row basis: $\{(1,2,4),(2,5,8)\}$

column basis: $\{(1,2),(2,5)\}$

4.

a) $\begin{bmatrix}1&0\\1&0\\0&1\end{bmatrix}$ b) not possible

11.

$r<m$ and $r=n$

$r<m$ and $r<n$

A^Ty is short and wide, which always has solutions

12.

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

$\dim(C(A^T))=r=1$,

then $\dim(N(A))=n-r=3-1=2$

14. row space, left nullspace

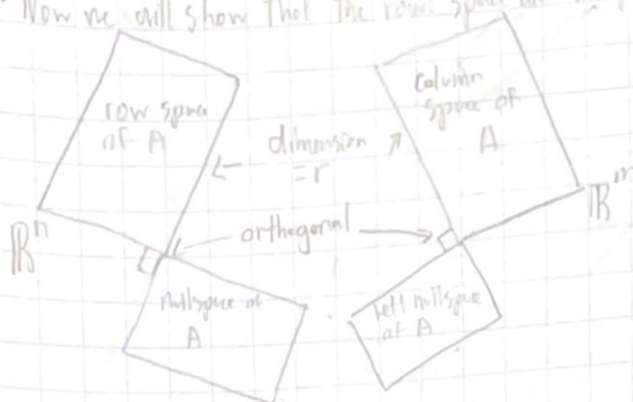
Chapter 4: Orthogonality

Chapter 4.1: Orthogonality of the Four Subspaces

- Two vectors are orthogonal when their dot products are zero: $v^Tw=v\cdot w=0$
- Orthogonal vectors also satisfy $\|v\|^2+\|w\|^2=\|v+w\|^2$

  $(v+w)^T(v+w)=v^Tv+w^Tw+2v^Tw=\|v\|^2+\|w\|^2+\underbrace{2v^Tw}_{\text{must be zero}}$
- The row space is perpendicular to the nullspace. Every row of A is perpendicular to every solution of $Ax=0$
- The column space is perpendicular to the left nullspace. When b is outside the column space, then this nullspace of A^T comes into its own. It contains the error $e=b-Ax$ in the "least-square" solution, a key application of linear algebra.

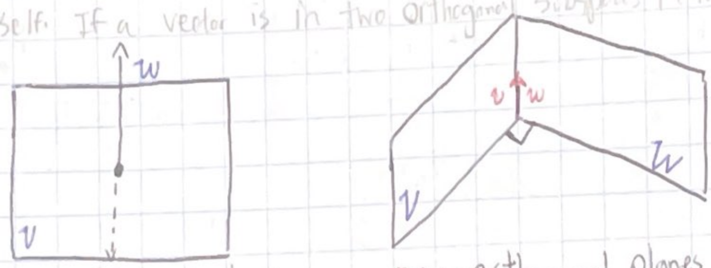
- Part one of the Fundamental Theorem gives the dimension of the subspaces.
- Now we will show that the row space and nullspace are orthogonal subspaces inside $\mathbb{R}^n$



- Two subspaces V and W are orthogonal if every vector v in V is perpendicular to every w in W. That is:

$$v^T w = v \cdot w = 0 \;\forall\, v \in V, w \in W$$

- Examples (see fig 4.1.1)
  - The floor of your room (extended to infinity) is a subspace V. The line where two walls meet is a subspace W. The subspaces are orthogonal
  - The two walls may look like it, but they are not orthogonal subspaces. The intersection line is in both V and W, and a line is not perpendicular to itself. If a vector is in two orthogonal subspaces it must be the zero vector



orthogonal line and plane　　non-orthogonal planes

Fig 4.1.1

- Orthogonality is impossible when dim V + dim W > dimension of the whole space
- Zero is the only point where the nullspace meets the row space. More than that, the nullspace and row space of A meet at 90°. This fact comes directly from $Ax=0$
  - Every vector x in the nullspace is perpendicular to every row of A because $Ax=0$

$$Ax=0 \Rightarrow \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} x = \begin{bmatrix} a_1 \cdot x \\ a_2 \cdot x \\ \vdots \\ a_n \cdot x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \;\forall\, x \in N(A)$$

Then, x is also perpendicular to all linear combinations of the rows

$$x \cdot (c_1 a_1 + c_2 a_2 + \dots c_n a_n) = c_1 \overbrace{(x \cdot a_1)}^{0} + c_2 \overbrace{(x \cdot a_2)}^{0} + \dots + c_n \overbrace{(x \cdot a_n)}^{0} = 0$$

  - The nullspace $N(A)$ and the row space are orthogonal subspaces of $\mathbb{R}^n$

- Example
  - The rows of A are perpendicular to $v=(1,1,-1)$

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

- Likewise, the columns of A and the left nullspace are orthogonal.
  - Every vector y in the nullspace of A^T is perpendicular to every column of A
  - The left nullspace $N(A^T)$ and the column space $C(A)$ are orthogonal in $\mathbb{R}^m$
  - It is proved similarly:

$$A^Ty=0 \Rightarrow \begin{bmatrix}(\text{column }1)^T\\ \vdots \\ (\text{column }n)^T\end{bmatrix}y=\begin{bmatrix}(\text{column }1)^T\cdot y\\ \vdots \\ (\text{column }2)^T\cdot y\end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix} \quad \forall y \in N(A^T)$$

Then y is perpendicular to all linear combination of the columns

$$[c_1(\text{column }1)+\dots+c_n(\text{column }n)]\cdot y = c_1(y\cdot \text{column }1)+\dots+c_n(y\cdot\text{column }n)=0$$

Orthogonal Complements

- The fundamental subspaces are more than just orthogonal (in pairs). Their dimensions are also right.
- Two lines could be perpendicular in $\mathbb{R}^3$ but those lines could not be the row space and nullspace of a 3 by 3 matrix. The lines have dimensions 1 and 1, summing to 2. The correct dimensions r and $n-r$ must sum to n.
  - The fundamental subspaces have dimensions 2 and 1 or 3 and 0. These subspaces are not only orthogonal, they are orthogonal complements.
- The orthogonal complement of a subspace V contains every vector perpendicular to V. It is noted as $V^\perp$ (pronounced "V perp")
- By this definition, the nullspace is the orthogonal complement of the row space since every x perpendicular to the rows satisfies $Ax=0$
- The reverse is also true. If V is orthogonal to the nullspace it must be in the row space. Otherwise we could add this V as an extra row without changing the nullspace. The row space would grow, which breaks the law $r+(n-r)=n$
  - Thus $N(A)^\perp = C(A^T)$ and $C(A^T)^\perp = N(A)$
- Fundamental Theorem of Linear Algebra, Part Two
  - $N(A)$ is the orthogonal complement of the row space $C(A^T)$ (in $\mathbb{R}^n$)
  - $N(A^T)$ is the orthogonal complement of the column space $C(A)$ (in $\mathbb{R}^m$)
- The point of complements is that every x can be split into a row space component x_r and a nullspace component x_n. When A multiplies x

$$Ax=A(x_r+x_n)=Ax_r+Ax_n=Ax_r+0=Ax_r=Ax$$

- Every vector goes to the column space.
  - The nullspace component goes to zero: $Ax_n = 0$
  - The row space component goes to the column space: $Ax_r = Ax$

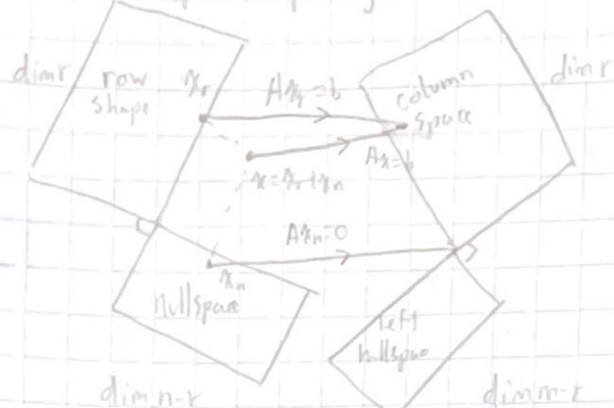


Fig 4.1.2

- Fig 4.1.2 shows the true action of A on $x = x_r + x_n$. A transforms (or "maps") x_r to the column space and the nullspace vector x_n to 0.
  - More than that, every vector b in the column space comes from one and only one vector in the row space. If $Ax_r = Ax_r'$, the difference $x_r - x_r'$ is in the nullspace. It is also in the row space since x_r and x_r' also did, and $x_r - x_r'$ is a linear combination. This difference $x_r' - x_r$ must be zero since the nullspace and row space are perpendicular. Therefore $x_r' = x_r$.
- There is an r by r invertible matrix hiding inside A if we throw away the two nullspaces. From the row space to the column space, A is invertible. The "pseudoinverse" will invert it in Section 7.3
- Example

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{contains the submatrix} \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

  - The other 11 zeroes are responsible for the nullspaces

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix} \quad \text{contains} \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \text{ in the pivot rows and columns}$$

Combining Bases from Subspaces

- Any n independent vectors in $\mathbb{R}^n$ must span $\mathbb{R}^n$. So they are a basis.
- Any n vectors that span $\mathbb{R}^n$ must be independent. So they are a basis
- Starting with the correct number of vectors, one property of a basis produces the other
- When the vectors go into the columns, the above two facts are:
  - If the n columns of A are independent, they span $\mathbb{R}^n$ and $Ax = b$ is solvable
  - If the n columns span $\mathbb{R}^n$ they are independent, $Ax = b$ has only 1 solution
- Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad \text{split } x = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \text{ into } x_r + x_n = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Problem Set 4.1

1.

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

basis for row space: $\{(1,2,3)\}$
basis for null space: $\{(-2,1,0),(-3,0,1)\}$

3.

c) $\begin{bmatrix} 1 & 2 & -3 \\ 2 & -3 & 1 \\ -3 & 5 & -2 \end{bmatrix}$

b) not possible
row space vectors should all be orthogonal to $(1,1,1)$
$(2,-3,5)\cdot(1,1,1) \neq 0$

8.
since $x = x_p + x_n$

$Ax = A(x_p + x_n) = Ax_r + Ax_n$
$= Ax_r$

13.

If $V^T W = 0$, then the corresponding columns of V and W are perpendicular, which imply the spans of the columns of V and W are orthogonal

15.
$p+q \leq n$

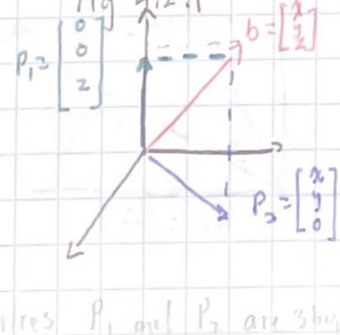
19.
Plane, line, L

24.
row 2 1 to n

Chapter 4.2 Projections

- Two introductory questions!
  1. What are projections of $b = (2,3,4)$ onto the z axis and the xy plane?
  2. What matrices produce these projections onto a line and plane?
- When b is projected onto a line, its projection p is the part of b along that line. If b is projected onto a plane, p is the part in that plane. The projection p is Pb. The projection matrix P multiplies b to give p.
- The projection onto the z-axis we'll call p_1 and the projection onto the xy-plane we'll call p_2. The image in your mind should be that of fig 4.2.1.
  $p_1 = (0,0,4)$, $p_2 = (2,3,0)$. These are the parts of b along the z-axis and xy-plane

fig 4.2.1



- The projection matrices P_1 and P_2 are 3 by 3. They multiply b with 3 components to produce p with 3 components.
  - Projection onto a line comes from a rank one matrix. Projection onto a plane comes from a rank two matrix.

- For the second question:

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- In this case the projections p_1 and p_2 are perpendicular. The xy plane and z-axis are orthogonal subspaces. More than that, they are orthogonal complements. Their dimensions add to $1+2=3$. Every vector b in the whole space can be expressed as a sum of its parts in the two subspaces.
  - The projection p_1 and p_2 are exactly those parts.

    The vectors give $p_1+p_2=b$. The matrices give $P_1+P_2=I$

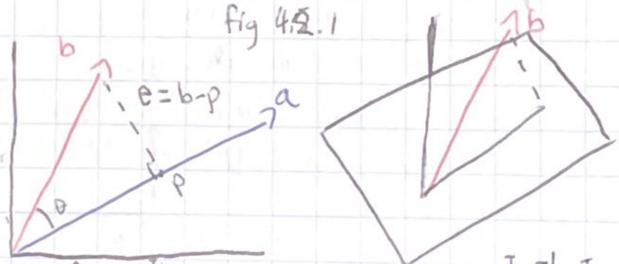
- We can use this to express x as x_r+x_n. The goal is to find the part p in each subspace and the corresponding projection matrix that produces $p=Pb$
  - Every subspace of $\mathbb{R}^m$ has its own m by m projection matrix
- To compute P, we need a good description of the subspace it projects onto. The best such description is a basis!
  - We put the basis vectors into the columns of A. Now we are projecting onto the column space of A.

    For z-axis: $A_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ For xy plane: $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}$ Both are bases. Both work.
  - We want to be able to project any b onto the column space of any m by n matrix.

Projection onto a Line

- A line goes through the origin in the direction of $a=(a_1, a_2, a_3, \dots, a_n)$. Along that line, we want the point p closest to $b=(b_1, b_2, b_3, \dots, b_n)$. The key to projection is orthogonality. The line from b to p is perpendicular to the vector a.
- The projection p is some scalar multiple of a. Call it $p=\hat{x}a$. Computing the scalar $\hat{x}$ will give the vector p. Then from the formula for p, we read off the projection matrix P.

fig 4.2.1



$p=\hat{x}a = \frac{a^Tb}{a^Ta}a$ $\qquad p = A\hat{x} = A(A^TA)^{-1}A^Tb = Pb$

- The line from b to p is denoted by a dotted line in fig 4.2.1
  - It is $e=b-p=b-\hat{x}a$, which is perpendicular to a.

$$a\cdot(b-\hat{x}a)=0 \Rightarrow a\cdot b - \hat{x}a\cdot a = 0$$

$$\Rightarrow \hat{x}a\cdot a = a\cdot b$$

$$\hat{x} = \frac{a\cdot b}{a\cdot a} = \frac{a^Tb}{a^Ta}$$

  - Then $p=\hat{x}a = \frac{a^Tb}{a^Ta}a$ is the projection of b onto the line through a.

- Special Cases
  - If $b = a$, then $\hat{x} = 1$. The projection a onto a is itself. $Pa = a$
  - If $b \perp a$, then $a^T b = 0$. The projection is $p = 0$
- Example
  - project $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find p

$$p = \frac{a^T b}{a^T a} a = \frac{5}{9} a = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \text{ and } e = b - p = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right)$$

- Now for the projection matrix:
  - What matrix is multiplying b in $p = Pb$?

$$p = a\hat{x} = a\frac{a^T b}{a^T a} = Pb \Rightarrow P = \frac{aa^T}{a^T a}$$

  - P is a column times a row (outer product). The column is a, the row is a^T, and then we divide by $a^T a$. The projection matrix is m by m, but its rank is one. We are projecting onto a one-dimensional subspace
- Example
  - Find the projection matrix $P = \frac{aa^T}{a^T a}$ onto the line through $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$$P = \frac{aa^T}{a^T a} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$$

Check:

$$p = Pb = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{9} \\ \frac{10}{9} \\ \frac{10}{9} \end{bmatrix}$$

- If the vector a is doubled, the matrix P stays the same. It still projects onto the same line (the line through $2a$). If the matrix is squared, $P^2 = P$. Projecting twice doesn't change anything, so $P^2 = P$.
- The matrix $I - P$ is another projection matrix. It produces the part of b perpendicular to a (the "e" part). Note that $(I-P)b = Ib - Pb = b - p$. When P projects onto one subspace, $I - P$ projects onto the perpendicular subspace. Here $I - P$ projects onto the plane perpendicular to a.
- Projection onto an n-dimensional subspace of $\mathbb{R}^m$ takes more effort.

Projection Onto a Subspace
- Start with n vectors $a_1, \ldots, a_n$ in $\mathbb{R}^m$. Assume these a's are linearly independent
- We want to find the combination $p = \hat{x}_1 a_1 + \ldots + \hat{x}_n a_n$ closest to a given vector b. We are projecting each b in $\mathbb{R}^m$ onto the subspace spanned by the a's to get p
- With $n = 1$, this is projection onto a line. The line is the column space of $A = [a_1]$ with only 1 column
- The combinations for p in $\mathbb{R}^m$ are the vectors Ax in the column space. We are looking to the particular combination $p = A\hat{x}$ closest to b.

- We compute projections onto n-dimensional subspaces in three steps as before: Find the vector $\hat{x}$, find the projection $p = A\hat{x}$, then find the matrix P.
- The dotted line in Fig 4.2.1 goes from b to the nearest point $A\hat{x}$ in the subspace. This error vector $b - A\hat{x}$ is perpendicular to the subspace.
- The error $b - A\hat{x}$ makes a right angle with all the vectors $a_1, \dots, a_n$. That gives us n equations

$$a_1^T(b - A\hat{x}) = 0 \quad \vdots \quad a_n^T(b - A\hat{x}) = 0 \qquad \text{or} \qquad \begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} \begin{bmatrix} b - A\hat{x} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

  - The matrix with those rows is A^T. The n equations are exactly $A^T(b - A\hat{x}) = 0$. Rewrite to $A^Tb - A^TA\hat{x} = 0 \Rightarrow A^TA\hat{x} = A^Tb$. This is the equation for $\hat{x}$ and the coefficient matrix is A^TA. Now we can find $\hat{x}$, p, and P.
- The combination $p = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n = A\hat{x}$ that is closest to b comes from:

$$A^T(b - A\hat{x}) = 0 \quad \text{or} \quad A^TA\hat{x} = A^Tb$$

  - This symmetric matrix A^TA is n by n. It is invertible if the a's are independent. The solution is $\hat{x} = (A^TA)^{-1}A^Tb$. Then the projection is:

$$p = A\hat{x} = A(A^TA)^{-1}A^Tb$$

  - The projection matrix P is then

$$P = A(A^TA)^{-1}A^T$$

- Looking at the respective equations for $n=1$:

$$\hat{x} = \frac{a^Tb}{a^Ta} \qquad p = a\frac{a^Tb}{a^Ta} \qquad P = \frac{aa^T}{a^Ta}$$

they are very similar. Instead of a^Ta, we have A^TA. Instead of dividing by the scalar, we invert the matrix: $(A^TA)^{-1}$ instead of $\frac{1}{a^Ta}$. Since $a_1, \dots, a_n$ are linearly independent, this inverse matrix is guaranteed to exist.
- The key step is $A^T(b - A\hat{x}) = 0$:
  1. Our subspace is the column space of A
  2. The error vector $b - A\hat{x}$ is perpendicular to that column space
  3. Therefore $b - A\hat{x}$ is in the nullspace of A^T (by the Fundamental Theorem, part 2). Then $A^T(b - A\hat{x}) = 0$
- The left nullspace is important in projection. It contains the error vector $e = b - A\hat{x}$. The vector b is split into the error $e = b - p$ and the projection p.

- Example
  - If $A=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}$ and $b=\begin{bmatrix}6\\0\\0\end{bmatrix}$, find $\hat{x}$, p, P

$$A^TA=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}=\begin{bmatrix}3&3\\3&5\end{bmatrix} \text{ and } A^Tb=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}$$

Now solve

$$A^TA\hat{x}=A^Tb \Rightarrow \begin{bmatrix}3&3\\3&5\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix} \text{ gives } \begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}5\\-3\end{bmatrix}$$

$$p=A\hat{x}=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}\begin{bmatrix}5\\-3\end{bmatrix}=\begin{bmatrix}5\\2\\-1\end{bmatrix}$$ The error is $e=b-p=\begin{bmatrix}1\\-2\\1\end{bmatrix}$

To find P, compute $P=A(A^TA)^{-1}A^T$

$$(A^TA)^{-1}=\frac{1}{\det(A^TA)}\begin{bmatrix}5&-3\\-3&3\end{bmatrix}=\frac{1}{6}\begin{bmatrix}5&-3\\-3&3\end{bmatrix} \text{ and } P=\frac{1}{6}\begin{bmatrix}5&2&-1\\2&2&2\\-1&2&5\end{bmatrix}$$

- Warning: $P=A(A^TA)^{-1}A^T$ is deceptive. $(A^TA)^{-1}\neq A^{-1}A^{T^{-1}}$. If we do that, then $P=AA^{-1}(A^T)^{-1}A^T=I$. This is not true.
  A is a rectangular and it has no inverse matrix
- A^TA is invertible if and only if A has linearly independent columns.
  - A^TA is a square matrix. For every matrix A, we will show that A^TA has the same nullspace as A. When the columns of A are linearly independent, its nullspace contains only the zero vector. Then A^TA with this same nullspace is invertible.
  - Let A be any matrix. If x is in its nullspace, then $Ax=0$. Multiplying by A^T gives $A^TAx=0$, so x is also in the nullspace of A^TA
  - Now we need to prove the opposite, that from $A^TAx=0$ we get $Ax=0$. We can't multiply by $(A^T)^{-1}$, which generally doesn't exist. Just multiply by x^T

    $(x^T)A^TAx \Rightarrow (Ax)^T(Ax)=0 \Rightarrow \|Ax\|^2=0$

  - This says if $A^TAx=0$ then Ax has length zero. Therefore $Ax=0$. Every vector x in one nullspace is in the other.
- When A has independent columns, A^TA is square symmetric and invertible
  - A^TA is (n by m) times (m by n). Then A^TA is square (n by n)
  - It is symmetric because $(A^TA)^T=A^T(A^T)^T=A^TA$

$$\underset{A^T}{\begin{bmatrix}1&1&0\\2&2&0\end{bmatrix}}\underset{A}{\begin{bmatrix}1&2\\1&2\\0&0\end{bmatrix}}=\underset{A^TA}{\begin{bmatrix}2&4\\4&8\end{bmatrix}} \qquad \underset{A^T}{\begin{bmatrix}1&1&0\\2&2&1\end{bmatrix}}\underset{A}{\begin{bmatrix}1&2\\1&2\\0&1\end{bmatrix}}=\underset{A^TA}{\begin{bmatrix}2&4\\4&9\end{bmatrix}}$$

dep. singular indep. invertible

Problem Set 4.2

1.

a)

$p = \frac{a^Tb}{a^Ta}a$

$= \frac{5}{3}(1,1,1)$

$= (\frac{5}{3}, \frac{5}{3}, \frac{5}{3})$

$e = b - p$

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

b)

$p = \frac{a^Tb}{a^Ta}a$

$= \frac{-11}{11}(-1,-3,-1)$

$= (1,3,1)$

$e = b - p$

$= (0,0,0)$

2.

a)

$p = \frac{a^Tb}{a^Ta}a$

$= \frac{\cos\theta}{\cos^2\theta + \sin^2\theta}(1,0)$

$= \cos\theta(1,0)$

$= (\cos\theta, 0)$

$e = b - p$

$= (0, \sin\theta)$

3.

a)

$P = \frac{aa^T}{a^Ta}$

$= \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

$= \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

b)

$P = \frac{aa^T}{a^Ta}$

$= \frac{1}{11}\begin{bmatrix} 1 & 3 & 1 \\ 3 & 9 & 3 \\ 1 & 3 & 1 \end{bmatrix}$

6.

$P_1 = \frac{a_1a_1^T}{a_1^Ta_1}$

$= \frac{1}{9}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{bmatrix}$

$P_2 = \frac{a_2a_2^T}{a_2^Ta_2}$

$= \frac{1}{9}\begin{bmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{bmatrix}$

$P_1P_2 = \frac{1}{81}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{bmatrix}$

$= \frac{1}{81}\begin{bmatrix} -8 & -8 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

9.

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

$A^TA = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$

$(A^TA)^{-1} = \frac{1}{4}\begin{bmatrix} 5 & -1 \\ -1 & 1 \end{bmatrix}$

$P = A(A^TA)^{-1}A^T = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 5 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$

$= \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$

$= \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$

11

a)

$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

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

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

$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\hat{x} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$

$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$

$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

$p = A\hat{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 3 \end{bmatrix}$

$= \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$

11

b)

$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

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

$A^Tb = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \end{bmatrix}$

$A^TA\hat{x} = A^Tb$

$\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \end{bmatrix}$

$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$

$p = A\hat{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

$e = b - p = (0, 0, 0)$

12.

$P = A(A^TA)^{-1}A^T$

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

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

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

$Pb = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 8 \\ 8 \\ 12 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

13.

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

$p=(1,2,3,0)$

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

14.

$$A=\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix} \quad b=\begin{bmatrix}0\\2\\4\end{bmatrix}$$

$$A^TA=\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}=\begin{bmatrix}5&2\\2&5\end{bmatrix}$$

$$A^Tb=\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}\begin{bmatrix}0\\2\\4\end{bmatrix}=\begin{bmatrix}10\\4\end{bmatrix}$$

$A^TA\hat{x}=A^Tb$

$$\begin{bmatrix}5&2\\2&5\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}10\\4\end{bmatrix}$$

$$\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}$$

$A\hat{x}=p$

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

$$(A^TA)^{-1}=\frac{1}{21}\begin{bmatrix}5&-2\\-2&5\end{bmatrix}$$

$P=A(A^TA)^{-1}A^T$

$$=\frac{1}{21}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}\begin{bmatrix}5&-2\\-2&5\end{bmatrix}\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}$$

$$=\frac{1}{21}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}\begin{bmatrix}-2&1&10\\5&8&-4\end{bmatrix}$$

$$=\frac{1}{21}\begin{bmatrix}5&8&-4\\8&17&-8\\-4&2&20\end{bmatrix}$$

17.

$(I-P)^2=I^2-IP-PI+P^2$

$=I-P-P+P$

$=I-P$

left nullspace

21. itself

$P^2=[A(A^TA)^{-1}A^T][A(A^TA)^{-1}A^T]$

$=A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T$

$=A(A^TA)^{-1}A^T$

$=P$

27. A^T, A

23.

$P=I_n$

A covers the entire $\mathbb{R}^n$ space, so a projection of any vector onto $\mathbb{R}^n$ is itself. $E=0$

Chapter 4.3: Least Squares Approximations

- It often happens that $Ax=b$ has no solution, and usually it's because there are too many equations. The matrix has more rows than columns. There are more equation than unknowns $(m>n)$. The n columns spans a small part of $\mathbb{R}^m$ and b happens to be outside the column space.
- We can't always get the error $e=b-Ax$ down to zero, but it is practical to minimize it. When the length of e is minimized $\hat{x}$ is a least squares solution.
- When $Ax=b$ has no solution, multiply by A^T and solve $A^TA\hat{x}=A^Tb$
- Example 1: Fitting a straight line
  - Given $(0,6),(1,0),(2,0)$; no straight line $b=C+Dt$ goes through the three points. We are asking for two numbers C and D that satisfies 3 equations
    - $t=0$ The first point is on the line $b=C+Dt$ if $C+D\cdot 0=6$
    - $t=1$ The second point is on the line $b=C+Dt$ if $C+D\cdot 1=0$
    - $t=2$ The third point is on the line $b=C+Dt$ if $C+D\cdot 2=0$

- We have three equations

$$\begin{cases} C + 0D = 6 \\ C + 1D = 0 \\ C + 2D = 0 \end{cases} \quad \text{Rewrite} \quad \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad Ax = b \text{ is not solvable}$$

Find $\hat{x}$ via $A^TA\hat{x} = A^Tb$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad ; \quad A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \Rightarrow \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} C \\ D \end{bmatrix}$$

$\therefore$ $5 - 3t$ is the line of best fit for those three lines

Minimizing the Error

- How do we minimize $e = b - Ax$, we can find $\hat{x}$ (the best choice) by geometry, linear algebra, or calculus
- By geometry
  - Every Ax lies within the column space of A. In that space, we look for the point closest to b, which is the projection p.
  - The best choice for $A\hat{x}$ is p. The smallest possible error is $e = b - p$.
  - The three points of height (p_1, p_2, p_3) do lie on a line because p is in the column space. In fitting a straight line, $\hat{x}$ gives the best choice for (C, D)
- By algebra
  - Every vector b splits into two parts. The part in the column space is p. The perpendicular part in the nullspace of A^T is e.
  - We cannot solve $Ax = b$, but we can solve $A\hat{x} = p$ since p is in the column space

    $Ax = b = p + e$ is impossible ; $A\hat{x} = p$ is solvable

  - The solution to $A\hat{x} = p$ leaves the least possible error (which is e)
    Squared length for any x: $\|Ax - b\|^2 = \|Ax - p\|^2 + \|e\|^2$
    This is from the pythagorean theorem. The vector $Ax - p$ in the column space is perpendicular to e in the left nullspace.
  - We reduce $Ax - p$ to zero by choosing x to be $\hat{x}$. This leaves the smallest possible error $e = (e_1, e_2, e_3)$
- We minimize the squared length of $Ax - b$, hence the name
- The least squares solution $\hat{x}$ makes $E = \|Ax - b\|^2$ as small as possible

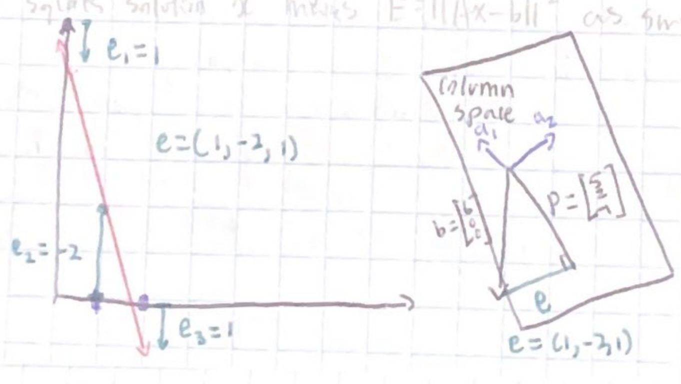


fig 4.3.1
A line of best fit vs. its linear algebra counterpart

- fig 4.3.1 shows the closest line, which minimizes the vertical distance between corresponding points on the line.
  It also shows the linear algebra approach
- By Calculus
  - We want to minimized the error function $E = e_1^2 + e_2^2 + e_3^2$

$$E = \|Ax - b\|^2 = (C + 0D - 6)^2 + (C + D - 0)^2 + (C + 2D - 0)^2$$

$$= (C-6)^2 + (C+D)^2 + (C+2D)^2$$

  - Since we have 2 unknowns (C and D), we have two derivatives.
    These are partial derivatives with respect to each unknown which give us two equations when set to zero

$$\frac{\partial E}{\partial C} = 2(C-6) + 2(C+D) + 2(C+2D) \quad \text{set } \frac{\partial E}{\partial C} = 0 \implies 3C + 3D = 6$$

$$\frac{\partial E}{\partial D} = 2(C+D) + 4(C+2D) \quad \text{set } \frac{\partial E}{\partial D} = 0 \implies 3C + 5D = 0$$

Write equations as matrix

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ is our } A^T A$$

$$A^T A \quad \hat{x} \quad = A^T b$$

  - These equations are the same as the ones we got through linear algebra
    The partial derivatives of $\|Ax - b\|^2$ are zero when $A^T A \hat{x} = A^T b$

The Big Picture

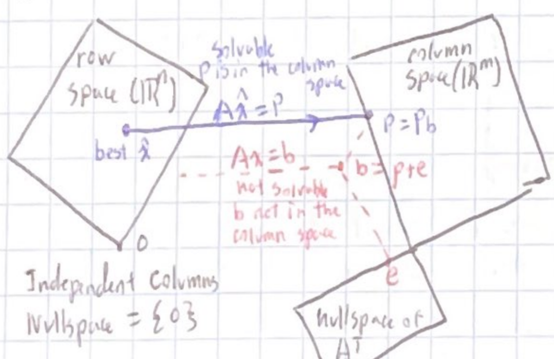


Fig 4.3.2
The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$

- In previous chapters, we split x into $x_r + x_n$. There were many solutions to $Ax = b$. Now, we do the opposite. There are no solutions to $Ax = b$. Instead of splitting x, we are splitting b into p (in the column space) and an orthogonal e (in the left nullspace). Now we solve $A\hat{x} = p$; the error $e = b - p$ is unavoidable.
- Notice how the nullspace N(A) is very small; just Z. This is because the columns are independent. Then $A^T A$ is invertible. The equation $A^T A \hat{x} = A^T b$ describes the best vector $\hat{x}$. The error has $A^T e = 0$

Fitting a Straight Line

- Fitting a clear line is the clearest application of least squares. It starts with $m > 2$ points. At times $t_1, \dots, t_m$ those m points are at heights $b_1, \dots, b_m$. The best line $C + Dt$ misses the points by vertical distances $e_1, \dots, e_m$. Least squares minimizes $e_1^2 + \dots + e_m^2$.
- A line goes through the m points when we exactly solve $Ax = b$. Generally we can't do it.
- To fit the m points, we try to solve m equations with two unknowns each.

$$Ax = b \text{ is } \begin{cases} C + Dt_1 = b_1 \\ \vdots \\ C + Dt_m = b_m \end{cases} \quad \text{with } A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$$

- The column space is so thin that b is almost certainly outside of it. When b happens to be in the column space, the points lie on a line and $b = p$. Then $Ax = b$ is solvable and the errors are $e = (0, \dots, 0)$.
- The closest line $C + Dt$ has heights $p_1, \dots, p_m$ with errors $e_1, \dots, e_m$. Solve $A^TA\hat{x} = A^Tb$ for $\hat{x} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$
  - The two columns of A are independent unless all times t_i are the same.

Dot-product Matrix $A^TA = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum_1^m t_i \\ \sum_1^m t_i & \sum_1^m t_i^2 \end{bmatrix}$

On the right side:

$A^Tb \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_1^m b_i \\ \sum_1^m t_i b_i \end{bmatrix}$

Making the substitutions

The line $C + Dt$ minimizes $e_1^2 + \dots + e_m^2 = \|Ax - b\|^2$ when $A^TA\hat{x} = A^Tb$

$$\begin{bmatrix} m & \sum_1^m t_i \\ \sum_1^m t_i & \sum_1^m t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum_1^m b_i \\ \sum_1^m t_i b_i \end{bmatrix}$$

- The vertical errors at the m points of the line are the components of $e = b - p$. This error vector (the residual) $b - A\hat{x}$ is perpendicular to the columns of A.
- The best $\hat{x} = (C, D)$ minimizes the sum of the errors squared

$$E = \|Ax - b\|^2 = (C + Dt_1 - b_1)^2 + \dots + (C + Dt_m - b_m)^2$$

  - When the derivatives $\frac{\partial E}{\partial D}$ and $\frac{\partial E}{\partial C}$ are set to zero, it produces $A^TA\hat{x} = A^Tb$
- Other least squares can have more parameters. Fitting a parabola takes 3 coefficients. In general we are fitting m data points by n parameters $x_1, \dots, x_n$. The matrix has n columns and $n < m$.
- The derivative of a square is linear, which is why we can approach this with calculus and linear algebra

· Example 2

- A has orthogonal columns when the measurement times t_i add to zero. Suppose $b = 1, 2, 4$ at times $t = -2, 0, 2$. Those times add to zero. The columns of A here a zero dot product.

$$\begin{cases} C + D(-2) = 1 \\ C + D(0) = 2 \\ C + D(2) = 4 \end{cases} \quad \text{or } Ax = b \Rightarrow \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \text{ is } \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

- Notice how A^TA is now diagonal. We can solve for $C = \frac{7}{3}$ and $D = \frac{3}{4}$ separately. The diagonal matrix A^TA with entries $m = 3$ and $\sum_1^3 t_i^2 = t_1^2 + t_2^2 + t_3^2 = 8$ is virtually as good as the identity matrix.
- Orthogonal columns are so useful it is worth moving the time origin to produce them. To do that, we subtract the average time $\hat{t} = (t_1 + t_2 + \dots + t_m)/m$ from each time. The shifted times $T_i = t_i - \hat{t}$ sum to $\sum_1^m T_i = m\hat{t} - m\hat{t} = 0$, which makes A^TA diagonal. A^TA's entries are now m and $\sum_1^m T_i^2$, giving the best C and D direct formulas

$$T_i = t_i - \hat{t} \qquad C = \frac{b_1 + \dots + b_m}{m} \quad \text{and} \quad D = \frac{b_1T_1 + \dots + b_mT_m}{T_1^2 + \dots + T_m^2}$$

and the best line is $C + DT$ or $C + D(t - \hat{t})$.
- This time shift which makes A^TA diagonal is an example of the Gram-Schmidt process: orthogonalize the columns in advance.

Fitting a Parabola

- A parabola $b = C + Dt + t^2$, even though it is quadratic, it coefficients are still linear.
- Problem: Fit heights $b_1, \dots, b_m$ at times $t_1, \dots, t_m$ by a parabola $C + Dt + Et^2$
Solution: with $m > 3$ points, the m equations for an exact fit are generally unsolvable

$$\begin{cases} C + Dt_1 + Et_1^2 = b_1 \\ \vdots \\ C + Dt_m + Et_m^2 = b_m \end{cases} \quad \text{or } Ax = b \Rightarrow \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}\begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Least Squares: The closest parabola $C + Dt + Et^2$ chooses $\hat{x} = (C, D, E)$ to satisfy the three normal equations $A^TA\hat{x} = A^Tb$

- The column space of A has dimension 3. We project b onto the column space of A. Our error is $e_i = b_i - C - Dt_i - Et_i^2$. Total error $= \sum_1^m e_i^2 = \sum_1^m (b_i - C - Dt_i - Et_i^2)^2$
- By calculus, we take partial derivatives with respect to C, D, and E and solve the 3 by 3 system generated

- Example 3
  - Given the points $(0,6),(1,0),(2,0)$

$$\begin{cases} C + D(0) + (0)^2 = 6 \\ C + D(1) + E(1)^2 = 0 \\ C + D(2) + E(2)^2 = 0 \end{cases} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x = \begin{bmatrix} 6 \\ -9 \\ 3 \end{bmatrix}$$

we can solve this directly since $m=3$.

Problem Set 4.3

1.

$$\begin{cases} C + (0)D = 0 \\ C + (1)D = 8 \\ C + (3)D = 8 \\ C + (4)D = 20 \end{cases} \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

2.

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 13 \\ 17 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \Rightarrow \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$$

$$\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \Rightarrow b = 1 + 4t$$

4.

$$E = \|Ax - b\|^2 = (C + (0)D - 0)^2 + (C + (1)D - 8)^2 + (C + (3)D - 8)^2 + (C + (4)D - 20)^2$$

$$\frac{\partial E}{\partial C} = 2C + 2(C + D - 8) + 2(C + 3D - 8) + 2(C + 4D - 20) = 0$$

$$= 8C + 16D - 72 = 0$$

$$4C + 8D = 36$$

$$\frac{\partial E}{\partial D} = 2(C + D - 8) + 6(C + 3D - 8) + 8(C + 4D - 20)$$

$$= 16C + 52D - 224 = 0$$

$$8C + 26D = 112$$

5.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [C] = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 4 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = [4]$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \\ 8 \\ 4 \end{bmatrix} = 20$$

$$4C = 20$$
$$C = 5$$

6.

$$\hat{x} = \frac{a^Tb}{a^Ta} = \frac{36}{4} = 9$$

$$p = \hat{x}a = (9,9,9,9)$$
$$e = b - p = (0,8,8,20) - (9,9,9,9)$$
$$= (-9,-1,-1,11)$$

9.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \\ 400 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} = \begin{bmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \\ 400 \end{bmatrix} \Rightarrow \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{4}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$y = \frac{2}{3}x^2 + \frac{4}{3}x + 2$$

12. a)

$$\hat{x} = \frac{a^Tb}{a^Ta}$$

$$= \frac{b_1 + b_2 + \dots + b_m}{m}$$

b)

$$P = a\hat{x}$$

$$= \begin{bmatrix} \frac{b_1+b_2+\dots+b_m}{m} \\ \vdots \\ \frac{b_1+b_2+\dots+b_m}{m} \end{bmatrix}$$

$e = b - p$

$$= \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} - \begin{bmatrix} \frac{b_1+\dots+b_m}{m} \\ \vdots \\ \frac{b_1+\dots+b_m}{m} \end{bmatrix}$$

$$= \begin{bmatrix} b_1 - \frac{b_1+\dots+b_m}{m} \\ \vdots \\ b_m - \frac{b_1+\dots+b_m}{m} \end{bmatrix}$$

$$\|e\|^2 = \sum_1 (b_i - \bar{b})^2$$

17.

$$\vec{PQ} = (y-x, 3y-x, -1-x)$$

$$\vec{PQ} \cdot (1,1,1) = 0$$

$$y - x + 3y - x - 1 - x = 0$$

$$-3x + 4y = 1$$

$$\vec{PQ} \cdot (1,3,0) = 0$$

$$y - x + 9y - 3x = 0$$

$$-4x + 10y = 0$$

$$x = \frac{10}{4}y$$

$$-3\left(\frac{10}{4}y\right) + 4y = 1$$

$$y = -\frac{2}{7}$$

$$x = -\frac{5}{7}$$

13.

$$(A^TA)^{-1}A^Te = (A^TA)^{-1}A^Tb - (A^TA)^{-1}A^TAx$$

$$= \hat{x} - x$$

17.

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix} \text{ no solution}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix} = \begin{bmatrix} 35 \\ 42 \end{bmatrix}$$

$$A^TAx = A^Tb \Rightarrow \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 35 \\ 42 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 3 & 2 \\ 0 & \frac{14}{3} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 35 \\ \frac{56}{3} \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 9 \\ 4 \end{bmatrix} \quad y = 9 + 4x$$

18.

$$p = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 13 \\ 17 \end{bmatrix}$$

$$e = b - p = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix} - \begin{bmatrix} 5 \\ 13 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ -6 \\ 14 \end{bmatrix}$$

$$P = A(A^TA)^{-1}A^T$$

$$(A^TA)^{-1} = \frac{1}{14}\begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$$

$$P = \frac{1}{14}\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

$$= \frac{1}{14}\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 8 & 4 & 2 \\ -5 & 1 & 4 \end{bmatrix}$$

$$= \frac{1}{14}\begin{bmatrix} 13 & 3 & -2 \\ 3 & 5 & 6 \\ -2 & 6 & 10 \end{bmatrix}$$

$$Pe = \frac{1}{14}\begin{bmatrix} 13 & 3 & -2 \\ 3 & 5 & 6 \\ -2 & 6 & 10 \end{bmatrix} \begin{bmatrix} 2 \\ -6 \\ 14 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} \; \end{bmatrix}$$

Chapter 4.4: Orthogonal Bases and Gram-Schmidt

- This section has two goals: to illustrate the usefulness of orthogonal bases in finding $\hat{x}$, p, and P. The dot products are zero, so A^TA becomes diagonal. Secondly, is to construct orthogonal vectors. We will pick combinations of the original vectors to produce right angle. The original vectors are columns of A, probably not orthogonal. Orthogonal vectors will be columns of a new matrix Q.
- The vectors $q_1, q_2, \ldots, q_n$ are orthogonal when their dot products $q_i \cdot q_j$ are zero. More specifically $q_i^T q_j = 0$ when $i \neq j$. If we divide all vectors by its length to get unit vectors, we get orthogonal unit vectors. Then this basis is called orthonormal.
- The vectors $q_1, \ldots, q_n$ are orthonormal if:

$$q_i^T q_j = \begin{cases} 0 & \text{when } i \neq j \text{ (orthogonal vectors)} \\ 1 & \text{when } i = j \text{ (unit vectors: } \|q_i\| = 1) \end{cases}$$

A matrix with orthonormal columns is assigned the letter Q

- The matrix Q is easy to work with because $Q^TQ = I$. Q does not have to be square.

$$Q^TQ = \begin{bmatrix} - & q_1^T & - \\ - & q_2^T & - \\ & \vdots & \\ - & q_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- When Q is square, $Q^TQ = I$ means $Q^T = Q^{-1}$: transpose = inverse.
- If the columns are only orthogonal, dot products still give a diagonal matrix.
- $Q^TQ = I$ even when Q is rectangular, but it is only a left-inverse. For square matrices, we have $Q^TQ = QQ^T = I$. The rows of a square Q is also orthonormal columns. We call a square Q an orthogonal/orthonormal matrix
- Here are three examples of orthogonal matrices:

1. Rotation
   - Q rotates every vector in the plane clockwise by the angle θ

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ and } Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

   - These columns give an orthonormal basis for $\mathbb{R}^2$
   - Q^T rotates by $-\theta$

2. Permutation
   - Q changes the order of the entries of a vector

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

   - The inverse of any permutation matrix is its transpose and vice versa

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

3. Reflection

- If u is any unit vector, set $Q = I - 2uu^T$. Then $Q^T = Q^{-1} = Q$

$$Q^T = I - 2uu^T = Q \quad \text{and} \quad Q^TQ = I - 4uu^T + 4uu^Tuu^T = I$$

- Reflection matrices $I - 2uu^T$ are symmetric and also orthogonal. Squaring them gets the identity matrix $Q^2 = Q^TQ = I$. Reflecting twices gets back the original vector
- Example:

Given $u = (1,0)$

$$I - 2uu^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = Q$$

$\rightarrow$ Q reflects $(x,0)$ across the y axis to $(-x,0)$

- Multiplication by any orthogonal matrix leaves lengths and angles unchanged.
- If Q has orthonormal columns $(Q^TQ = I)$, it leaves lengths unchanged:
  - Same length: $\|Qx\| = \|x\| \; \forall x \in V$
  - Dot products preserved: $(Qx)^T(Qy) = x^TQ^TQy = x^Ty$
- Orthogonal matrices are very useful in computation because it keeps numbers relatively small.

Projections using Orthogonal bases: Q replaces A

- In all our formulas for $\hat{x}$, p, and P onto subspaces, we see A^TA. Imagine if they were orthonormal: Then Q^TQ simplifies to I. We have:

$$Q^TQ\hat{x} = Q^Tb \Rightarrow \hat{x} = Q^Tb, \quad p = Q\hat{x} \quad P = Q(Q^TQ)^{-1}Q^T = QQ^T$$

  - The least squares solution of $Qx = b$ is $\hat{x} = Q^Tb$. The project matrix is $P = QQ^T$

$$P = QQ^Tb = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix}\begin{bmatrix} q_1^Tb \\ \vdots \\ q_n^Tb \end{bmatrix} = q_1(q_1^Tb) + \dots + q_n(q_n^Tb)$$

- When Q is square, the subspace is the whole space. Then $Q^T = Q^{-1}$ and $\hat{x} = Q^Tb = Q^{-1}b$ and we have an exact solution. In this case $P = QQ^T$ because b is being projected onto a subspace containing it.
  - When $p = b$, our formula constructs b out of one dimensional projections

$$q_i\left(\frac{q_i^Tb}{q_i^Tq_i}\right) = p \quad q_i^Tq_i = 1 \text{ because } \|q_i\| = 1. \text{ Therefore, we get } q_i(q_i^Tb) = p$$

- This is the foundation of famous transforms, like the fourier transform, which break vectors/functions into perpendicular pieces. The inverse transform then builds these pieces together

• Example 4:

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$$

The separate projection of $b = (0,0,1)$ onto q_1 and q_2 and q_3 are p_1 and p_2 and p_3.

$p_1 = q_1(q_1^T b) = \frac{2}{3}q_1$ $\quad p_2 = q_2(q_2^T b) = \frac{2}{3}q_2$ $\quad p_3 = q_3(q_3^T b) = -\frac{1}{3}q_3$

The sum of the first two is the projection of b onto the plane of q_1 and q_2.
The sum of all three is the projection onto the whole space, which is b itself

$$b = P_1 + P_2 + P_3 \qquad \frac{2}{3}q_1 + \frac{2}{3}q_2 - \frac{1}{3}q_3 = \frac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = b$$

The Gram-Schmitt Process

- Orthonormal vectors "uncouple" the projections. A projection can be though of as the "part" of a vector in a subspace. If our vectors are orthogonal to each other, no vector "bleeds into" another.
- How do we create orthonormal vectors?
  - We start with three independent vectors a, b, c. We intend to construct three orthogonal vectors A, B, and C. Then we divide by the lengths to get unit vectors.
  - Gram-Schmit: Begin by choosing $A = a$. The next direction B must be perpendicular to A. We start with b and subtract its projection along A. This leaves us the perpendicular part, the vector B (the error vector e)
  - A and B are then orthogonal. $A^T B = A^T b - A^T b = 0$.
  - The third direction starts with C. This is not a combination of A and B (because C is not a combination of a and b). Most likely C is not perpendicular to A and B, so we subtract off its components in those 2 directions to get C

$$A = a$$

$$B = b - \mathrm{proj}_A(b) = b - \frac{A^T b}{A^T A}A$$

$$C = c - \mathrm{proj}_A(c) - \mathrm{proj}_B(c) = c - \frac{A^T c}{A^T A}A - \frac{B^T c}{B^T B}B$$

$$U_k = V_k - \sum_{j=1}^{k-1} \mathrm{proj}_{U_j}(V_k) = V_k - \sum_{j=1}^{k-1} \frac{U_j^T V_k}{U_j^T U_j}$$

- Example

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$$

$A = a \qquad q_1 = \frac{A}{\|A\|} = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0)$

$B = b - \frac{A^Tb}{A^TA}A = b - \frac{2}{2}A = (1, 1, -2) \qquad q_2 = \frac{B}{\|B\|} = (\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{3})$

$C = c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B = c - \frac{6}{2}A + \frac{6}{6}B = (1,1,1) \qquad q_3 = \frac{C}{\|C\|} = (\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3})$

The Factorization $A = QR$

- We started with a matrix A with columns a, b, c. We ended with a matrix Q with columns q_1, q_2, q_3. Since the vectors a, b, c are combinations of q_1, q_2, q_3 and vice versa, there must be a third matrix connecting A to Q. This matrix is a triangular R in $A = QR$.
- The first step is $q_1 = \frac{a}{\|a\|}$ (other vectors not involved). Then we iterate, subtract from each new vector its projections onto previous vectors. This non-involvement of later vectors is the key point of Gram-Schmidt.
- The vectors a, A, q_1 are along one line
- The vectors a, b, A, B, q_1, q_2 are in the same plane
- The vectors $a, b, c, A, B, C, q_1, q_2, q_3$ are in one subspace ($\mathbb{R}^3$)
- At every step $a_1, \dots, a_k$ are combinations of $q_1, \dots, q_k$. Later q's are not involved. Therefore the connecting matrix R is triangular and we have $A = QR$

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^Ta & q_1^Tb & q_1^Tc \\ & q_2^Tb & q_2^Tc \\ & & q_3^Tc \end{bmatrix} \quad \text{or} \quad A = QR$$

- From independent vectors $a_1, \dots, a_n$, Gram-Schmidt constructs orthonormal vectors $q_1, \dots, q_n$. The matrices with these columns satisfy $A = QR$. Then $R = Q^TA$ is upper triangular because the later q's are orthogonal to earlier a's.
- From the example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & -\frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ & \sqrt{6} & -\sqrt{6} \\ & & \sqrt{3} \end{bmatrix}$$

- Any m by n matrix A with independent columns can be factored into QR. The m by n matrix Q has orthonormal columns and the matrix R is upper triangular with a positive diagonal.
- For least squares, $A^TA = R^TQ^TQR = R^TR$ and the least squares equation simplifies $A^TA\hat{x} = A^Tb \Rightarrow R^TR\hat{x} = R^TQ^Tb \Rightarrow R\hat{x} = Q^Tb$ or $\hat{x} = R^{-1}Q^Tb$.
- We solve $R\hat{x} = Q^Tb$ by back substitution, which is really fast. The real cost is mn^2 multiplications for the Gram-Schmidt process

Problem Set 4.4

11

a)

$A = a = (1,3,4,5,7)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b - \frac{100}{100}A$

$= (-7,3,4,-5,1)$

$q_1 = \frac{1}{100}A = (\frac{1}{100}, \frac{3}{100}, \frac{1}{25}, \frac{1}{20}, \frac{7}{100})$

$q_2 = \frac{1}{100}B = (-\frac{7}{100}, \frac{3}{100}, \frac{1}{25}, -\frac{1}{20}, \frac{1}{100})$

b)

$Q^TQ\hat{x} = Q^Tb$

$\hat{x} = Q^Tb$

$$\hat{x} = \begin{bmatrix} \frac{1}{100} & \frac{3}{100} & \frac{1}{25} & \frac{1}{20} & \frac{7}{100} \\ -\frac{7}{100} & \frac{3}{100} & \frac{1}{25} & \frac{1}{20} & \frac{1}{100} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{100} \\ -\frac{7}{100} \end{bmatrix}$$

$$p = A\hat{x} = \begin{bmatrix} \frac{1}{100} & -\frac{7}{100} \\ \frac{3}{100} & \frac{3}{100} \\ \frac{1}{25} & \frac{1}{25} \\ \frac{1}{20} & \frac{1}{20} \\ \frac{7}{100} & \frac{1}{100} \end{bmatrix} \begin{bmatrix} \frac{1}{100} \\ -\frac{7}{100} \end{bmatrix} = \begin{bmatrix} \frac{1}{200} \\ \frac{1}{500} \\ \\ \\ \\ \end{bmatrix}$$

15.

$a = (1,2,-2)$

$b = (1,-1,4)$

$A = a = (1,2,-2)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b - \frac{-9}{9}A$

$= b + A$

$= (2,1,2)$

$q_1 = (\frac{1}{3}, \frac{2}{3}, -\frac{2}{3})$

$q_2 = (\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$

$q_3 = (\frac{2}{3}, -\frac{2}{3}, -\frac{1}{3}) = q_1 \times q_2$

$q_3 \in N(A^T)$

$$Q^Tb = \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} -9 \\ 18 \end{bmatrix} \quad R = Q^{-1}A$$

$Q^TA = ~~Q^TQR~~$

$Ax = b$

$Rx = Q^Tb$

$R = Q^TA$

$$= \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 9 & -9 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 3 & -3 \\ 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -3 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -9 \\ 18 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

17.

$proj_a(b) = \frac{a^Tb}{a^Ta}a = \frac{9}{3}a = 3a = (3,3,3)$

$e = b - p = (1,3,5) - (3,3,3) = (-2,0,2)$

$q_1 = \frac{1}{\sqrt{3}}(1,1,1)$

$q_2 = \frac{1}{2\sqrt{2}}(-2,0,2)$

18.

$A = a = (1,-1,0,0)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b + \frac{1}{2}A$

$= (\frac{1}{2}, \frac{1}{2}, -1, 0)$

$C = c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B$

$= c - 0A + \frac{2}{3}B$

$= (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1)$

19.

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

$a_1 = (-1,2,2),\ a_2 = (1,1,4)$

$v_1 = a_1 = (-1,2,2)$

$v_2 = a_2 - \frac{a_1^Ta_2}{a_1^Ta_1}a_1$

$= a_2 - \frac{9}{9}a_1$

$= a_2 - a_1$

$= (2,-1,2)$

$q_1 = \frac{1}{\|v_1\|}v_1 = \frac{1}{\sqrt{9}}v_1$

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

$q_2 = \frac{1}{\|v_2\|}v_2 = \frac{1}{\sqrt{9}}v_2$

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

$$\rightarrow Q = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

$A = QR$

$A^TA = (QR)^T(QR)$

$= R^TQ^TQR$

$= R^TR$

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

$$= \begin{bmatrix} 9 & 9 \\ 9 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 9 \\ 0 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & \\ & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$$

21.

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

$a_1 = (1,1,1,1)$

$a_2 = (-2,0,1,3)$

$v_1 = a_1 = (1,1,1,1)$

$$v_2 = a_2 - \frac{v_1^T a_2}{v_1^T v_1} v_1$$

$$= a_2 - \frac{2}{4} v_1$$

$$= a_2 - \tfrac{1}{2} v_1$$

$$= (-2,0,1,3) - \tfrac{1}{2}(1,1,1,1)$$

$$= \left(-\tfrac{5}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{5}{2}\right)$$

$$q_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{2}(1,1,1,1) = \left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right)$$

$$q_2 = \frac{1}{\|v_2\|} v_2 = \frac{1}{\sqrt{13}}\left(-\tfrac{5}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{5}{2}\right) = \left(-\tfrac{5\sqrt{13}}{26}, -\tfrac{\sqrt{13}}{26}, \tfrac{\sqrt{13}}{26}, \tfrac{5\sqrt{13}}{26}\right)$$

$$Q = \begin{bmatrix} \frac{1}{2} & -\frac{5\sqrt{13}}{26} \\ \frac{1}{2} & -\frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{5\sqrt{13}}{26} \end{bmatrix}$$

$$Q^T Q \hat{x} = Q^T b$$

$$\hat{x} = Q^T b$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{5\sqrt{13}}{26} & -\frac{\sqrt{13}}{26} & \frac{\sqrt{13}}{26} & \frac{5\sqrt{13}}{26} \end{bmatrix} \begin{bmatrix} -4 \\ -3 \\ 3 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -2 \\ \sqrt{13} \end{bmatrix}$$

$$p = Q\hat{x}$$

$$= \begin{bmatrix} \frac{1}{2} & -\frac{5\sqrt{13}}{26} \\ \frac{1}{2} & -\frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{5\sqrt{13}}{26} \end{bmatrix} \begin{bmatrix} -2 \\ \sqrt{13} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{7}{2} \\ -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{bmatrix}$$

23.

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

$a_1 = (1,0,0)$

$a_2 = (2,0,3)$

$a_3 = (4,5,6)$

$v_1 = a_1 = (1,0,0)$

$$v_2 = a_2 - \frac{v_1^T a_2}{v_1^T v_1} v_1$$

$$= a_2 - \frac{2}{1} v_1$$

$$= a_2 - 2v_1$$

$$= (0,0,3)$$

$$v_3 = a_3 - \frac{v_1^T a_3}{v_1^T v_1} v_1 - \frac{v_2^T a_3}{v_2^T v_2} v_2$$

$$= a_3 - \frac{4}{1} v_1 - \frac{18}{9} v_2$$

$$= a_3 - 4v_1 - 2v_2$$

$$= (0,5,0)$$

$q_1 = (1,0,0)$

$q_2 = (0,0,1)$

$q_3 = (0,1,0)$

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

$A = QR$

$R = Q^{-1}A = Q^T A$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

30.

$$W = \frac{1}{4} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} = \frac{1}{4} Q$$

$$W^{-1} = 4Q^T$$

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

Chapter 5: Determinants

Chapter 5.1: The Properties of Determinants

- The determinant of a square matrix, denoted det(A), det A, or |A| is a single scalar value.
  - For 2 by 2 matrices:
    $$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad-bc$$
  - For 3 by 3 matrices
    $$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$$
  - And so on...
- The determinant is zero if and only if the matrix is singular.
  When A is invertible, $\det(A^{-1}) = \frac{1}{\det(A)}$
  - For 2 by 2 matrices:
    $$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ has inverse } A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
- The product of the pivots is the determinant.
  - For a 2 by 2 matrix:
    $$\det A = ad-bc = a(d - \tfrac{c}{a}b)$$
- The determinant of a matrix can be found in three ways
  1. Multiply the n pivots (times 1 or −1) (Pivot formula)
  2. Add up n! (times 1 or −1) ("big" formula)
  3. Combine n smaller determinants (times 1 or −1) (Cofactor formula)
  - The (times 1 or −1) comes from the following rule:
    The determinant changes signs when two rows (or two columns) are exchanged.
  - The identity matrix has determinant +1. Exchange the 2 rows and det P = −1.
    Half of all permutations are even (det P = 1) and half are odd (det P = −1)
    Starting from I, half of P's involve an even number of exchanges and half involve an odd number
    $$\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1 \quad \text{and} \quad \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1$$
- Another rule is linearity:
  $\det(cA) = c^n \det(A)$ for an n by n matrix

- Applications of determinants:
  - Determinants give A^{-1} and solutions to $A^{-1}b$ (this formula is called Cramer's Rule)
  - When the edges of a box are the rows of A, the volume is $|\det A|$
  - For n special numbers λ, called eigenvalues, the determinant of $A - \lambda I$ is zero.

The Properties of the Determinant

- Determinants have three basic properties

1. The determinant of the n by n identity matrix is one

$$\begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1 \quad \text{and} \quad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

2. The determinant changes signs when two rows are exchanged

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{(both sides equal } bc-ad\text{)}$$

   - We can find the determinant of any permutation matrix by counting the number of row exchanges, k. Then $\det P = (-1)^k$

3. The determinant is a linear function of each row separately (all other rows remain fixed)
   - If the first row is multiplied by t, the determinant is also multiplied by t. If the first rows of 2 matrices are added, the determinants are added.
   - This rule only applies when the other rows remain the same

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

   - Combining multiplication and addition, we can get linear combinations in one row
   - This does not mean $\det(2I) = 2\det(I)$. To get $2I$, we multiply both rows by 2 and the factor 2 comes out both times

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2^2 = 4 \quad \text{and} \quad \begin{vmatrix} t & 0 \\ 0 & t \end{vmatrix} = t^2$$

   - This is like area and volume. Expanding a rectangle by 2 increases the area by 4.

4. If two rows of A are equal, then $\det A = 0$

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0$$

   - This follows from rule 2. Exchanging the two equal rows changes the sign of the determinant, yet the matrix has not changed. $\det A = -\det A$ implies $\det A = 0$

5. Subtracting a multiple of a row from another row leaves $\det A$ unchanged.

$$\begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

   - This follows from rule 3

$$\begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ -la & -lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + 0 = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

- Elimination steps don't change the determinant. Since permuting the rows only inverts the sign of the determinant, $\det A = \pm \det U$

6. A matrix with a row of zeroes has $\det A = 0$

$$\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = 0$$

- Add some other row in the matrix to the zero row. The determinant has not changed and there are now 2 equal rows so $\det A = 0$

7. If A is triangular then $\det A = a_{11} a_{22} \cdots a_{nn}$ = product of diagonal entries

$$\begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad \quad \text{and} \quad \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad$$

- For a proof, eliminate until rref is reached (this does not change the determinant), and factor out each pivot so that $\det A = a_{11} a_{22} a_{33} \cdots a_{nn} \det I$

8. If A is singular then $\det A = 0$. If A is invertible then $\det A \neq 0$
   - If A is invertible then it must have n pivots. If pivots are zero (singular), then its determinant is also zero

9. The determinant of AB is $\det A$ times $\det B$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & q \\ r & s \end{vmatrix} = \begin{vmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{vmatrix}$$

- When $B = A^{-1}$, then from this rule we get $\det A^{-1} = \frac{1}{\det A}$
  $AA^{-1} = I$ so $\det(A)\det(A^{-1}) = \det I = 1$
- Proof: When $|B| \neq 0$, consider the ratio $D(A) = |AB|/|B|$.
  This ratio has properties 1, 2, and 3 and then $D(A)$ is the determinant
  $|A| = |AB|/|B| \Rightarrow |AB| = |A||B|$

10. The transpose A^T has the same determinant as A.
   - If A is singular so is A^T, so $\det A = \det A^T = 0$
   - If A is invertible, it has the factorization $PA = LU$. Transposing both sides, $A^T P^T = U^T L^T \Rightarrow |A^T| = \frac{|U^T||L^T|}{|P^T|}$, $\det L = \det L^T = 1$ (both have 1's on diagonal). $\det U = \det U^T$ (same diagonal), $\det P = \det P^T$ (permutations have $P^T P = I$). So L, U, P have the same determinants as L^T, U^T, P^T so $\det A = \det A^T$.

- Every rule above also applies to columns, since $\det A = \det A^T$

Problem Set 5.1

1.

$\det A = \frac{1}{2}$

$\det(2A) = 2^4 \det(A) = 16(\frac{1}{2}) = 4$

$\det(-A) = (-1)^4 \det(A) = \frac{1}{2}$

$\det(A^2) = \det(AA) = \det(A)\det(A) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$

$\det(A^{-1}) = \frac{1}{\det A} = 2$

2.

$\det(A) = -1$

$\det(\frac{1}{2}A) = \frac{1}{2}^3 \det A = \frac{1}{8}(-1) = -\frac{1}{8}$

$\det(-A) = (-1)^3 \det A = 1$

$\det(A^2) = (\det A)(\det A) = 1$

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

3.

a) False

b) True

c) False

d) True

4.

$J_3 = \underbrace{P_{13}}_{\text{one exchange}} I$

$\therefore \det(J_3) = -1$

$J_4 = \underbrace{P_{14} P_{23}}_{\text{two exchanges}} I$

$\therefore \det(J_4) = 1$

5.

$\det(J_n) = (-1)^n, n \in \mathbb{N}, n \geq 3$

8.

a)

$|Q^T Q| = |I|$

$|Q^T||Q| = 1$

$|Q|^2 = 1$

$|Q| = \pm 1$

9.

$\det A = 1$

$\det B = 2$

$\det C = 0$

13.

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

$\det A = 1$

19.

$\det U = 6$

$\det U^{-1} = \frac{1}{6}$

$\det U^2 = 36$

30.

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

Chapter 5.2: Permutations and Cofactors

- A computer finds the determinant from the pivots. This section explores the two other formulas: the "big" formula and the cofactor formula.
- Example

$$A = \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} \quad \text{has } \det A = 5$$

we can compute the determinant in three ways:

1. The product of the pivots $2 \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4} = 5$
2. The "big" formula has $4! = 24$ terms. Only 5 terms are non zero

   $\det A = 16 - 4 - 4 - 4 + 1 = 5$

3. The numbers $2, -1, 0, 0$ in the first row multiply their cofactors $4, 3, 2, 1$ from the other rows. That gives $2 \cdot 4 - 1 \cdot 3 = 5$. Those cofactors are 3 by 3 determinants. Cofactors use the rows and column not used by the entry in the first row.

The Pivot Formula

- Elimination leaves the pivots $d_1, \ldots, d_n$ on the diagonal of the upper triangular U. If no row exchanges are involved, multiply those pivots to find the determinant.

$$\det A = (\det L)(\det U) = (1)(d_1 d_2 \cdots d_n)$$

- For the factorization $PA = LU$, the determinant of P is either $+1$ or -1, so

$$(\det P)(\det A) = (\det L)(\det U) \Rightarrow \det A = \pm(d_1 d_2 \ldots d_n)$$

- When A has fewer than n pivots, $\det A = 0$ and A is singular.
- Example 1:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad PA = \begin{bmatrix} 4 & 5 & 6 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} \quad \det A = -(4)(2)(1) = -8$$

P has one exchange

• Example 2

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & \ddots & \ddots & \\ & & & -\frac{n-1}{n} & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \ddots & \ddots \\ & & & & \frac{n+1}{n} \end{bmatrix}$$

$$\det A = (\det L)(\det U) = (1)(2)(\tfrac{3}{2})(\tfrac{4}{3})\ldots(\tfrac{n+1}{n}) = n+1$$

◦ The first pivots depend only on the upper left corner of the original matrix A. This is the rule for all matrices without row exchanges.

• The first k pivots come from the k by k submatrix A_k in the top corner of the original matrix A. The determinant of that corner submatrix is $d_1 d_2 \ldots d_k$.

$$A_1 = [2] \Rightarrow \det A_1 = d_1 = 2 \qquad A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \det A_2 = d_1 d_2 = 3$$

$$A_3 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \Rightarrow \det A_3 = d_1 d_2 d_3 = 4$$

• Elimination deals with the corner matrix A_k while starting on the whole matrix. We assume no row exchanges — then $A = LU$ and $A_k = L_k U_k$. Dividing one determinant by the previous gives us the latest pivot d_k:

The kth pivot is $d_k = \dfrac{d_1 d_2 \ldots d_k}{d_1 d_2 \ldots d_{k-1}} = \dfrac{\det A_k}{\det A_{k-1}}$

The Big Formula for Determinants

• We want to derive a single explicit formula for the determinant from the entries a_{ij}.

• The formula has $n!$ terms.

• For a 3 by 3 matrix

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

• Each term has one entry from each row and one from each column. The order of the permuted columns tells us the sign. For example: the column order 1, 3, 2 is negative, and 3, 1, 2 is positive.

◦ Down the main diagonal, the column order 1, 2, 3, 4 is always positive. That order is the identity permutation.

• Start with $n=2$. The goal is to reach $ad-bc$ in a systematic way

$$[a\ b] = [a\ 0] + [0\ b] \quad \text{and} \quad [c\ d] = [c\ 0] + [0\ d]$$

Now apply linearity:

0 b/c dependent rows

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}$$

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = ad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + bc \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = ad - bc$$

- Now try $n=3$; we again split into permutation matrices

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

  - We get 27 determinants, only 6 of which are non zero (3 positions per row: $3 \times 3 \times 3$)
  - There are $3! = 6$ ways to order the columns, so 6 determinants
  - The determinants are non zero only when the non zero terms come from different columns.
  - The 6 permutations of $(1,2,3)$ are:
    $(1,2,3), (2,3,1), (3,1,2), (1,3,2), (2,1,3), (3,2,1)$
  - The last 3 are odd permutations (1 exchange) and the other 3 are even, (0 or 2 exchanges). When the column sequence is (α, β, ω), we have chosen entries $a_{1\alpha} a_{2\beta} a_{3\omega}$ and the column sequence comes with a plus or minus sign.

$$\det A = a_{11}a_{22}a_{33} \begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31} \begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32} \begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix} +$$
$$a_{11}a_{23}a_{32} \begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33} \begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31} \begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}$$

- The above goes for any n by n matrix.
  - There are $n!$ permutations of the columns $(1, 2, \ldots, n)$. Taking $(\alpha, \beta, \ldots, \omega)$ takes $a_{1\omega}, a_{2\beta}, \ldots a_{n\omega}$. The determinant contains the product $a_{1\alpha}a_{2\beta}\ldots a_{n\omega}$ times $+1$ or -1, which comes from the permutation matrix

$$\det A = \text{sum over all } n! \text{ column permutations } P = (\alpha, \beta, \ldots, \omega)$$
$$= \sum (\det P) a_{1\alpha} a_{2\beta} \ldots a_{n\omega}$$

Determinant by Cofactors

- For a 3 by 3 matrix, linearity becomes clear if you factor a_{11}, a_{12}, or a_{13} that comes from the first row.

$$\det A = a_{11}\underbrace{(a_{22}a_{33} - a_{23}a_{32})} + a_{12}\underbrace{(a_{23}a_{31} - a_{21}a_{33})}_{\text{cofactors}} + a_{13}\underbrace{(a_{21}a_{32} - a_{22}a_{31})}$$

- These cofactors are 2 by 2 determinants coming from submatrices in row 2 and 3
  - The first row contributes the factors a_{11}, a_{12}, a_{13}. The lower rows contribute the cofactors C_{11}, C_{12}, C_{13}.
  - The cofactor of a_{11} is $C_{11} = a_{22}a_{33} - a_{23}a_{32}$

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

  - We're still choosing one entry per row and column. Since each entry in the first row takes up a row and column, that leaves a 2 by 2 submatrix as the cofactor

- We need to watch signs. The 2 by 2 determinant that goes with a_{12} looks like $a_{21}a_{33} - a_{23}a_{31}$, but in the cofactor C_{12}, its sign is reversed. Then $a_{12}C_{12}$ is the correct 3 by 3 determinant.
- The sign pattern for cofactors along the first row is plus-minus-plus-minus-etc. You cross out row 1 and column j to get a submatrix M_{1j} of size $(n-1)$ by $(n-1)$. Multiply the determinant of M by $(-1)^{1+j}$ to get the cofactor.
- The cofactors along row 1 are $C_{1j} = (-1)^{1+j} \det M_{1j}$. The cofactor expansion is $\det A = a_{11}C_{11} + a_{12}C_{12} + \dots + a_{1n}C_{1n}$
- This is possible for any row, not just row 1. The entries a_{ij} in that row have cofactors C_{ij}. These are determinants of order $n-1$ multiplied by $(-1)^{i+j}$. Since a_{ij} accounts for row i and column j, the submatrix M_{ij} throws out row i and column j.

$$A = \begin{vmatrix} \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ & & a_{43} & \end{vmatrix} \qquad \text{Signs } (-1)^{i+j} = \begin{vmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{vmatrix}$$

- The determinant is the dot product of any row i of A with its cofactors using other rows.

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in}$$

Each cofactor C_{ij} (order $n-1$, without i and column j)

$$\boxed{C_{ij} = (-1)^{i+j} \det M_{ij}}$$

- A determinant of order n is a combination of determinants of order $n-1$. We can recursively break it down until we reach order 1 (determ. [a] of 1 by 1 matrix $= a$)
- Example 6

$$\begin{vmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{vmatrix} = 2\begin{vmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{vmatrix} - (-1)\begin{vmatrix} -1 & -1 & \\ & 2 & -1 \\ & -1 & 2 \end{vmatrix} = 2\begin{vmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{vmatrix} + \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix}$$

Problem Set 5.2

1.

$$\det A = (1)(1)(1) + (2)(2)(3) + (3)(3)(2) - (3)(1)(3) - (3)(2)(1) - (1)(2)(2)$$
$$= 1 + 12 + 18 - 9 - 6 - 4$$
$$= 12$$

3.

$$\det A = x\begin{vmatrix} 0 & x \\ 0 & x \end{vmatrix} + x\begin{vmatrix} 0 & x \\ 0 & x \end{vmatrix} + x\begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}$$
$$= 0$$

5.

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

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

11.

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 0 & 0 \end{bmatrix}$

$C = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$ $\quad D = \begin{bmatrix} 0 & 42 & -35 \\ 0 & -21 & 14 \\ -3 & 6 & -3 \end{bmatrix}$

$$AC = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix} = \begin{bmatrix} ad - b^2 & -ac + ab \\ cd - bd & -c^2 + ad \end{bmatrix}$$

15

a)

$$E_n = 1 \underbrace{\begin{vmatrix} 1 & 1 & & \\ 1 & 1 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & 1 \end{vmatrix}}_{n-1} - 1 \underbrace{\begin{vmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & 1 \end{vmatrix}}_{n-1}$$

$$= \underbrace{\begin{vmatrix} 1 & 1 & & \\ 1 & 1 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & 1 \end{vmatrix}}_{n-1} - \underbrace{\begin{vmatrix} 1 & 1 & & \\ 1 & 1 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & 1 \end{vmatrix}}_{n-2} + \underbrace{\begin{vmatrix} & 1 & & \\ & 1 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & \end{vmatrix}}_{0}$$

$$= E_{n-1} - E_{n-2}$$

28.

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

$$= (1)(5)(9) + (2)(6)(7) + (3)(4)(8) - (3)(5)(7) - (2)(4)(9) - (1)(6)(8)$$
$$= 0$$

Chapter 5.3: Cramer's Rule, Inverses, and Volumes

- Cramer's Rule solves $Ax = b$. A neat idea gives the first column x_1.
  Replacing the first column of I by x gives a matrix with determinant x_1.
  When you multiply by A, the first column becomes Ax which is b. The other columns are copied from A.

Key Idea
$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}\begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1$$

Take determinants of the three matrices
$$(\det A)(x_1) = \det B_1 \quad \text{or} \quad x_1 = \frac{\det B_1}{\det A}$$

- We follow the same principle to find all components of x
$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \dots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$
  Where A_i denotes the matrix created by replacing the ith column of A with b
- Example
$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \Rightarrow x_1 = \frac{\det\begin{bmatrix} 2 & 4 \\ 4 & 6 \end{bmatrix}}{\det\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}} = \frac{-4}{-2} = 2, \quad x_2 = \frac{\det\begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}}{\det\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}} = \frac{2}{-2} = -1$$
- Cramer's Rule requires the computation of $n+1$ determinants, each of which is $n!$ terms each $(n+1)!$ terms!! Inefficient but still an explicit formula

- Example 2
  - Finding the inverse matrix
    $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ can be split into $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

    We need 5 determinants:

    $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ and $\begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix}, \begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix}, \begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix}, \begin{vmatrix} a & 0 \\ c & 1 \end{vmatrix}$

    The last four are $d, -c, -b, a$ (They are the cofactors)

    $x_1 = \frac{d}{|A|}, x_2 = -\frac{c}{|A|}, y_1 = -\frac{b}{|A|}, y_2 = \frac{a}{|A|}$ then $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

- When the right side is a column of the identity matrix, the determinant of each A_i in Cramer's Rule is a cofactor.
  - For the first column of a 3 by 3 matrix

    $\begin{vmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} = C_{11}$ $\begin{vmatrix} a_{11} & 1 & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} = C_{12}$ $\begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix} = C_{13}$

    *These 3 determinants solve the first column of A (i.e. vertically), while the identity matrix column moves horizontally)

  - These cofactors go in their transpose (j, i) position.* So we transpose the cofactor matrix (this is called the adjugate matrix of A, adj(A)), and we divide by the determinant

- Formula for A^{-1}:

  $(A^{-1})_{ij} = \frac{C_{ji}}{\det A}$ and $A^{-1} = \frac{C^T}{\det A} = \frac{adj(A)}{\det(A)}$, where adj denotes the adjugate.

- In solving $AA^{-1} = I$, the columns of I lead to the columns of A^{-1}. Then Cramer's Rule using b = columns of I gives the formula for A^{-1}.
- Direct Proof: Multiply A by C^T

  $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix}$

  $AC^T = (\det A) I$

  $A^{-1} = \frac{C^T}{\det A}$

  - The reason we get zeros for expressions like:
    $a_{21}C_{11} + a_{22}C_{12} + a_{23}C_{13} = 0$
  - This is the cofactor rule for a new matrix A^* with the second row of A is copied into its first row, and so $\det A^* = 0$ b/c there are two identical rows.

- Example:

  $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ has inverse $A^{-1} = \frac{C^T}{1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$

  "sum matrix"

Area of a triangle

- Determinants allow us to compute the area of a triangle.
- The area of a triangle with corners (x_1, y_1), (x_2, y_2), and (x_3, y_3) have the area

$$\text{Area} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \qquad \text{Area} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \text{ when } (x_3, y_3) = (0,0)$$

- The 3 by 3 determinant can be broken down by cofactors

$$\text{Area} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \frac{1}{2}x_1\begin{vmatrix} y_2 & 1 \\ y_3 & 1 \end{vmatrix} - \frac{1}{2}y_1\begin{vmatrix} x_2 & 1 \\ x_3 & 1 \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}$$

$$= \frac{1}{2}x_1y_2 - \frac{1}{2}x_1y_3 - \frac{1}{2}x_2y_1 + \frac{1}{2}x_3y_1 + \frac{1}{2}x_2y_3 - \frac{1}{2}x_3y_2$$

$$= \frac{1}{2}(x_1y_2 - x_2y_1) + \frac{1}{2}(x_2y_3 - x_3y_2) + \frac{1}{2}(x_3y_1 - x_1y_3)$$

- The 3 by 3 determinant is split into two 2 by 2 determinants just as triangles can be split into 3 triangles with a point at $(0,0)$ (see fig 5.3.1)

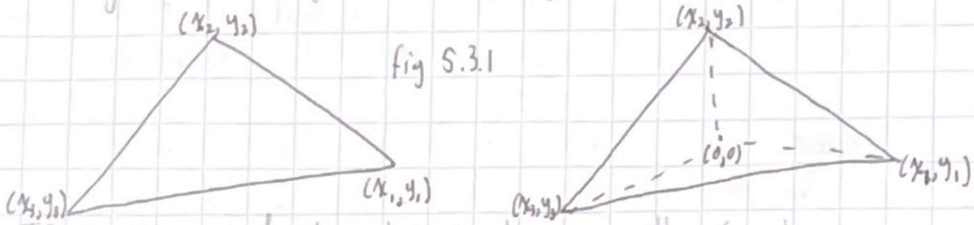


fig 5.3.1

- If $(0,0)$ is outside the triangle, two of the special areas can be negative — but the sum is still correct. The real problem is to explain the special area $\frac{1}{2}(x_1y_2 - x_2y_1)$
- If $\frac{1}{2}(x_1y_2 - x_2y_1)$ is the area of a triangle, then $x_1y_2 - x_2y_1$ is the area of a parallelogram
- Proof that a parallelogram starting from $(0,0)$ has area = 2 by 2 determinant
  - The area has the same properties as the determinant
    1. When $A = I$, the parallelogram is the unit square, $\det I = 1$
    2. When the rows are exchanged, the determinant reverses sign. The absolute value (the positive area) remains the same.
    3. If row 1 is multiplied by t, the area is also multiplied by t. Suppose a new row (x_1', y_1') is added to (x_1, y_1). The area is then added

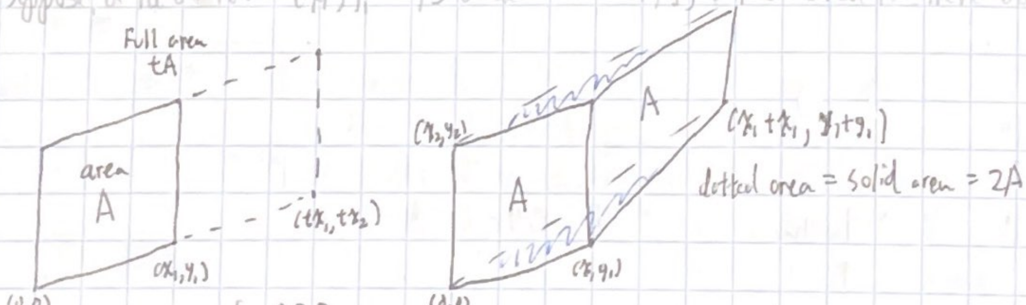
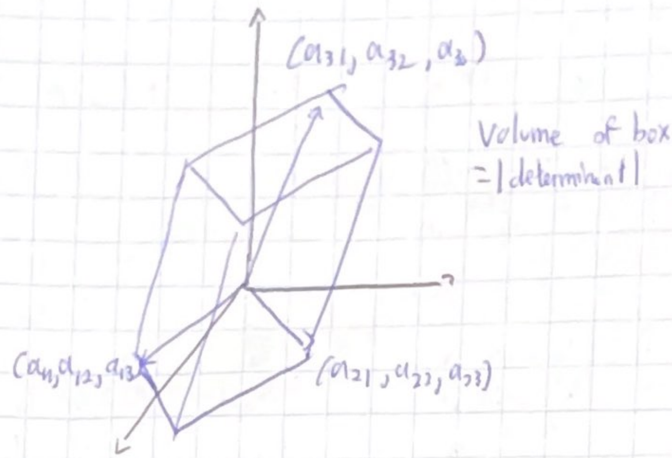


fig 5.3.2

- The n edges going out from the origin are given by the rows of an n by n square matrix.

- The volume of that parallelepiped is the absolute value of $\det A$.
- The 3 rules for determinants are also obeyed by volumes



- Example
  - Suppose a rectangular box has side lengths $r, s,$ and t. The diagonal matrix with entries r, s, t produces those 3 sides. Then $\det A = rst$
- Example
  - In calculus, to integrate over a circle, we might use polar coordinates. $x = r\cos\theta$, $y = r\sin\theta$. The area of a "polar box" is a determinant J times $dr\,d\theta$

$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

The determinant is the r in the differential $dA = r\,dr\,d\theta$.

The Cross Product

- In three dimensions, a cross product between 2 vectors is defined as

$$a \times b = \begin{vmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = (a_2b_3 - a_3b_2)i + (a_3b_1 - a_1b_3)j + (a_1b_2 - b_1a_2)k$$

This vector is perpendicular to both u and v. The cross product of $b \times a = -(a \times b)$
- The above formula is a very useful mnemonic, but is not particularly legal, since a and b are scalars and $i, j,$ and k are vectors
- Properties
  1. $v \times u$ reverses rows 2 and 3 in the determinant and so it equals $-(u \times v)$
  2. The cross product $u \times v$ is perpendicular to u and v.
  3. The cross product of any vector with itself is 0 (2 equal rows)

     $\|u \times v\| = \|u\|\|v\|\sin\theta$ $\quad\quad$ $\|u \cdot v\| = \|u\|\|v\||\cos\theta|$
- The magnitude $u \times v$ is the area of the parallelogram with sides u and v
- Example

$u = (1,1,1)$, $v = (1,1,2)$

$$\begin{vmatrix} i & j & k \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = i\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} - j\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} + k\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = i - j = (1,-1,0)$$

Triple Product = Determinant = Volume

- A triple product is a specific product of 3 3-dimensional vectors
  - Scalar triple product : $(u \times v) \cdot w$
  - vector triple product : $(u \times v) \times w$
- The scalar triple product is a determinant and gives the volume of the parallelepiped formed by $u, v,$ and w.

$$(u \times v) \cdot w = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \quad \text{(two row exchanges)}$$

Problem Set 5.3

1. a)

$$x_1 = \frac{\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix}}{\begin{vmatrix} 2 & 5 \\ 1 & 4 \end{vmatrix}} = \frac{-6}{3} = -2 \qquad x_2 = \frac{\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 5 \\ 1 & 4 \end{vmatrix}} = \frac{3}{3} = 1$$

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

2. a)

$$y = \frac{\begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{-c}{ad-bc}$$

b)

$$x_1 = \frac{\begin{vmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix}} = \frac{3}{4} \qquad x_2 = \frac{\begin{vmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix}} = \frac{-2}{4} = -\frac{1}{2} \qquad x_3 = \frac{\begin{vmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix}} = \frac{1}{4}$$

6.

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 7 & 1 \end{bmatrix} \qquad A^{-1} = \frac{C^T}{\det A}$$

$$C = \begin{bmatrix} 3 & 0 & 0 \\ -2 & 1 & -7 \\ 0 & 0 & 3 \end{bmatrix} \Rightarrow C^T = \begin{bmatrix} 3 & -2 & 6 \\ 0 & 1 & 0 \\ 0 & -7 & 3 \end{bmatrix}$$

$$\det A = 1(3) + 2(0) + 0(0) = 3$$

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

7.

$\det A = 0$ by cofactor expansion, so

9.

$$AC^T = (\det A)I$$
$$= I$$
$$C^T = A^{-1}$$

Find inverse & $C^T = A^{-1}$

$$[\,C^T \;\; I\,] \to [\,I \;\; C^{T^{-1}}\,]$$
$$C^{T^{-1}} = (A^{-1})^{-1} = A$$

10.

$$AA^{-1} = I$$
$$\underbrace{(\det A)(\det A^{-1})}_{\text{must be both 1 or -1}} = 1$$

15.

$24 \times 4 \times 25 = 2400$

16.

$$\begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} = 10$$

b)

17.

$$\begin{vmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{vmatrix} = 27 + 1 + 1 - 3 - 3 - 3 = 20$$

36.

$$(x, y, z) \cdot ((1,1,0) \times (1,2,1)) = 0$$

$$\begin{vmatrix} x & 1 & 1 \\ y & 1 & 2 \\ z & 0 & 1 \end{vmatrix} = x + 2z - z - y = 0$$

$$x - y + z = 0$$

Chapter 6: Eigenvalues and Eigenvectors

Chapter 6.1: Introduction to Eigenvalues

- Linear equations $Ax=b$ come from steady state problems. Eigenvalues have their greatest importance in dynamic problems. The solution of $\frac{du}{dt}=Au$ is changing with time — growing, decaying, or oscillating. We can't find it by elimination.
- This chapter enters a new part of linear algebra, based on $Ax=\lambda x$. All matrices in this chapter are square.
- A good model comes from the powers $A, A^2, A^3, \ldots$ of a matrix. Suppose you need the hundredth power A^{100}

$$\underset{A}{\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}} \quad \underset{A^2}{\begin{bmatrix} .70 & .45 \\ .30 & .55 \end{bmatrix}} \quad \underset{A^3}{\begin{bmatrix} .650 & .525 \\ .350 & .475 \end{bmatrix}} \quad \cdots \quad \underset{A^{100}}{\begin{bmatrix} .60000 & .60000 \\ .40000 & .4000 \end{bmatrix}}$$

  - A^{100} approaches $[.6 \; .6; \; .4 \; .4]$
  - A^{100} is found by using the eigenvalues of A, not by multiplying 100 matrices
- To explain eigenvalues, we must first explain eigenvalues. Almost all vectors change direction when multiplied by A. Certain exceptional vectors x remain in the same direction; that is $Ax=\lambda x$, where λ is a scalar.
  - The vector x is an "eigenvector"
  - The scalar λ (lambda) is an "eigenvalue". It tells how much a specific eigenvector is scaled in Ax. It can be any real number
- If $A=I$, then all vectors are eigenvectors of I, and all eigenvalues are 1
- We can find eigenvalues by using $\det(A-\lambda I)=0$*
- Example:

$A=\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$ has eigenvalues $\lambda_1=1,\ \lambda_2=\frac{1}{2}$

$\det\left(\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}-\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)=\det\left(\begin{bmatrix} -.2 & .3 \\ .2 & -.3 \end{bmatrix}\right)=0$

$\det\left(\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}-\frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)=\det\left(\begin{bmatrix} .3 & .3 \\ .2 & .2 \end{bmatrix}\right)=0$

$x_1=\begin{bmatrix} .6 \\ .4 \end{bmatrix}$ and $Ax_1=\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} .6 \\ .4 \end{bmatrix}=x_1 \quad (\lambda_1=1)$

$x_2=\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $Ax_2=\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}=\begin{bmatrix} .5 \\ -.5 \end{bmatrix}=\frac{1}{2}x_2 \quad (\lambda_2=\frac{1}{2})$

Solving for λ:

$\det\left(\begin{bmatrix} .8-\lambda & .3 \\ .2 & .7-\lambda \end{bmatrix}\right)=\overbrace{\lambda^2-\frac{3}{2}\lambda+\frac{1}{2}}^{\text{"characteristic}}=\overbrace{(\lambda-1)(\lambda-\frac{1}{2})}^{\text{polynomial"}}$

The eigenvectors x_1 and x_2 are in the nullspace of $A-I$ and $A-\frac{1}{2}I$

$(A-I)x_1=0 \Rightarrow Ax_1=x_1 \Rightarrow x_1=(.6, .4)$

$(A-\frac{1}{2}I)x_2=0 \Rightarrow Ax_2=\frac{1}{2}x_2 \Rightarrow x_2=(1,-1)$

*

$Ax=\lambda x \Rightarrow Ax-\lambda x=0 \Rightarrow (A-\lambda I)x=0$

$(A-\lambda I)x=0$ has a nonzero sol. iff $\det(A-\lambda I)=0$

  - Apply A over and over on the eigenvectors do not change its direction. The eigenvalues are exponentiated. $AAx=A\lambda x=\lambda^2 x$, etc
- All other vectors are combinations of the eigenvectors's

- The first column of the example matrix A can be split into $x_1 + (.2)x_2$

$$\begin{bmatrix} .8 \\ .2 \end{bmatrix} = x_1 + (.2)x_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} .2 \\ -.2 \end{bmatrix}$$

Multiplying by A gives $(.7, .3)$, the first column of A^2. Do it separately for x_1 and $(.2)x_2$. A multiplies each eigenvector by its eigenvalue

$$A\begin{bmatrix} .8 \\ .2 \end{bmatrix} = A\left(\begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} .2 \\ -.2 \end{bmatrix}\right) = A\begin{bmatrix} .6 \\ .4 \end{bmatrix} + (0.2)A\begin{bmatrix} 1 \\ -1 \end{bmatrix} = (1)\begin{bmatrix} .6 \\ .4 \end{bmatrix} + \tfrac{1}{2}(0.2)\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} .7 \\ .3 \end{bmatrix}$$

Then,

$$A^{99}\begin{bmatrix} .8 \\ .2 \end{bmatrix} \text{ is really } \overbrace{x_1}^{\text{multiply by } \lambda_1^{99} = 1^{99} = 1} + \underbrace{(.2)\left(\tfrac{1}{2}\right)^{99}x_2}_{\text{multiply by } \lambda_2^{99} = (\frac{1}{2})^{99}} = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} \text{very small} \\ \text{vector} \end{bmatrix}$$

  - The eigenvector x_1 is a "steady state" that doesn't change (b/c $\lambda_1 = 1$), while the eigenvector x_2 is a "decaying mode" that virtually disappears (b/c $\lambda_2 = .5$)
- This particular matrix A is a Markov matrix. Its entries are positive and each column sums to 1. These facts guarantee the largest eigenvalue as 1. Its eigenvector $x_1 = (.6, .4)$ is the steady state that all columns of A^k will approach.
- For projections we can spot the steady state ($\lambda = 1$) and the nullspace ($\lambda = 0$)
- Example:

The projection matrix $P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$ has eigenvalues $\lambda = 1$ and $\lambda = 0$

Its eigenvectors are $x_1 = (1, 1)$ and $x_2 = (1, -1)$. For those vectors, $Px_1 = x_1$ (steady state) and $Px_2 = 0$ (nullspace)

1. Each column of P sums to 1, so $\lambda = 1$ is an eigenvalue
2. P is singular so $\lambda = 0$ is an eigenvalue
3. P is symmetric, so its eigenvectors $(1,1)$ and $(1,-1)$ are perpendicular

The eigenvectors for $\lambda = 0$ (which means $Px = 0x$) fill up the nullspace. The eigenvectors for $\lambda = 1$ (which means $Px = x$) fill up the column space
The nullspace is projected to 0. The column space projects to itself. The projection keeps the column space and destroys the nullspace

$$v = \overbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}^{x_2} + \overbrace{\begin{bmatrix} 2 \\ 2 \end{bmatrix}}^{2x_1} \Rightarrow Pv = \overbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}^{\text{multiply by } \lambda_1 = 0} + \overbrace{\begin{bmatrix} 2 \\ 2 \end{bmatrix}}^{\text{multiply by } \lambda_2 = 1}$$

- Permutation matrices have all $|\lambda| = 1$
- Example:

$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$ and eigenvectors $x_1 = (1,1)$ and $x_2 = (1,-1)$

- The eigenvalues for R are the same as P because reflection $= 2$(projection) $- I$

$$R = 2P - I \Rightarrow \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = 2\begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- If $Px = \lambda x$, then $2Px = 2\lambda x$. The eigenvalues are doubled when the matrix is doubled. Subtracting $Ix = x$, we get $(2P - I)x = 2Px - Ix = 2\lambda x - x = (2\lambda - 1)x$
  When a matrix is shifted by I, each λ is shifted by 1.

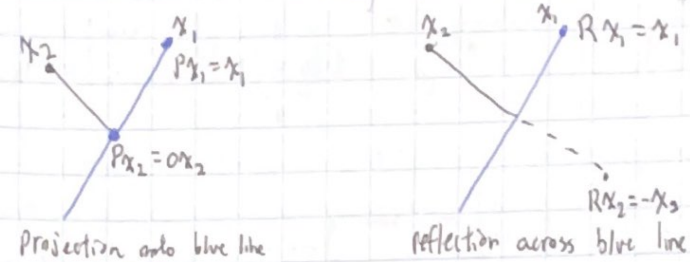


fig 6.1.1
Projections P have eigenvalues 0 and 1
Reflections R have eigenvalues -1 and 1

- Key idea: The eigenvalues of R and P are related exactly as the matrices are related.
  - The eigenvalues of $R = 2P - I$ are $2(1) - 1 = 1$ and $2(0) - 1 = -1$
  - The eigenvalues of R^2 are λ^2. In this case $R^2 = I$. Check $(1)^2 = 1$ and $(-1)^2 = 1$

The Equation for the Eigenvalues

- Start with $Ax = \lambda x$ and move λx to the left side

$$Ax - \lambda x = 0 \Rightarrow (A - \lambda I)x = 0$$

- The eigenvectors x are in the nullspace of $A - \lambda I$. When we know an eigenvalue λ, we find an eigenvector by solving $(A - \lambda I)x = 0$.
- We want a nonzero solution to $(A - \lambda I)x = 0$ (0 is not considered an eigenvector), which occurs if and only if $\det(A - \lambda I) = 0$ (i.e. the matrix is singular)
- Using the big formula for determinants, we get a "characteristic polynomial" in λ, whose roots give us the eigenvalues. Then,

  For each eigenvalue λ, solve $(A - \lambda I)x = 0$ or $Ax = \lambda x$ to find an eigenvector x

- Example:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow A - \lambda I = \begin{bmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix} \Rightarrow \det(A - \lambda I) = (1-\lambda)(4-\lambda) - (2)(2) = \lambda^2 - 5\lambda$$

$$\lambda^2 - 5\lambda = 0 \Rightarrow \lambda(\lambda - 5) = 0 \Rightarrow \lambda = 0, 5$$

- Note: If A is already singular, then $\lambda = 0$ is an eigenvalue

The matrices $A - 0I$ and $A - 5I$ are singular

$$(A - 0I)x = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields an eigenvector } \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \text{ for } \lambda_1 = 0$$

$$(A - 5I)x = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields an eigenvector } \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ for } \lambda_2 = 5$$

The eigenvectors $(2, -1)$ and $(1, 2)$ are in the nullspace of $(A - \lambda I)x = 0$

· Summary:
1. Compute $\det(A-\lambda I)$, which is a polynomial in λ of degree n
2. Find the roots of this "characteristic polynomial" $\det(A-\lambda I)=0$.
The n roots are the n eigenvalues of A. They make $A-\lambda I$ singular.
3. For each eigenvalue λ, solve $(A-\lambda I)x=0$ to find an eigenvector x

- A note on the eigenvectors of 2 by 2 matrices. When $A-\lambda I$ is singular, both rows are multiples of a vector (a,b). The eigenvector is any multiple of $(b,-a)$.
  - $\lambda=0$: rows of $A-0I$ in the direction $(1,2)$; eigenvector in the direction of $(2,-1)$
  - $\lambda=5$: rows of $A-5I$ in the direction $(-4,2)$; eigenvector in the direction of $(2,4)$
- Note that if x is an eigenvector, so is cx, where c is a constant. There is a whole line of eigenvectors in the direction of x.
- Some 2 by 2 matrices only have one line of eigenvectors, which happens only if two eigenvalues are equal.
- n by n matrices don't always have n independent eigenvectors, which doesn't give us a basis. We can't write every v as a combination of eigenvectors.

Good News, Bad News
- Elimination does not preserve eigenvalues.
$U=\begin{bmatrix}1 & 3\\ 0 & 0\end{bmatrix}$ has $\lambda=0,1$ $A=\begin{bmatrix}1 & 3\\ 2 & 6\end{bmatrix}$ has $\lambda=0,7$
  - Row exchanges and adding rows usually changes eigenvalues
- The products and sums of eigenvalues:
  - The product of the n eigenvalues equals the determinant of the matrix
  - The sum of the n eigenvalues equal the sum of the n diagonal entries.
    - The sum of the entries on the main diagonal is called the trace of A.
- The above are useful as checks but not for computation
- The determinant test makes the product of the λ's equal to the product of the pivots (assuming no row exchanges). But the sum of the λ's is not the sum of the pivots.
The individual λ's have almost nothing to do with the pivots.
In this new part of linear algebra, the key equation is really nonlinear: λ multiplies x.

Imaginary Eigenvalues
Eigenvalues are generally complex numbers.
- Example
$Q=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$ has no real eigenvectors. $(-\lambda)(-\lambda)-(-1)(1)=0 \Rightarrow \lambda^2+1=0$
- Its eigenvalues are $\lambda=i$ and $\lambda=-i$, which matches with the sum and product checks for eigenvalues.
- Q is a $90°$ rotation matrix. No vector Qx keeps its direction after a rotation besides 0. We must go into the complex numbers.
- If we look at Q^2, which is $-I$. Its eigenvalues are -1 and -1. These values are the squares of the eigenvalues of Q, and i^2 and $(-i)^2=-1$ and -1.

- Somehow, these complex vectors $x_1 = (1, i)$ and $x_2 = (i, 1)$ keep their directions upon rotation.
- Q is an orthonormal matrix, so the absolute value of each λ is $|\lambda| = 1$
- Q is a skew-symmetric matrix so each λ is pure imaginary.
- The symmetric matrix ($A^T = A$) can be compared to a real number and a skew-symmetric matrix ($A^T = -A$) can be compared to an imaginary number. An orthogonal matrix ($A^T A = I$) can be compared to a complex number with $|\lambda| = 1$

Problem set 6.1

1

a)

$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$

$A - \lambda I = \begin{bmatrix} .8-\lambda & .3 \\ .2 & .7-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (\frac{4}{5} - \lambda)(\frac{7}{10} - \lambda) - (\frac{1}{5})(\frac{3}{10})$

$= \frac{14}{25} - \frac{3}{2}\lambda + \lambda^2 - \frac{3}{50}$

$= \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2}$

Set $\det(A - \lambda I) = 0$

$\lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = 0$

$2\lambda^2 - 3\lambda + 1 = 0$

$(\lambda - 1)(2\lambda - 1) = 0$

$\lambda - 1 = 0$ or $2\lambda - 1 = 0$ (zpp)

$\lambda = 1$ or $\lambda = \frac{1}{2}$

$A' = \begin{bmatrix} .2 & .7 \\ .8 & .3 \end{bmatrix}$

$A' - \lambda I = \begin{bmatrix} .2-\lambda & .7 \\ .8 & .3-\lambda \end{bmatrix}$

$\det(A' - \lambda I) = (\frac{1}{5} - \lambda)(\frac{3}{10} - \lambda) - (\frac{4}{5})(\frac{7}{10})$

$= \frac{3}{50} - \frac{1}{2}\lambda + \lambda^2 - \frac{14}{25}$

$= \lambda^2 - \frac{1}{2}\lambda - \frac{1}{2}$

Set $\det(A - \lambda I) = 0$

$\lambda^2 - \frac{1}{2}\lambda - \frac{1}{2} = 0$

$2\lambda^2 - \lambda - 1 = 0$

$(\lambda - 1)(2\lambda + 1) = 0$

$\lambda - 1 = 0$ or $2\lambda + 1 = 0$ (zpp)

$\lambda = 1$ or $\lambda = -\frac{1}{2}$

b)

The $\lambda = 0$ eigenvalue leads to the equation

$(A - 0I)x = 0$

$Ax = 0$

elimination does not change the nullspace of A

2.

$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ $A + I = \begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix} = A'$

$A - \lambda I = \begin{bmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (1-\lambda)(3-\lambda) - (4)(2)$

$= 3 - 4\lambda + \lambda^2 - 8$

$= \lambda^2 - 4\lambda - 5$

Set $\det(A - \lambda I) = 0$

$\lambda^2 - 4\lambda - 5 = 0$

$(\lambda - 5)(\lambda + 1) = 0$

$\lambda - 5 = 0$ or $\lambda + 1 = 0$ (zpp)

$\lambda = 5$ or $\lambda = -1$

Case $\lambda = 5$

$(A - 5I)x_1 = 0$

$\begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$x_1 = (1, 1)$

Case $\lambda = 1$:

$(A + I)x_2 = 0$

$\begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$x_2 = (2, -1)$

Same, increased

$A' - \lambda I = \begin{bmatrix} 2-\lambda & 4 \\ 2 & 4-\lambda \end{bmatrix}$

$\det(A' - \lambda I) = (2-\lambda)(4-\lambda) - (4)(2)$

$= 8 - 6\lambda + \lambda^2 - 8$

$= \lambda^2 - 6\lambda$

Set $\det(A' - \lambda I) = 0$

$\lambda^2 - 6\lambda = 0$

$\lambda(\lambda - 6) = 0$

$\lambda = 0, 6$

$(A' - 0I)x_1 = 0$

$x_1 = (2, -1)$

$(A' - 6I)x_2 = 0$

$\begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix} x_2 = 0$

$x_2 = (1, 1)$

3.

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

$A - \lambda I = \begin{bmatrix} -\lambda & 2 \\ 1 & 1-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (-\lambda)(1-\lambda) - (2)(1)$
$= \lambda^2 - \lambda - 2$
$= (\lambda - 2)(\lambda + 1)$

set $\det(A - \lambda I) = 0$
$(\lambda - 2)(\lambda + 1) = 0$
$\lambda - 2 = 0$ or $\lambda + 1 = 0$ (zpp)
$\lambda = 2$ or $\lambda = -1$

Case $\lambda = 2$
$(A - 2I)x_1 = 0$
$\begin{bmatrix} -2 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$x_1 = (1, 1)$
Case $\lambda = -1$
$(A + I)x_2 = 0$
$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = 0$
$x_2 = (2, -1)$

$A^{-1} = \begin{bmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$
$A^{-1} - \lambda I = \begin{bmatrix} -\frac{1}{2} - \lambda & 1 \\ \frac{1}{2} & -\lambda \end{bmatrix}$
$\det(A^{-1} - \lambda I) = (-\frac{1}{2} - \lambda)(-\lambda) - (\frac{1}{2})(1)$
$= \lambda^2 + \frac{1}{2}\lambda - \frac{1}{2}$
set $\det(A^{-1} - \lambda I) = 0$
$0 = 2\lambda^2 + \lambda - 1$
$0 = (\lambda + 1)(2\lambda - 1)$
$\lambda = -1, \frac{1}{2}$
$\frac{1}{\lambda_1}, \frac{1}{\lambda}$

7.
determinant, 1, 0

26.
$Ax = \begin{bmatrix} Bx & Cx \\ 0 & Dx \end{bmatrix} = $

27.
A.
rank = 1

$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

$A - \lambda I = \begin{bmatrix} 1-\lambda & 1 & 1 & 1 \\ 1 & 1-\lambda & 1 & 1 \\ 1 & 1 & 1-\lambda & 1 \\ 1 & 1 & 1 & 1-\lambda \end{bmatrix}$

9.
a) $AAx = \lambda Ax = \lambda\lambda x$
b) $Ax = \lambda x$
$x = \lambda A^{-1}x$
$A^{-1}x = \frac{1}{\lambda}x$
c) $(A + I)x = Ax + Ix$
$= \lambda x + x$
$= (\lambda + 1)x$

16.
$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots(\lambda_n - \lambda)$
$\det(A - 0I) = \lambda_1\lambda_2\lambda_3\dots\lambda_n$

17.
$\det(A - \lambda I) = \lambda^2 - (a+d)\lambda + ad - bc = 0$
$\lambda = \frac{(a+d) \pm \sqrt{(a+d)^2 - 4(1)(ad - bc)}}{2(1)}$
$\lambda_{sum} = (a+d) = $ trace

Chapter 6.2: Diagonalizing a Matrix

- When x is an eigenvector, multiplication by A is just multiplication by a number λ: $Ax = \lambda x$. We can follow the eigenvectors separately instead of having an interconnected system.
- It's like having a diagonal matrix, with no off-diagonal interconnections. The 100th power of a diagonal matrix is easy.
- The matrix A turns into a diagonal matrix Λ (uppercase lambda) when we use the eigenvectors properly.
- Diagonalization: Suppose the n by n matrix A has n linearly independent eigenvectors $x_1, \dots, x_n$. Put them into the columns of an eigenvector matrix S. Then $S^{-1}AS$ is the eigenvalue matrix Λ: Matrix A is "diagonalized"

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- Proof: Multiply A by its eigenvectors, which are the columns of S. The first column of AS is Ax_1, which is $\lambda_1 x_1$. Each column of S is multiplied by its eigenvalue λ_i:

$$AS = A\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} Ax_1 & \cdots & Ax_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix}$$

The trick is to split this matrix AS into S times Λ

$$S\Lambda = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Then:

$AS = S\Lambda$ or $\Lambda = S^{-1}AS$ or $\underbrace{A = S\Lambda S^{-1}}_{\text{this is called the eigendecomposition of A}}$

S^{-1} exists because we assume S contains n independent eigenvectors as columns.
Without n independent eigenvectors, we can't diagonalize.

- A and Λ have the same eigenvalues $\lambda_1, \ldots, \lambda_n$. The eigenvectors are different.
The job of the original eigenvectors $x_1, \ldots, x_n$ was to diagonalize A. Those eigenvectors in S produce $A = S\Lambda S^{-1}$.

- Example

Eigenvectors $\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$ $\underset{S^{-1}}{\begin{bmatrix}1&-1\\0&1\end{bmatrix}}\underset{A}{\begin{bmatrix}1&5\\0&6\end{bmatrix}}\underset{S}{\begin{bmatrix}1&1\\0&1\end{bmatrix}} = \underset{\Lambda}{\begin{bmatrix}1&0\\0&6\end{bmatrix}}$

$A = S\Lambda S^{-1}$ so $A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$

In general

$A^n = S\Lambda^n S^{-1}$ (Same eigenvectors in S)
(exponentiated eigenvalues in Λ)

For this A,

$A^k = \begin{bmatrix}1&5\\0&6\end{bmatrix}^k = \begin{bmatrix}1&1\\0&1\end{bmatrix}\begin{bmatrix}1^k&\\&6^k\end{bmatrix}\begin{bmatrix}1&-1\\0&1\end{bmatrix} = \begin{bmatrix}1&6^k-1\\0&6^k\end{bmatrix}$

$K=1 \Rightarrow \begin{bmatrix}1&5\\0&6\end{bmatrix}$

$K=0 \Rightarrow \begin{bmatrix}1&0\\0&1\end{bmatrix}$

$K=-1 \Rightarrow \begin{bmatrix}1&-\frac{5}{6}\\0&\frac{1}{6}\end{bmatrix} = A^{-1}$

- Notes:
  - If all eigenvalues are unique, then it is automatic that the eigenvectors are independent.
  Any matrix with unique eigenvalues is diagonalizable
  - Eigenvectors can be scaled arbitrarily without changing their eigenvector status.
  - The eigenvalues in Λ come in the same order as the eigenvector in S.
  - Matrices with too few eigenvectors cannot be diagonalized.
  - There is no connection between diagonalizability and invertibility.
    - Invertibility is concerned with the eigenvalues ($\lambda = 0$ or $\lambda \neq 0$)
    - Diagonalizability is concerned with the eigenvectors (too few or too many for S)
- Eigenvectors $x_1, \ldots, x_n$ that correspond to distinct eigenvalues are linearly independent.
  - Proof: suppose $c_1x_1 + c_2x_2 = 0$. Multiply by A to get $c_1\lambda_1x_1 + c_2\lambda_2x_2 = 0$. Multiply the original equation by λ_2 to get $c_1\lambda_2x_1 + c_2\lambda_2x_2 = 0$. Subtract them.
  We get $(\lambda_1 - \lambda_2)c_1x_1 = 0 \Rightarrow c_1 = 0$ since λ are all different and c_2 also equals 0.
  No other combination gives $c_1x_1 + c_2x_2 = 0$ so x_1 and x_2 must be independent.
  This proof extends to j eigenvectors

- Suppose $c_1x_1 + \dots + c_jx_j = 0$. Multiply by A and λ_j separately and subtract. This removes x_j. Repeat until $(\lambda_1-\lambda_2)(\lambda_1-\lambda_j)c_1x_1 = 0 \Rightarrow c_1 = 0$.
- Example:

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & .5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix} = S\Lambda S^{-1}$$

$$A^k = S\Lambda^k S^{-1} = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} 1^k & 0 \\ 0 & .5^k \end{bmatrix}\begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix}$$

$$\lim_{k\to\infty} A^k = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix} = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} .6 & .6 \\ .4 & .4 \end{bmatrix}$$

- $\lim_{k\to\infty} A^k = 0$ when all $|\lambda| < 1$

Fibonacci Numbers

- Challenge: Find F_{100}, the 100th Fibonacci number.
- Start with the equation $u_{k+1} = Au_k$

  Let $u_k = (F_{k+1}, F_k)$. The rule $\begin{cases} F_{k+2} = F_{k+1} + F_k \\ F_{k+1} = F_{k+1} \end{cases}$ is $u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k$
- Each step multiplies by $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. $u_{100} = A^{100}u_0$

$$u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ u_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ u_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},\ u_3 = \begin{bmatrix} 3 \\ 2 \end{bmatrix},\ \dots,\ u_{100} = \begin{bmatrix} F_{101} \\ F_{100} \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \Rightarrow \det(A-\lambda I) = (1-\lambda)(-\lambda) - (1)(1) = \lambda^2 - \lambda - 1$$

Using the quadratic formula, $\lambda = \frac{1+\sqrt{5}}{2}$, $\lambda = \frac{1-\sqrt{5}}{2}$

- Then our eigenvectors are $x_1 = (\lambda_1, 1)$ and $x_2 = (\lambda_2, 1)$.
- Then we find the combination of eigenvectors which produce $u_0 = (1,0)$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\lambda_1 - \lambda_2}\left(\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} - \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}\right) \quad \text{or} \quad u_0 = \frac{x_1 - x_2}{\lambda_1 - \lambda_2}$$

Then

$$u_{100} = A^{100}u_0 = \frac{A^{100}x_1 - A^{100}x_2}{\lambda_1 - \lambda_2} = \frac{(\lambda_1)^{100}x_1 - (\lambda_2)^{100}x_2}{\lambda_1 - \lambda_2}$$

We want F_{100} the second component of u_{100}. The second components of x_1 and x_2 are 1. Then

$$F_{100} = \frac{1}{\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{100} - \left(\frac{1-\sqrt{5}}{2}\right)^{100}\right] = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{100} - \left(\frac{1-\sqrt{5}}{2}\right)^{100}\right]$$

Matrix Powers of A^k

- Fibonacci's example is a typical difference equation $u_{k+1} = Au_k$. Each iterative step multiplies by A. The solution is $u_k = A^k u_0$.
  Diagonalizing matrices allow computation as $S\Lambda^k S^{-1} u_0 = u_k$
- Procedure
  1. Write u_0 as a combination $c_1 x_1 + \dots + c_n x_n$ of the eigenvectors. Then $c = S^{-1} v_0$
  2. Multiply each eigenvector x_i by $(\lambda_i)^k$. Now we have $\Lambda^k S^{-1} u_0$
  3. Add up all the pieces $c_i(\lambda_i)^k x_i$ to find the solution $u_k = A^k u_0$. This is $S\Lambda^k S^{-1} u_0$

  $u_{k+1} = Au_k \Rightarrow u_k = A^k u_0 = c_1(\lambda_1)^k x_1 + \dots + c_n(\lambda_n)^k x_n$
- Example 1

  Start from $u_0 = (1,0)$ with

  $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$ has $\lambda_1 = 2$ and $x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\lambda_2 = -1$ and $x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

  $u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ so $c_1 = c_2 = \frac{1}{3}$ (1)

  Multiply the two parts by $(\lambda_1)^k = 2^k$ and $(\lambda_2)^k = (-1)^k$

  $\frac{1}{3}(2)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3}(-1)^k \begin{bmatrix} 1 \\ -1 \end{bmatrix} = u_k$

Nondiagonalizable Matrices

- Eigenvalues may be repeated and we want to know the multiplicity. There's two different ways to count these: ($GM \leq AM \ \forall \lambda$)
  1. Geometric multiplicity = GM
     - Count the independent eigenvectors for λ. This is the dimension of $N(A-\lambda I)$
  2. Algebraic multiplicity = AM
     - Count the repetitions of λ in the roots of $\det(A-\lambda I) = 0$
- If A has $\lambda = 4,4,4$, then it has $AM = 3$, $GM = 1, 2,$ or 3, depending on whether each eigenvalue corresponds to an independent eigenvector or not.
- For

  $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\det(A-\lambda I) = \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2 \Rightarrow \lambda = 0, 0$ but only 1 eigenvector $\begin{pmatrix} AM=2 \\ GM=1 \end{pmatrix}$

  $A = \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$ has $\det(A-\lambda I) = (\lambda-5)^2 \Rightarrow \lambda = 5, 5$ but $A - 5I$ has only rank 1 $\begin{pmatrix} AM=2 \\ GM=1 \end{pmatrix}$

Eigenvalues of AB and $A+B$

- The following is false because we are assume A and B share the same eigenvalue x

  ~~$ABx = A\beta x = BAx = \beta\lambda x$~~
- Eigenvalues do not generally add either for $A+B$
- When all n eigenvectors are shared we can multiply eigenvalues
- Suppose both A and B are diagonalizable. They share S iff $AB = BA$

Problem Set 6.2

1.

a)

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

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

$\det(A - \lambda I) = (1-\lambda)(3-\lambda) - (2)(0)$

$= (\lambda-3)(\lambda-1)$

Set $\det(A-\lambda I) = 0$

$(\lambda-3)(\lambda-1) = 0$

$\lambda = 1, 3$

case $\lambda = 1$: $(A - I)x = 0$ $\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} x_1 = 0$ $x_1 = (1, 0)$

case $\lambda = 3$: $(A-3I)x = 0$ $\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} x_2 = 0$ $x_2 = (1, 1)$

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

$A = S \Lambda S^{-1}$

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

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

$A - \lambda I = \begin{bmatrix} 1-\lambda & 1 \\ 3 & 3-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (1-\lambda)(3-\lambda) - (1)(3)$

$= \lambda^2 - 4\lambda + 3 - 3$

$= \lambda^2 - 4\lambda$

$= \lambda(\lambda - 4)$

Set $\det(A - \lambda I) = 0$

$\lambda(\lambda-4) = 0$

$\lambda = 0, 4$

Case $\lambda = 0$: $(A - 0I)x_1 = 0$ $\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} x_1 = 0$ $x_1 = (1, -1)$

Case $\lambda = 4$: $(A - 4I)x_2 = 0$ $\begin{bmatrix} -3 & 1 \\ 3 & -1 \end{bmatrix} x_2 = 0$ $x_2 = (1, 3)$

$S = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$ $S^{-1} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

$A = S \Lambda S^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

b)

$A^3 = S \Lambda^3 S^{-1}$

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 27 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 27 \end{bmatrix}$

$= \begin{bmatrix} 1 & 26 \\ 0 & 27 \end{bmatrix}$

$A^{-1} = S \Lambda^{-1} S^{-1}$

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

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

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

2.

$A = S \Lambda S^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$

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

$= \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}$

4.

a) Maybe

b) True

c) True

d) Maybe

8.

$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{\lambda_2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{\lambda_1}{\sqrt{5}} \end{bmatrix}$

15.

$|\lambda| < 1$

$A_1 - \lambda I = \begin{bmatrix} \frac{3}{5} - \lambda & \frac{9}{10} \\ \frac{2}{5} & \frac{1}{10} - \lambda \end{bmatrix}$

$\det(A_1 - \lambda I) = (\frac{3}{5} - \lambda)(\frac{1}{10} - \lambda) - (\frac{9}{10})(\frac{2}{5})$

$= \lambda^2 - \frac{7}{10}\lambda - \frac{3}{10} = 0$

$10\lambda^2 - 7\lambda - 3 = 0$

$(\lambda - 1)(10\lambda + 3) = 0 \implies \lambda = 1, -\frac{3}{10}$

11.

a) F b) F c) T $\to \begin{bmatrix} 3 & -3 \\ 3 & -3 \end{bmatrix} x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

12.

a) F b) T c) T $x_1 = (1, 1)$

13.

$A = \begin{bmatrix} 8 & -3 \\ 3 & 2 \end{bmatrix}$

$Ax = 5x$

$(A - 5I)x$

14.

$A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix} \implies A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ A_2

$r = 0$

16.

$$A_1 = \begin{bmatrix} .6 & .9 \\ .4 & .1 \end{bmatrix}$$

$\lambda_1 = 1,\ \lambda_2 = -\frac{3}{10}$

$(A - I)x_1 = 0$ | $(A + \frac{3}{10}I)x_2 = 0$

$\begin{bmatrix} -.4 & .9 \\ .4 & -.9 \end{bmatrix} x_1 = 0$ $\begin{bmatrix} .9 & .9 \\ .4 & .4 \end{bmatrix} x_2 = 0$

$x_1 = (.9, .4)$ $x_2 = (1, -1)$

$S = \begin{bmatrix} .9 & 1 \\ .4 & -1 \end{bmatrix}$, $S^{-1} = \begin{bmatrix} \frac{10}{13} & \frac{10}{13} \\ \frac{4}{13} & -\frac{9}{13} \end{bmatrix}$

$\lim_{k\to\infty} \Lambda^k = \begin{bmatrix} 1^k & 0 \\ 0 & (-\frac{3}{10})^k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

$\lim_{k\to\infty} A^k = \lim_{k\to\infty} S\Lambda^k S^{-1}$

$= \begin{bmatrix} .9 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{10}{13} & \frac{10}{13} \\ \frac{4}{13} & -\frac{9}{13} \end{bmatrix}$

$= \begin{bmatrix} .9 & 1 \\ .4 & -1 \end{bmatrix}\begin{bmatrix} \frac{10}{13} & \frac{10}{13} \\ 0 & 0 \end{bmatrix}$

$= \begin{bmatrix} \frac{9}{13} & \frac{9}{13} \\ \frac{4}{13} & \frac{4}{13} \end{bmatrix}$

20.

$\det A = (\det S)(\det \Lambda)(\det S^{-1})$

$= (\det S)(\sum_i \lambda_i)(\frac{1}{\det S})$

$= \sum_i \lambda_i$

25.

$N(A - I)$ or $C(A), N(A)$

26.

~~$\lambda \neq 0$ / doesn't occupy the full column space~~

$C(A) \neq N(A)$ ∴ overlap

17.

$$A_2 = \begin{bmatrix} .6 & .9 \\ .1 & .6 \end{bmatrix}$$

$A - \lambda I = \begin{bmatrix} .6-\lambda & .9 \\ .1 & .6-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (.6-\lambda)(.6-\lambda) - (.1)(.9)$
$= \lambda^2 - \frac{6}{5}\lambda + \frac{9}{25} - \frac{9}{100}$
$= \lambda^2 - \frac{6}{5}\lambda + \frac{27}{100}$

set $\det(A - \lambda I) = 0$

$\lambda_1 = \frac{9}{10},\ \lambda_2 = \frac{3}{10}$

Case $\lambda_1 = \frac{9}{10}$ | Case $\lambda_2 = \frac{3}{10}$

$(A - \frac{9}{10}I)x_1 = 0$. $(A - \frac{3}{10}I)x_2 = 0$

$\begin{bmatrix} -.3 & .9 \\ .1 & -.3 \end{bmatrix} x_1 = 0$ $\begin{bmatrix} .3 & .9 \\ .1 & .3 \end{bmatrix} x_2 = 0$

$x_1 = (3, 1)$ $x_2 = (-3, 1)$

$S = \begin{bmatrix} 3 & -3 \\ 1 & 1 \end{bmatrix}$ $S^{-1} = \begin{bmatrix} \frac{1}{6} & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{2} \end{bmatrix}$ $\Lambda = \begin{bmatrix} \frac{9}{10} & \\ & \frac{3}{10} \end{bmatrix}$

a)

$u_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1x_1 + 0x_2 = x_1$

$(A_2)^{10} u_0 = (A_2)^{10}(x_1) = (\lambda_1)^{10}(x_1) = \left(\frac{9}{10}\right)^{10} x_1 = \begin{bmatrix} 3(\frac{9}{10})^{10} \\ (\frac{9}{10})^{10} \end{bmatrix}$

b)

$u_0 = \begin{bmatrix} 3 \\ -1 \end{bmatrix} = -x_2$

$(A_2)^{10} u_0 = (\lambda_2)^{10}(-x_2) = (\frac{3}{10})^{10}(-x_2) = \begin{bmatrix} -3(\frac{3}{10})^{10} \\ (\frac{3}{10})^{10} \end{bmatrix}$

c)

$u_0 = \begin{bmatrix} 6 \\ 0 \end{bmatrix} = x_1 + x_2$

$A^{10} u_0 = A^{10}(x_1 + x_2) = A^{10}x_1 + A^{10}x_2 = \lambda_1^{10}x_1 + \lambda_2^{10}x_2$
$= (\frac{9}{10})^{10}x_1 + (\frac{3}{10})^{10}x_2$
$= \begin{bmatrix} 3(\frac{9}{10})^{10} + 3(\frac{3}{10})^{10} \\ (\frac{9}{10})^{10} - (\frac{3}{10})^{10} \end{bmatrix}$

6.3: Applications to Differential Equations

- Eigen-values, -vectors, and -decompositions are also perfect for differential equations $du/dt = Au$. Our most important derivative will be $\frac{d}{dt}(e^{\lambda t}) = \lambda e^{\lambda t}$
- $\frac{du}{dt} = \lambda u$ is solved by $u = Ce^{\lambda t}$. We can solve for C given an initial condition
- Now we want to solve n equations

  $\frac{du}{dt} = Au$ with initial condition $u(0)$ at $t=0$

- These differential equations are linear. If $u(t)$ and $v(t)$ are solutions, so is $Cu(t) + Dv(t)$
- We will need n constants like C and D to match the n components of $u(0)$
- Note that A is a constant matrix. We usually have A change as t changes (linear) or A change when u changes (nonlinear). Here we have $\frac{du}{dt} = Au$ as "linear with constant coefficients"
- Our main point will be:

<u>Solve linear constant coefficient equations by exponentials $e^{\lambda t}x$ when $Ax = \lambda x$.</u>

Solution of $du/dt = Au$

- Our pure exponential solution will be $e^{\lambda t}$ times a fixed vector x. λ is an eigenvalue of A and x is the eigenvector.

$$\frac{d}{dt}(e^{\lambda t}x) = A(e^{\lambda t}x) \Rightarrow \lambda e^{\lambda t}x = \lambda e^{\lambda t}x$$

- All components of this solution $u = e^{\lambda t}x$ share the same $e^{\lambda t}$ part. The solution grows when $\lambda > 0$, decays if $\lambda < 0$. If λ is complex, its real parts decides growth or decay, and its imaginary part gives oscillation.
- Example

Solve $\frac{du}{dt} = Au = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with initial condition $u(0) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$

$\frac{du}{dt} = Au \quad \frac{d}{dt}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix}$ means that $\frac{dy}{dt} = z$ and $\frac{dz}{dt} = y$

The idea of eigenvectors is to combine those equations in a way that gets it to a 1 equation ("1 by 1") problem

The matrix A has eigenvalues 1 and -1 and eigenvectors $(1,1)$, $(1,-1)$. The pure exponential solutions takes the form

$$u_1(t) = e^{\lambda_1 t}x_1 = e^{t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad u_2(t) = e^{\lambda_2 t}x_2 = e^{-t}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- Note: These u's are eigenvectors. They satisfy $Au_1 = u_1$ and $Au_2 = -u_2$ just like x_1 and x_2. The factors e^{t} and e^{-t} change with time. These factors give $du_1/dt = u_1 = Au_1$ and $du_2/dt = -u_2 = Au_2$. We now have 2 solutions. To get all solutions, take all linear combinations of those 2

$$u(t) = Ce^{t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + De^{-t}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} Ce^{t} + De^{-t} \\ Ce^{t} - De^{-t} \end{bmatrix}$$

$u(0)$ gives C, D

- Procedure, summarized:

1. Write $u(0)$ as a combination $c_1x_1 + \ldots + c_nx_n$ of the eigenvectors of A
2. Multiply each eigenvector x_i by $e^{\lambda_i t}$
3. The solution is the combination of pure solutions $e^{\lambda t}x$

$$u(t) = c_1e^{\lambda_1 t}x_1 + \ldots + c_ne^{\lambda_n t}x_n$$

- If the two λ's are equal, with only one eigenvector, another solution is needed (It will be $te^{\lambda t}x$). Step 1 needs $A = S\Lambda S^{-1}$ to be diagonalizable
- Example

Solve $du/dt = Au$ knowing the eigenvalues $\lambda = 1, 2, 3$ of A

$$\frac{du}{dt} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix} u \text{ with } u(0) = \begin{bmatrix} 9 \\ 7 \\ 4 \end{bmatrix}$$

The eigenvectors are $x_1 = (1, 0, 0)$, $x_2 = (1, 1, 0)$, and $x_3 = (1, 1, 1)$

$$u(0) = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2x_1 + 3x_2 + 4x_3 \text{. Thus } (c_1, c_2, c_3) = (2, 3, 4)$$

Then,

The pure exponential solutions are $e^{t}x_1$, $e^{2t}x_2$, and $e^{3t}x_3$

The solution is

$$u(t) = 2e^{t}x_1 + 3e^{2t}x_2 + 4e^{3t}x_3$$

Second order Equations

- The equation $my'' + by' + ky = 0$ is very important in mechanics.
  - The first term $my'' = ma$ balances the force F. The force includes the damping $-by'$ and the elastic restoring force $-ky$.
- In a differential equations course, the method of solution is to substitute $y = e^{\lambda t}$

Each derivative brings a factor λ. We want $y = e^{\lambda t}$

$$m\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = 0 \text{ becomes } (m\lambda^2 + b\lambda + k)e^{\lambda t} = 0$$

Everything depends on $m\lambda^2 + b\lambda + k = 0$. The equation for λ has 2 roots λ_1 and λ_2.

Then the equation for y has 2 pure solutions $y_1 = e^{\lambda_1 t}$ and $y_2 = e^{\lambda_2 t}$.

Their combinations $c_1y_1 + c_2y_2$ give the complete solution unless $\lambda_1 = \lambda_2$

- We can convert this equation into a vector one. Suppose $m = 1$

$$\begin{cases} \frac{dy}{dt} = y' \\ \frac{dy'}{dt} = -ky - by' \end{cases} \text{ converts to } \frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k & -b \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix}$$

  - The second equation connects y'' to y' and y. Together, the equation connects u' to u.
  - $A - \lambda I = \begin{bmatrix} -\lambda & 1 \\ -k & -b-\lambda \end{bmatrix}$ has determinant $\lambda^2 + b\lambda + k = 0$. The equation for the λ are the same. The roots λ_1 and λ_2 are now the eigenvalues of A.
  - The eigenvectors and the solution are

$$x_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \quad x_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} \quad u(t) = c_1e^{\lambda_1 t}\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} + c_2e^{\lambda_2 t}\begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$$

- Example 3 Motion around a circle with $y'' + y = 0$ and $y = \cos t$

  $y'' + y = 0 = my'' + by' + ky$ with $m = 1, b = 0, k = 1$

  Substitute $y = e^{\lambda t}$

  $\lambda^2 e^{\lambda t} + e^{\lambda t} = 0$

  $e^{\lambda t}(\lambda^2 + 1) = 0$

  The roots are $\lambda = i$ and $\lambda = -i$. Then half of $e^{it} + e^{-it}$ gives the solution $y = \cos t$

  As a matrix system

  $y(0) = 1$, $y'(0) = 0$ go into $u(0) = (y(0), y'(0)) = (1, 0)$

  use $y'' = -y$ $\quad \frac{du}{dt} = \frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix} = Au$

  The eigenvalues of A are again $\lambda = i$ and $\lambda = -i$. A is anti-symmetric with the eigenvectors $x_1 = (1, i)$ and $x_2 = (1, -i)$. The combination that matches $u(0) = (1, 0)$ is $\frac{1}{2}x_1 + \frac{1}{2}x_2$.

  $$u(t) = \frac{1}{2}e^{it}\begin{bmatrix} 1 \\ i \end{bmatrix} + \frac{1}{2}e^{-it}\begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} \cos t \\ -\sin t \end{bmatrix} = \begin{bmatrix} y(t) \\ y'(t) \end{bmatrix}$$

  The vector $u = (\cos t, -\sin t)$

Stability of 2 by 2 Matrices

- For the solution $du/dt = Au$, there is a fundamental question. Does the solution approach $u = 0$ as $t \to \infty$. Is the problem stable? The examples in previous sections included e^t (unstable). Stability depends on the eigenvalues of A.
- The complete solution $u(t)$ is built from pure solutions $e^{\lambda t}x$. If $\lambda \in \mathbb{R}$, $e^{\lambda t}$ will approach zero if λ is negative
- If $\lambda \in \mathbb{C}$, $\lambda = r + is$, the real part r must be negative. When $e^{\lambda t}$ splits into $e^{rt}e^{ist}$, the factor e^{ist} has absolute value fixed at 1.

  $e^{ist} = \cos st + i\sin st$ has $|e^{ist}|^2 = \cos^2 st + \sin^2 st = 1$

  The factor e^{rt} controls growth ($r > 0$ is instability) or decay ($r < 0$ is stability)
- The question is when are the real parts of λ all negative
- Stability

  A is stable and $\lim_{t \to \infty} u(t) = 0 \iff \mathrm{Re}(\lambda) < 0 \ \forall \lambda_i$

  The 2 by 2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ must pass two tests:

  1. $\lambda_1 + \lambda_2 < 0$. The trace $a + d$ must be negative
  2. $\lambda_1 \lambda_2 > 0$. The determinant must be positive

The Exponential of a Matrix

- We want to write the solution $u(t)$ in a new form $e^{At}u(0)$
- To do this, we use the Taylor series

$$e^{x} = \sum_{n=1}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

$$e^{x} = \sum_{1} \frac{x^n}{n!} \quad \text{(Maclaurin)}$$

$$e^{At} = I + At + \tfrac{1}{2}(At)^2 + \tfrac{1}{6}(At)^3 + \dots$$

Its derivative is

$$Ae^{At} = A + A^2t + \tfrac{1}{2}A^3t^2 + \dots$$

Its eigenvalues are $e^{\lambda t}$

$$(I + At + \tfrac{1}{2}(At)^2 + \dots)x = (1 + \lambda t + \tfrac{1}{2}(\lambda t)^2 + \dots)x$$

- $e^{At}u(0)$ solves the differential equation even if there is a shortage of eigenvectors
- Assume A is diagonalizable, then

$$e^{At} = e^{S\Lambda S^{-1}t} = I + S\Lambda S^{-1}t + \tfrac{1}{2}(S\Lambda S^{-1}t)(S\Lambda S^{-1}t) + \dots$$

$$= S[I + \Lambda t + \tfrac{1}{2}(\Lambda t)^2 + \dots]S^{-1}$$

$$= Se^{\Lambda t}S^{-1}$$

- $e^{\Lambda t}$ equals $Se^{\Lambda t}S^{-1}$. Then Λ is a diagonal matrix and so is $e^{\Lambda t}$. The numbers $e^{\lambda_i t}$ are on its diagonal. Multiply by $u(0)$ to get $u(t)$

$$e^{At}u(0) = Se^{\Lambda t}S^{-1}u(0) = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

- Example

Solve $du/dt = Au = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} u$ starting from $u(0) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$\lambda_1 = 1 \Rightarrow x_1 = (1, 0)$

$\lambda_2 = 2 \Rightarrow x_2 = (1, 1)$

$u(0) = x_1 + x_2 \Rightarrow u(t) = e^{t}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + e^{2t}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For every $u(0)$

$$u(t) = Se^{\Lambda t}S^{-1}u(0) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{t} & \\ & e^{2t} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} u(0) = \begin{bmatrix} e^{t} & e^{2t} - e^{t} \\ 0 & e^{2t} \end{bmatrix} u(0)$$

Chapter 6.4: Symmetric Matrices

- For projection onto a plane in $\mathbb{R}^3$, the plane is full of eigenvectors (where $Px = x$). The other eigenvectors are perpendicular to the plane (where $Px = 0$)
- The eigenvalues $\lambda = 1, 1, 0$ are real. Three eigenvectors can be chosen perpendicular to each other
- What is special about $Ax = \lambda x$ when A is symmetric?
  If we diagonalize $A = S\Lambda S^{-1}$,
  $A^T = (S^{-1})^T(\Lambda)^T(S)^T = (S^{-1})^T(\Lambda)(S^T)$.

  Since
  $A^T = A$,
  $A = (S^{-1})^T \Lambda S^T = S\Lambda S^{-1} \Rightarrow S^{-1} = S^T$ and $S^TS = I$, which makes S orthonormal
- Fact:
  1. A symmetric matrix has only real eigenvalues
  2. The eigenvalues can be chosen orthonormal
- Those n orthonormal eigenvectors go into the columns of S. Every symmetric matrix can be diagonalized. Its eigenvector matrix S becomes an orthonormal matrix Q.
- We "chose" because any scalar multiple of an eigenvector is itself an eigenvector. We pick eigenvectors of unit length. Then for symmetric matrices, $S\Lambda S^{-1}$ is in its special form $Q\Lambda Q^T$
- Spectral Theorem:
  Every symmetric matrix has the factorization $A = Q\Lambda Q^T$ with real eigenvalues in Λ and orthonormal eigenvectors in $S = Q$:

  Symmetric diagonalization $\quad A = Q\Lambda Q^{-1} = Q\Lambda Q^T$ with $Q^{-1} = Q^T$

- It's easy to see $Q\Lambda Q^T$ is symmetric $(Q\Lambda Q^T)^T = (Q^T)^T\Lambda^T Q^T = Q\Lambda Q^T$.
  The harder part is to prove every symmetric matrix has real λ's and orthonormal eigenvectors.
- First, an example!
  Find the λ's and x's when $A = \begin{bmatrix}1&2\\2&4\end{bmatrix}$
  $\det(A - \lambda I) = \lambda^2 - 5\lambda = 0$
  $\lambda = 0, 5$ (both real)
- Two eigenvectors are $(2, -1)$ and $(1, 2)$ — orthogonal but not yet orthonormal.
- The eigenvector for $\lambda = 0$ is in the nullspace of A. The eigenvector for $\lambda = 5$ is in the column space. The fundamental Theorem says the row space and nullspace are orthogonal. Since $A^T = A$, the row space is equivalent to the column space, so our 2 eigenvectors must be orthogonal.
- To make them orthonormal, we just divide each one by its length ($\sqrt{5}$)
- Then, to get Λ

$$\Lambda = Q^{-1}AQ = \frac{1}{\sqrt5}\begin{bmatrix}2&-1\\1&2\end{bmatrix}\begin{bmatrix}1&2\\2&4\end{bmatrix}\frac{1}{\sqrt5}\begin{bmatrix}2&1\\-1&2\end{bmatrix} = \begin{bmatrix}0&0\\0&5\end{bmatrix} = \Lambda$$

- Proof: All eigenvalues of a real symmetric matrix are real.
  Suppose that $Ax = \lambda x$, with $\lambda \in \mathbb{C}$. Its complex conjugate is $\bar{\lambda} = a - ib$.
  x is also a complex vector ($\mathbb{C}^n$), and taking the complex conjugate of each component gives $\bar{x}$. $\bar{\lambda}\bar{x} = \overline{(\lambda x)}$. We take the conjugate of both sides, with A being a real symmetric matrix

  $Ax = \lambda x$ leads to $A\bar{x} = \bar{\lambda}\bar{x}$. Transpose to $\bar{x}^T A = \bar{x}^T \bar{\lambda}$.

  Now take the dot product of the first equation with $\bar{x}$ and the second equation with x:

  $\bar{x} A x = \bar{x}^T \lambda x$ and $\bar{x}^T A x = \bar{x}^T \bar{\lambda} x$

  The left sides are equal so the right sides are also equal. One equation has λ and another has $\bar{\lambda}$. They multiply $\bar{x}^T x = |x_1|^2 + |x_2|^2 =$ length squared which is not zero.
  Therefore λ must equal $\bar{\lambda}$ and $b = 0$ (QED)
- Orthogonal Eigenvectors: Eigenvectors of a real symmetric matrix (when they correspond to different eigenvalues) are always perpendicular
  - Proof: Suppose $Ax = \lambda_1 x$ and $Ay = \lambda_2 y$. We are assuming that $\lambda_1 \neq \lambda_2$
    Take dot products of the first equation with y and the second with x.

    $(\lambda_1 x)^T y = (Ax)^T y = x^T A^T y = x^T (Ay) = x^T \lambda_2 y$

    The left side is $x^T \lambda_1 y$, the right side is $x^T \lambda_2 y$. Since $\lambda_1 \neq \lambda_2$, then $x^T y = 0$, which means the two eigenvectors x and y are orthogonal.
- Example 2:
  The eigenvectors of a 2 by 2 symmetric matrix have a special form

  $$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \text{ has } x_1 = \begin{bmatrix} b \\ \lambda_1 - a \end{bmatrix} \text{ and } x_2 = \begin{bmatrix} \lambda_2 - c \\ b \end{bmatrix}$$

  $$x_1^T x_2 = \begin{bmatrix} b & \lambda_1 - a \end{bmatrix} \begin{bmatrix} \lambda_2 - c \\ b \end{bmatrix} = b(\lambda_2 - c) + b(\lambda - a) = b(\lambda + \lambda_2 - a - c) = 0$$

  $\lambda_1 + \lambda_2$ is the trace, which is equal to $a + c$
- From $A = Q\Lambda Q^T$ with $Q^T Q = I$, every 2 by 2 symmetric matrix looks like

  $$A = Q\Lambda Q^T = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix}$$

  The columns x_1 and x_2 multiply the rows $\lambda_1 x_1^T$ and $\lambda_2 x_2^T$ to produce A.
  We can write A as the sum of rank one matrices
  $A = \lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T$

- When the symmetric matrix is n by n, there are n columns in Q multiplying n rows in Q^T. The n products $x_i x_i^T$ are projection matrices

- Including the eigenvalues the spectral theorem $A = Q\Lambda Q^T$ for symmetric matrices says that A is a combination of projection matrices

$$A = \lambda_1 P_1 + \dots + \lambda_i P_i + \dots + \lambda_n P_n \quad \lambda_i = \text{eigenvalue} \quad P_i = \text{projection onto eigenspace}$$

Complex Eigenvalues of Real Matrices

- If A is symmetric, then its eigenvalues are real. A nonsymmetric matrix can easily produce complex eigenvalues and eigenvectors. In this case $A\bar{x} = \lambda\bar{x}$ is different from $Ax = \lambda x$. It gives us a new eigenvalue $(\bar{\lambda})$ and a new eigenvector $(\bar{x})$.
- For real matrices complex λ's and x's come in conjugate pairs.
- Example:

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ has } \lambda_1 = \cos\theta + i\sin\theta \text{ and } \lambda_2 = \cos\theta - i\sin\theta$$

$$\lambda x: \quad Ax = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 \\ -i \end{bmatrix} = (\cos\theta + i\sin\theta)\begin{bmatrix} 1 \\ -i \end{bmatrix}$$

$$\bar{\lambda}\bar{x}: \quad A\bar{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 \\ i \end{bmatrix} = (\cos\theta - i\sin\theta)\begin{bmatrix} 1 \\ i \end{bmatrix}$$

The eigenvectors $(1, -i)$ and $(1, i)$ are complex conjugates because A is real.
- $|\lambda| = 1$ for the eigenvalues of every orthogonal matrix

Eigenvalues versus Pivots

- The eigenvalues of A are very different from pivots. The only connection we have thus far is:

product of pivots = determinant = product of eigenvalues

  - We are assuming n pivots and n real eigenvalues.
- For symmetric matrices the pivots and the eigenvalues have the same sign
  - The number of positive eigenvalues of $A = A^T$ equals the number of positive pivots
  - Special case: A has all $\lambda_i > 0$ if and only if all pivots are positive
- Example

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \text{ has } \begin{matrix} \text{Pivots } 1 \text{ and } -8 \\ \text{Eigenvalues } 4 \text{ and } -2. \end{matrix}$$

- Proof:

If $A^T = A$ and A is invertible, A can be factorized as $A = LDL^T$

$$\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -8 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

  - The eigenvalues of LDL^T are 4 and -2. The eigenvalues of IDI^T are 1 and -8 (the pivots). The eigenvalues are changing, as the "3" in L moves to zero, but to change sign, a real eigenvalue would have to cross 0, making it singular. Our matrix always has pivots 1 and -8 so it is never singular. The signs cannot change.

All Symmetric Matrices are Diagonalizable

- A repeated eigenvalue can lead to a shortage of eigenvectors for diagonalization. This happens sometimes for nonsymmetric matrices. It never happens for symmetric matrices. There are always enough eigenvectors to diagonalize $A = A^T$
- Proof:
  - Every square matrix factors into $A = QTQ^{-1}$ where T is upper triangular and $\bar{Q}^T = Q^{-1}$
  - If A has real eigenvalues then Q and T can chosen real: $Q^TQ = I$
  - This is Schur's Theorem, We are looking for $AQ = QT$. The first q_1 of Q must be a unit eigenvector of A. Then the first columns of AQ and QT are Aq_1 and $t_1 q_1$. But the other columns of Q need not be eigenvectors when T is only triangular.
  - So use any $n-1$ columns that complete q_1 to a Q_1 with orthonormal columns. At this point only the first columns of Q and T are set, where $Aq_1 = t_1 q_1$.

$$\bar{Q}_1^T A Q_1 = \begin{bmatrix} \bar{q}_1^T \\ \vdots \\ \bar{q}_n^T \end{bmatrix} \begin{bmatrix} Aq_1 & \cdots & Aq_n \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & \\ 0 & A_2 & \\ 0 & & \end{bmatrix}$$

  - By induction, we can repeat this for submatrix A_2
  - When A is symmetric, T is the diagonal Λ when A is symmetric. This is because T is both symmetric and triangular, so it is diagonal.

Problem set 6.4

1.

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 3 & 0 \\ 8 & 6 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 6 \\ 3 & 3 & 3 \\ 6 & 3 & 5 \end{bmatrix} + \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -3 \\ 2 & 3 & 0 \end{bmatrix}$$

2.

$(A^TCA)^T = A^TC^T(A^T)^T = A^TCA$

$C = 3$ by 3

$(A^T)(A) = 6$ by 6

3.

$$A - \lambda I = \begin{bmatrix} 2-\lambda & 2 & 2 \\ 2 & -\lambda & 0 \\ 2 & 0 & -\lambda \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & -\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{6} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{6} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \end{bmatrix}$$

$$\begin{aligned} \det(A - \lambda I) &= (2-\lambda)(-\lambda)(-\lambda) - (2)(2)(-\lambda) - (2)(2)(-\lambda) \\ &= -\lambda^3 + 2\lambda^2 + 8\lambda \\ &= -\lambda(\lambda^2 - 2\lambda - 8\lambda) \\ &= -\lambda(\lambda - 4)(\lambda + 2) \end{aligned}$$

set $\det(A - \lambda I) = 0$

$\lambda = 0, -2, 4$

case $\lambda = 0$

$(A - 0I)x_1 = 0$

$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} x_1 = 0$

$x_1 = (0, -1, 1)$

$q_1 = (0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$

case $\lambda = -2$:

$(A + 2I)x_2 = 0$

$\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} x_2 = 0$

$x_2 = (-1, 1, 1)$

$q_2 = (-\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3})$

case $\lambda = 4$

$(A - 4I)x_3 = 0$

$\begin{bmatrix} -2 & 2 & 2 \\ 2 & -4 & 0 \\ 2 & 0 & -4 \end{bmatrix} x_3 = 0$

$x_3 = (2, 1, 1)$

$q_3 = (\frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6})$

4.

$$\begin{aligned} \det(A - \lambda I) &= (-2-\lambda)(7-\lambda) - (6)(6) \\ &= \lambda^2 - 5\lambda - 14 - 36 \\ &= \lambda^2 - 5\lambda - 50 \\ &= (\lambda - 10)(\lambda + 5) \end{aligned}$$

$\lambda = -5, 10$

Case $\lambda = 10$:

$\begin{bmatrix} -12 & 6 \\ 6 & -3 \end{bmatrix} x_1 = 0$

$x_1 = (1, 2)$

$q_1 = (\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5})$

case $\lambda = -5$:

$\begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix} x_2 = 0$

$x_2 = (2, -1)$

$q_2 = (\frac{2\sqrt{5}}{5},$

$Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$

11.

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

$$\det(A-\lambda I) = (3-\lambda)(3-\lambda) - (1)(1)$$
$$= 9 - 6\lambda + \lambda^2 - 1$$
$$= \lambda^2 - 6\lambda + 8$$
$$= (\lambda-4)(\lambda-2)$$

$\det(A-\lambda I) = 0 \Rightarrow \lambda = 2, 4$

Case $\lambda = 2$: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x_1 = 0$, $x_1 = (1, -1)$, $q_1 = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$

Case $\lambda = 4$: $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} x_2 = 0$, $x_2 = (1, 1)$, $q_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$

$$A = Q\Lambda Q^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$
$$= 2 \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} + 4 \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

19.

orthonormal, $i, i, -i, -i$

21.

a) False

b) True

c) True

d) False

23.

A: Invertible, Permutation

B:

26.

$$\begin{bmatrix} 1 & 10^{-15} \\ 0 & 1+10^{-15} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$\lambda = 1$, $\lambda = 1 + 10^{-15}$

$$\begin{bmatrix} 10^{-15} & 10^{-15} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0$$

$$\theta = \cos^{-1}\left(\frac{(1)}{\sqrt{2}\sqrt{1}}\right) = 45°$$

Chapter 6.5: Positive Definite Matrices

- This chapter focuses on symmetric matrices that have positive eigenvalues.
- Symmetric matrices with positive eigenvalues are called positive definite matrices
- We can determine if a matrix is positive definite without calculating eigenvalues
- Remember, the eigenvalues are real because the matrix is symmetric.
- Start with 2 by 2. When does $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ have $\lambda_1 > 0$ and $\lambda_2 > 0$
  - The eigenvalues of A are positive if and only if $a > 0$ and $ac - b^2 > 0$

    $A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is not positive definite because $ac - b^2 = (1)(1) - (2)^2 = -3 < 0$

    $A_2 = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix}$ is positive definite because $ac - b^2 = (1)(6) - (-2)^2 = 2 > 0$

    $A_3 = \begin{bmatrix} -1 & 2 \\ 2 & -6 \end{bmatrix}$ is not positive definite because $a = -1$
  - Proof: The determinant $(ac - b^2)$ is the product of the pivots. If it is positive, λ_1 and λ_2 must be the same sign. The trace $a + c$ is the sum of the eigenvalues is also positive and a, c are both positive as well (or else $ac - b^2$ fails)
  - This test uses the 1 by 1 determinant and 2 by 2 determinant. A 3 by 3 matrix will use the 1 by 1, 2 by 2, and 3 by 3 determinant

- Another test:
  - The eigenvalues of $A = A^T$ are positive iff the pivots are all positive

    $$a > 0 \text{ and } \frac{ac-b^2}{a} > 0$$

  - This is a collarary of the first test, but we should recognize a and $\frac{ac-b^2}{a}$ are pivots

    $$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \longrightarrow \begin{bmatrix} a & b \\ 0 & -\frac{b^2}{a}+c \end{bmatrix} \quad -\frac{b^2}{a}+c = \frac{ac-b^2}{a}$$

  - Pivots are a lot easier and faster to compute than eigenvalues

Energy-based Definition

- From $Ax = \lambda x$, multiply by x^T to get $x^TAx = \lambda x^Tx$. The right is a positive λ times a positive number $x^Tx = \|x\|^2$. So x^TAx is positive for any eigenvector.
- The new idea is that x^TAx is positive for all nonzero vectors x, not just eigenvectors
- In many applications, the number x^TAx (or $\frac{1}{2}x^TAx$) is the energy in the system. The requirement of positive energy is another definition of a positive definite matrix.
  - Eigenvalues and pivots are two equivalent ways to test the new requirements $x^TAx > 0$
  - A is positive definite if $x^TAx > 0$ for every nonzero vector x:

    $$x^TAx = [x \; y]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = [x \; y]\begin{bmatrix} ax+by \\ bx+cy \end{bmatrix} = ax^2 + 2bxy + cy^2 > 0$$

  - If A and B are symmetric positive definite, so is $A+B$
    - $x^T(A+B)x = x^TAx + x^TBx$. The energies add, so the result has positive energy as well
- Start any matrix R, possibly rectangular. We know that $A = R^TR$ is square and symmetric.
  - If the columns of R are independent, then $A = R^TR$ is positive definite.

    $x^TAx = x^TR^TRx = (Rx)^T(Rx)$. The vector Rx is not zero when $x \neq 0$ (b/c independent columns). Then $x^TAx = \|Rx\|^2 > 0$ and A is positive definite.
- When a symmetric matrix has one of these five properties, it has them all
  1. All n pivots are positive
  2. All k by k, $k \in \mathbb{Z}$, $k \in [0, n]$ upper left determinants are positive
  3. All n eigenvalues are positive
  4. x^TAx is positive for all x except $x = 0$.
  5. A equals R^TR for a matrix R with independent columns
- Example

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

  - Pivots: $2, \frac{3}{2}, \frac{4}{3}$
  - Upper left determinants: $2, 3, 4$
  - Eigenvalues: $2-\sqrt{2}, 2, 2+\sqrt{2}$
  - $x^TAx = 2(x_1^2 - x_1x_2 + x_2^2 + x_2x_3 + x_3^2) = 2(x_1 - \frac{1}{2}x_2)^2 + \frac{3}{2}(x_2 - \frac{2}{3}x_3)^2 + \frac{4}{3}(x_3)^2$
    - All squares, so $x^TAx > 0$
  - $R = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

- We can also choose R using $A = LDL^T$

$$LDL^T = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ & 1 & -\frac{2}{3} \\ & & 1 \end{bmatrix} = (L\sqrt{D})(L\sqrt{D})^T = R^T R$$

- This R has square roots, but it is the only R that is 3 by 3 and upper triangular. It is the "Cholesky factor" of A
- In applications, the rectangular R is how we build A and this Cholesky R is how we break it apart
- Eigenvalues give the symmetric choice $R = Q\sqrt{\Lambda}Q^T$

Positive Semidefinite Matrices

- Often we are at the edge of positive definiteness. The determinant is zero. The smallest eigenvalue is zero. The energy in its eigenvector is $x^T A x = 0$. These matrices are called <u>positive semidefinite</u>
  - Here are two examples:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

A has eigenvalues 0 and 5. Its upper left determinants are 1 and 0. Its rank is only 1. This matrix A factors in $R^T R$ with dependent columns in R.

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

- Positive semidefinite matrices have all $\lambda \geq 0$ and $x^T A x \geq 0$.

First Application: The Ellipse $ax^2 + 2bxy + cy^2 = 1$

- Think of a tilted ellipse $x^T A x = 1$. Its center is $(0,0)$ as in fig 6.5.1. Turn it to align it with the x and y axes. These 2 ellipses show the geometry behind $A = Q \Lambda Q^T$.
  1. The tilted ellipse is associated with A. Its equation is $x^T A x = 1$.
  2. The lined-up ellipse is associated with Λ. Its equation is $X^T \Lambda X = 1$
  3. The rotation matrix that lines up the ellipse is the eigenvector matrix Q
- Example
  - Find the axes of the tilted ellipse $5x^2 + 8xy + 5y^2 = 1$
  - Start with the positive definite matrix that matches this equation

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1 \Rightarrow \text{the matrix is } A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

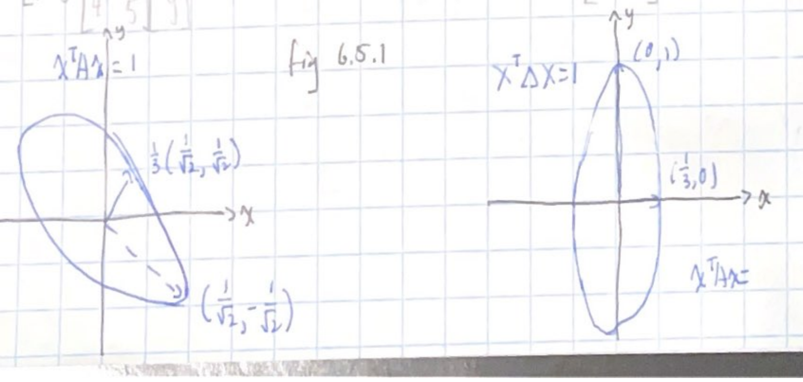


fig 6.5.1

- The eigenvectors are $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\-1\end{bmatrix}$. Divide by $\sqrt{2}$ to get unit vectors. Then $A=Q\Lambda Q^T$

$$\begin{bmatrix}5&4\\4&5\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}\begin{bmatrix}9&0\\0&1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$$

- Now multiply by $[x\ y]$ on the left and $\begin{bmatrix}x\\y\end{bmatrix}$ on the right to get to x^TAx

$$[x\ y]\begin{bmatrix}5&4\\4&5\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=5x^2+8xy+5y^2=9\left(\frac{x+y}{\sqrt{2}}\right)^2+1\left(\frac{x-y}{\sqrt{2}}\right)^2$$

- These coefficients are from the eigenvalues 9 and 1 from Λ. Inside the squares are the eigenvectors $\frac{(1,1)}{\sqrt{2}}$ and $\frac{(1,-1)}{\sqrt{2}}$.

- The axes of the tilted ellipse point along the eigenvectors. This $A=Q\Lambda Q^T$ is called the "Principal axis theorem". It also gives the axis lengths (from the eigenvalues).
  - To align the ellipse,

    Let $\frac{x+y}{\sqrt{2}}=X$ and $\frac{x-y}{\sqrt{2}}=Y$ and $9X^2+Y^2=1$
  - The largest value of $X^2=\frac{1}{9}$. The endpoint of the shorter axis has $X=\frac{1}{3}$ and $Y=0$. The bigger eigenvalue gives the shorter axis, of half-length $1/\sqrt{\lambda_1}=\frac{1}{3}$. The smaller eigenvalue $\lambda_2=1$ gives the greater length $1/\sqrt{\lambda_2}=1$.
  - In the xy system, the axes are along the eigenvectors of A. In the XY system, the axes are along the eigenvectors of Λ, the coordinate axes.
- Suppose $A=Q\Lambda Q^T$ is positive definite, so $\lambda_i>0$. The graph of $x^TAx=1$ is an ellipse:

$$[x\ y]Q\Lambda Q^T\begin{bmatrix}x\\y\end{bmatrix}=[X\,Y]\Lambda\begin{bmatrix}X\\Y\end{bmatrix}=\lambda_1X^2+\lambda_2Y^2=1$$

The axes point along eigenvectors. The half lengths are $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$

Problem Set 6.5

2.

A_4

3.

$$A=\begin{bmatrix}1&2\\2&9\end{bmatrix}=\begin{bmatrix}1&\\2&1\end{bmatrix}\begin{bmatrix}1&\\&4\end{bmatrix}\begin{bmatrix}1&2\\0&1\end{bmatrix}$$

4.

$$A=\begin{bmatrix}1&2\\2&9\end{bmatrix}$$

$$x^TAx=[x\,y]\begin{bmatrix}1&2\\2&9\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=[x\ y]\begin{bmatrix}x+2y\\2x+9y\end{bmatrix}=x^2+4xy+9y^2$$

6.

$f(x,y)=2xy$

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

$\lambda=-1,1$ $\quad(-\lambda)(-\lambda)-1=\lambda^2-1$

8.

$$A=\begin{bmatrix}3&6\\6&16\end{bmatrix}$$

a.

$x_1^2-2x_1(x_2+2x_3)+(x_2-2x_3)^2$

$x_1^2-2x_1x_2+4x_1x_3$

d,

$r=1$

$p=4$

15

$$[x_1\ x_2\ x_3]\begin{bmatrix}4&1&1\\1&0&2\\1&2&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = [x_1\ x_2\ x_3]\begin{bmatrix}4x_1+x_2+x_3\\x_1+2x_3\\x_1+2x_2+5x_3\end{bmatrix}$$

$= 4x_1^2 + x_1x_2 + x_1x_3 + x_1x_2 + 2x_2x_3 + x_1x_3 + 2x_2x_3 + 5x_3^2$

$= 4x_1^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3 + 5x_3^2$

choose $x = (1, -5, 0)$

$x^TAx = 4(1)^2 + 2(1)(-5) = -6$

18.

$Ax = \lambda x$

$x^TAx = x^T(\lambda x) = \lambda x^Tx = \lambda\|x\|^2$

$\text{sgn}(\lambda) = \text{sgn}(x^TAx)$

20.

a) All n pivots $a^2 - b < 0$

b)

P is symmetric, $ax^2 + 2bxy + cy^2 > 0$

$P = Q$ iff $A = \begin{bmatrix}a&b\\b&c\end{bmatrix}$ is positive definite

22.

$$A = \begin{bmatrix}5&4\\4&5\end{bmatrix}$$

$\lambda = 1, 9$

$x_1 = (1, -1)$

$x_2 = (1, 1)$

$q_1 = (\frac{1}{\sqrt2}, -\frac{1}{\sqrt2})$

$q_2 = (\frac{1}{\sqrt2}, \frac{1}{\sqrt2})$

$A = Q\Lambda Q^T$

$$= \begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}\begin{bmatrix}1&0\\0&9\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}$$

$$A^{\frac12} = \begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}\begin{bmatrix}1&0\\0&3\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}$$

25.

$A = C^TC$

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

$$= \begin{bmatrix}10&2\\2&4\end{bmatrix}$$

$A = LDL^T$

$$\begin{bmatrix}1&0\\2&1\end{bmatrix}\begin{bmatrix}4&0\\0&9\end{bmatrix}\begin{bmatrix}1&2\\0&1\end{bmatrix}$$

$$= \left(\begin{bmatrix}1&0\\2&1\end{bmatrix}\begin{bmatrix}2&0\\0&3\end{bmatrix}\right)\left(\begin{bmatrix}2&0\\0&3\end{bmatrix}\begin{bmatrix}1&2\\0&1\end{bmatrix}\right)$$

$$C = \begin{bmatrix}2&0\\0&3\end{bmatrix}\begin{bmatrix}1&2\\0&1\end{bmatrix}$$

$$= \begin{bmatrix}2&4\\0&3\end{bmatrix}$$

28

$A = Q\Lambda Q^T$

a) 10 $\quad$ ( )$\begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$

b) 2, 5

a) All eigenvalues > 0

19

$x = c_1x_1 + c_2x_2 + \dots + c_nx_n$

$x^TAx = (c_1x_1 + c_2x_2 + \dots + c_nx_n)A(c_1x_1 + c_2x_2 + \dots + c_nx_n)$

$(c_1x_1 + c_2x_2 + \dots + c_nx_n)(c_1\lambda_1x_1 + c_2\lambda_2x_2 + c_n\lambda_nx_n)$

$= c_1^2\lambda_1x_1^Tx_1 + c_1c_2\lambda_2x_1^T\cdot x_2 + \dots$

Since A is symmetric, all eigenvectors are orthogonal

$x^TAx = \sum_1^n c_i^2\lambda_ix_i^Tx_i > 0$

23.

$\frac{1}{a^2} = \lambda_1 \Rightarrow a = \frac{1}{\sqrt{\lambda_1}},\ b = \frac{1}{\sqrt{\lambda_2}}$

$a = \frac13,\ b = \frac12$

24.

$$A = \begin{bmatrix}1&\frac12\\\frac12&1\end{bmatrix}$$

$\lambda = \frac12, \frac32$

$a = \frac{1}{\sqrt{\lambda_1}} = \frac{1}{\sqrt{\frac12}} = \sqrt2$

$b = \frac{1}{\sqrt{\lambda_2}} = \frac{1}{\sqrt{\frac32}} = \frac{\sqrt6}{3}$

26

$$A = \begin{bmatrix}9&0&0\\0&1&0\\0&2&1\end{bmatrix}\begin{bmatrix}9&0&0\\0&1&0\\0&0&4\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&2\\0&0&1\end{bmatrix}$$

$$C = \sqrt{D}L = \begin{bmatrix}3&0&0\\0&1&0\\0&0&2\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&2\\0&0&1\end{bmatrix}$$

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

Chapter 6.6: Similar Matrices

- Matrices with n unique eigenvalues can be diagonalized as $S\Lambda S^{-1}$. However, we can't do this for all matrices. Some matrices have too few eigenvectors.
- In this section, the eigenvector matrix S remains the choice when it is available, but we now allow any invertible matrix M.
- Starting from A we go to $M^{-1}AM$. No matter which matrix M we choose, the eigenvalues remain the same. The matrices A and $M^{-1}AM$ are called similar.
- A typical matrix A is similar to a whole family of other matrices because M can be any invertible matrix.
- Let M be any invertible matrix. Then $B = M^{-1}AM$ is similar to A.
  - If $B = M^{-1}AM$ then immediately $A = MBM^{-1}$. If B is similar to A, A is similar to B.
- A diagonalizable matrix is similar to Λ. In that special case M is S.
- The combination $M^{-1}AM$ appears when we change variables in a differential equations. Start with an equation for u and set $u = Mv$

$$\frac{du}{dt} = Au \text{ becomes } M\frac{dv}{dt} = AMv \text{ which is } \frac{dv}{dt} = M^{-1}AMv$$

- When $M = S$ the system is diagonal - the maximal simplicity. Other choices of M could make a system triangular and easier to solve. Since we can always go back to u, similar matrices must give the same growth and decay. More precisely, the eigenvalues of A and B are the same
- Similar matrices A and $M^{-1}AM$ have the same eigenvalues. If x is an eigenvector of A, then $M^{-1}x$ is an eigenvector of $M^{-1}AM$.
  - Proof: Since $B = M^{-1}AM$ gives $A = MBM^{-1}$. Suppose $Ax = \lambda x$

$$MBM^{-1}x = \lambda x \text{ means } B(M^{-1}x) = \lambda(M^{-1}x)$$

The eigenvalue of B is the same λ. The eigenvector has changed to $M^{-1}x$
  - Two matrices can have the same repeated λ, and fail to be similar
- Example

The projection $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ is similar to $\Lambda = S^{-1}AS = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

Now choose $M = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$. The similar matrix $M^{-1}AM$ is $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

Also choose $M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. The similar matrix $M^{-1}AM$ is $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$

- Example 2 (repeated eigenvalues)

  $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is similar to $\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ and all $B = \begin{bmatrix} cd & d^2 \\ -c^2 & -cd \end{bmatrix}$ except $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

  - The above matrices are all of determinant 0 and rank 1.
  - These matrices can't be diagonalized. $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is the closest we can get. It is the "Jordan normal form" for the family of matrices B.

- What M changes
  - Not changed by M
    - Eigenvalues
    - Trace and determinant
    - Rank
    - Number of independent eigenvectors
    - Jordan form
  - Changed by M
    - Eigenvectors
    - Nullspace
    - Column Space
    - Row Space
    - Left Nullspace

    $\}$ All 4 Fundamental Subspaces
    - Singular values

Examples of the Jordan Form

- Example

  $J = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$ has $\lambda = 5,5,5$ and $x = (1,0,0)$. $AM = 3$, $GM = 1$

  $J - 5I = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ has rank 2

  - Every similar matrix $B = M^{-1}JM$ has $\lambda = 5,5,5$. Also $B - 5I$ must have the same rank 2. Its nullspace has dimension 1. So every B that is similar to this "Jordan block" J has only one independent eigenvector $M^{-1}x$.
  - The transpose matrix J^T has the same eigenvalues 5, 5, 5 and $J^T - 5I$ has the same rank 2. Jordan's theorem says J^T is similar J. The matrix M that produces this similarity happens to be the reverse identity.

    $J^T = M^{-1}JM$ is $\begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$
  - The eigenvector of J^T is $M^{-1}(1,0,0) = (0,0,1)$
  - This matrix J is similar to every matrix A with eigenvalues 5,5,5 and one line of eigenvectors.

- Example
  - Since J is not diagonal, the equation $du/dt = Ju$ cannot be simplified by changing variables.

    $\frac{du}{dt} = Ju = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is $\begin{matrix} dx/dt = 5x + y \\ dy/dt = 5y + z \\ dz/dt = 5z \end{matrix}$

- We solve this by back substitution
  $\frac{dz}{dt} = 5z$ yields $z = z(0)e^{5t}$

  $\frac{dy}{dt} = 5y + z$ yields $y = (y(0) + tz(0))e^{5t}$

  $\frac{dx}{dt} = 5x + y$ yields $x = (x(0) + ty(0) + \frac{1}{2}t^2 z(0))e^{5t}$
- The missing eigenvectors are responsible for the te^{5t} and t^2e^{5t} in y and x.

The Jordan Form

- For every A, we want to choose M so that $M^{-1}AM$ is as nearly diagonal as possible.
- When A has a full set of n eigenvectors, they go into the columns of $M = S$. Then the matrix $S^{-1}AS$ is diagonal, period.
  - The matrix Λ is the Jordan Form of A - when A can be diagonalized.
- In general, suppose A has s independent eigenvectors. Then it is similar to a matrix with s blocks. Each "Jordan block" has an eigenvalue on the diagonal and 1's above it. This block accounts for one eigenvector of A. When there are n eigenvectors and n blocks, each block is 1 by 1 and J is Λ.
- Jordan Form
  - If A has s independent eigenvectors, it is similar to a matrix J that has s Jordan blocks on its diagonal. Some matrix M puts A into Jordan form.

$$M^{-1}AM = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix} = J$$

  - Each block in J has one eigenvalue λ_i, one eigenvector, and 1's above the diagonal.

$$J_i = \begin{bmatrix} \lambda_i & 1 & \\ & \ddots & 1 \\ & & \lambda_i \end{bmatrix}$$

  - The Jordan block has one off-diagonal 1 for each missing eigenvector.
  - In each family of similar matrices, we have a Jordan Form which is diagonal or nearly so. For that J, we can solve $du/dt = Ju$, we can take powers J^k. Every other matrix in the family has the form $A = MJM^{-1}$
  - A is similar to B iff they have the same Jordan Form.
  - The proof for Jordan's theorem is very intricate and computations with the Jordan Form is not stable.
  - Example
    If A has $\lambda = 4, 2, 3, 3$, its Jordan Form is

$$J = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Problem Set 6.6

1.

$C = F^{-1}AF$

$B = GCG^{-1}$

$B = GF^{-1}AFG^{-1}$

$= GF^{-1}A(GF^{-1})^{-1}$

$M = GF^{-1}$

3.

$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\left(\frac{1}{ad-bc}\right)$

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

$= \frac{1}{ad-bc}\begin{bmatrix} ad+bd & -ab-b^2 \\ cd+d^2 & -bc-bd \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

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

7.

a)

$Ax = 0$

$(M^{-1}AM)(M^{-1}x) = 0$

$M^{-1}Ax = 0$

$Ax = 0$

b) dimension

9.

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

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

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

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

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

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

10.

$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$

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

$J^3 = \begin{bmatrix} \lambda^2 & 2\lambda \\ 0 & \lambda^2 \end{bmatrix}\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} \lambda^3 & 3\lambda^2 \\ 0 & \lambda^3 \end{bmatrix}$

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

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

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

11.

$\frac{du}{dt} = Ju = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} u$

$y = y(0)e^{\lambda t} = 5e^{\lambda t}$

$x = (x(0) + ty(0))e^{\lambda t} = (5+2t)e^{\lambda t}$

$u = \begin{bmatrix} (5+2t)e^{\lambda t} \\ 5e^{\lambda t} \end{bmatrix} \Rightarrow \frac{du}{dt} = \begin{bmatrix} 2e^{\lambda t} + \lambda e^{\lambda t}(5+2t) \\ 5\lambda e^{\lambda t} \end{bmatrix}$

13.

$J_1 = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix} \quad J_2 = \begin{bmatrix} 0 & & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix} \quad J_3 = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$

$J_4 = \begin{bmatrix} 0 & 1 & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} \quad J_5 = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & \\ & & & 0 \end{bmatrix}$

15.

$\det(M^{-1}AM - \lambda M^{-1}M)$

$= \det[M^{-1}(AM - \lambda M)]$

$= \det[M^{-1}M(A - \lambda I)]$

$= \det(A - \lambda I)$

d)

$\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ has $\lambda = 3, 3$

20.

a) $A = M^{-1}BM$

$A^2 = M^{-1}BMM^{-1}BM$

$= M^{-1}B^2M$

c) $\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$ has $\lambda = 3, 4$

$\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$ has $\lambda = 3, 4$

$J = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$

Chapter 6.7: Singular Value Decomposition

- A is any m by n matrix. Its rank is r. We diagonalize this matrix A, but not by $S^{-1}AS$.
- The eigenvectors in S have three big problems
  1. They are usually not orthogonal
  2. There may be insufficient eigenvectors
  3. $Ax = \lambda x$ requires A to be square
- Instead, we use the singular vectors of A.
- The price we pay is to have two sets of singular vectors, u's and v's. The u's are eigenvectors of AA^T and the v's are the eigenvectors of A^TA. Since those matrices are both symmetric, their eigenvectors can be chosen orthonormal.
- The fact that $A(A^TA)$ is the same as $(AA^T)A$ will lead to a remarkable property of u and v

"A is diagonalized" $Av_1 = \sigma_1 u_1 \quad Av_2 = \sigma_2 u_2 \quad \ldots, \quad Av_r = \sigma_r u_r$

- The singular vectors $v_1, v_2, v_3, \ldots$ are in the row space of of A. The outputs $u_1, \ldots, u_r$ are in the column space of A. The singular values $\sigma_1, \sigma_2, \ldots$ are positive scalars. When the v's and u's go into the columns of U and V respectively, orthogonality dictates $V^TV = I$ and $U^TU = I$. The σ's go into a diagonal matrix Σ_1.
- Just as $Ax_i = \lambda_i x_i$ led to the diagonalization $AS = S\Lambda$, the equations $Av_i = \sigma_i u_i$ tell us column by column that $AV = U\Sigma_1$

(m by n)(n by r) equals (m by r)(r by r)
$$A\begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

- The v's and u's account for the row and column space of A. We need $n-r$ more v's and $m-r$ more u's for the nullspace and left nullspace. They can be orthonormal bases for those two nullspaces (and then automatically orthogonal to the first r v's and u's.
- Include all v's and all u's in V and U so these matrices become square. We still have $AV = U\Sigma_1$

(m by n)(n by n) equals (m by m)(m by n)
$$A\begin{bmatrix} v_1 & \cdots & v_r & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r & \cdots & u_m \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

- The new Σ_1 is m by n. It is just the old r by r matrix (Σ_r) with $m-r$ new zero rows and $n-r$ new zero columns. The real change is in the shapes U and V and Σ_1. Still $V^TV = I$ and $U^TU = I$, with sizes n and m.
- V is now a square orthogonal matrix, with inverse $V^{-1} = V^T$. So $AV = U\Sigma_1$ can become $A = U\Sigma_1 V^T$. This is the Singular Value Decomposition

SVD $\quad A = U\Sigma_1 V^T = u_1\sigma_1 v_1^T + \ldots + u_r\sigma_r v_r^T = \sigma_1 u_1 v_1 + \ldots + \sigma_r u_r v_r$

- $A = U_r \Sigma_r V_r^T$ is equally true and is the "reduced SVD". It gives the same splitting of A as a sum of rank 1 matrices.
- We will see that $\sigma_i^2 = \lambda_i$ is an eigenvalue of A^TA and AA^T. When we put the singular values in descending order $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$, it gives the r rank one pieces of A in order of importance
- Example 1
  - When is $U\Sigma V^T$ (singular values) the same as $S\Lambda S^{-1}$ (eigenvalues)
    We need orthonormal eigenvectors in $S = U$. We need non negative eigenvalues in $\Lambda = \Sigma$, so A must be a positive semidefinite (or definite) symmetric matrix $Q\Lambda Q^T$
- Example 2
  - If $A = xy^T$ with unit vectors x and y, what is the SVD of A?
    The reduced SVD is exactly $A = xy^T$ with rank $r = 1$. It has $u_1 = x$, $v_1 = y$ and $\sigma_1 = 1$. For the full SVD, complete $u_1 = x$ to an orthonormal basis of u's and $v_1 = y$ to an orthonormal basis of v

Image Compression

- An image (specifically a raster image) can be described as a matrix of colors. Each color is a vector (RGB)
- Image compression is very useful. Here is a SVD approach
  - Say we had a 256 by 512 image.
  - We replace the 256 by 512 image by a matrix of rank 1. That is just a column times a row, which is only 256 + 512 pixels.
    This is a $(256)(512)/(256+512) \sim 170:1$ compression ratio.
  - This is our best case. Instead we could use a linear combination of 5 rank one matrices and still get a $\sim 34:1$ compression ratio.
  - Where does SVD come in? The best rank one approximation is given by $\sigma_1 u_1 v_1^T$, where σ_1 is the largest singular value. The best rank 5 approximation is $\sigma_1 u_1 v_1^T + \ldots + \sigma_5 u_5 v_5^T$, where the σ's are in descending order.
- A library compresses a different matrix, where rows correspond to key words and columns to titles in the library. The entry in this word-title matrix is $a_{ij} = 1$ if word i is in title j (otherwise it is zero). We normalize columns so long titles don't get an advantage
  - Once this indexing matrix is created, it has to be compressed, perhaps with SVD.

The Bases and the SVD

- Start with a 2 by 2 matrix. Let its rank be $r = 2$, so A is invertible. We want v_1 and v_2 to be perpendicular unit vectors. We also want $Av_1 = \sigma_1 u_1$ and $Av_2 = \sigma_2 u_2$ to be perpendicular. Then the corresponding unit vectors will be orthonormal

• Example

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

- No orthogonal matrix Q will make $Q^{-1}AQ$ diagonal. We need $\Sigma_1 = U^{-1}AV$.
  The two bases will be different. The singular values are the lengths $\|Av_1\|$ and $\|Av_2\|$

$AV = U\Sigma$ $\quad A[v_1 \; v_2] = [\sigma_1 u_1 \; \sigma_2 u_2] = [u_1 \; u_2]\begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix}$

$A = U\Sigma_1 V^T$

- To get V by itself, take $A^TA = (U\Sigma_1 V^T)^T(U\Sigma_1 V^T) = V\Sigma_1^T\Sigma_1 V^T$
  Multiplying those diagonal Σ_1^T and Σ_1 gives σ_1^2 and σ_2^2. Then
  Eigenvalues: σ_1^2, σ_2^2 $\quad A^TA = V\begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix}V^T$
  Eigenvectors: v_1, v_2
- This is exactly like $A = Q\Lambda Q^T$
  - The symmetric matrix is A^TA
  - The columns of V are the eigenvectors of A^TA

• Likewise, the columns of U are the eigenvectors of AA^T. We can also compute them by:
$Av_i = \sigma_i u_i \Rightarrow u_i = Av_i / \sigma_i$

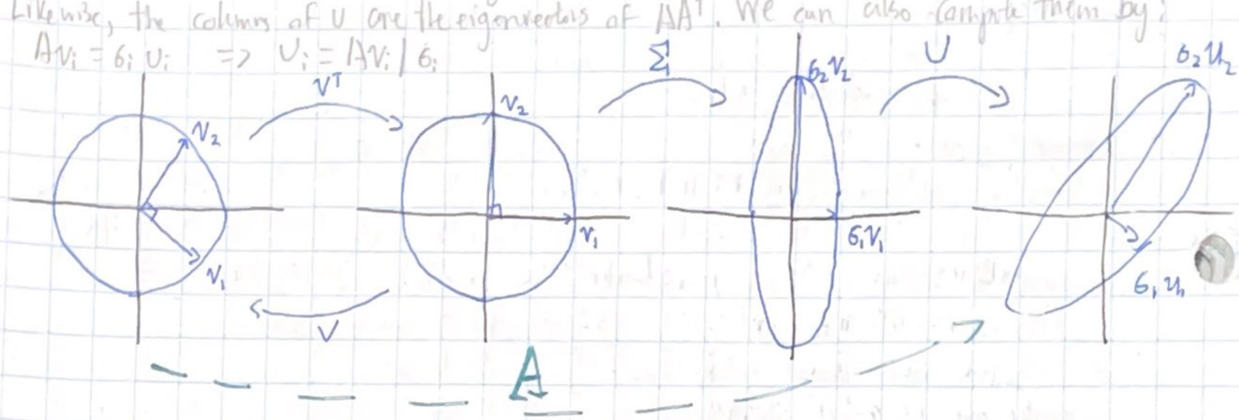


fig 6.7.1: U and V are rotations and reflections, Σ_1 stretches the circle tion ellipse.

• Example
- Find the SVD of $A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$
- Compute A^TA and its orthonormal eigenvectors
  $A^TA = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ has unit eigenvectors $v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$ and $v_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$
- The eigenvalues are 8 and 2. The v's are orthogonal because A^TA is automatically symmetric and symmetric matrices have perpendicular eigenvectors.
- We find u_i, we normalize Av_i

$Av_1 = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} \Rightarrow u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ The singular value is $2\sqrt{2}$. Note $\sigma_1^2 = (2\sqrt{2})^2 = 8 = \lambda_1$

$Av_2 = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix} \Rightarrow u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ The singular value is $\sqrt{2}$ Note $\sigma_2^2 = (\sqrt{2})^2 = 2 = \lambda_2$

- Then,

$A = U\Sigma_1 V^T$ is $\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$

- We could also have found the u's as the eigenvectors of AA^T

  Use $V^TV=I$ $\quad AA^T=(U\Sigma V^T)(V\Sigma^T U^T) = U\Sigma\Sigma^T U^T$

  $AA^T=\begin{bmatrix}8&0\\0&2\end{bmatrix} \Rightarrow u_1=(1,0),\ u_2=(0,1),\ \sigma_1^2=8,\ \sigma_2^2=2$

- Example
  - Find the SVD of $A=\begin{bmatrix}2&2\\1&1\end{bmatrix}$
  - The row space has only one basis vector $v_1=\frac{1}{\sqrt2}(1,1)$. Same with the column space; $u_1=\frac{1}{\sqrt5}(2,1)$. Then $Av_1=(4,2)/\sqrt2=\sigma_1 u_1$. It does, with $\sigma_1=\sqrt{10}$.

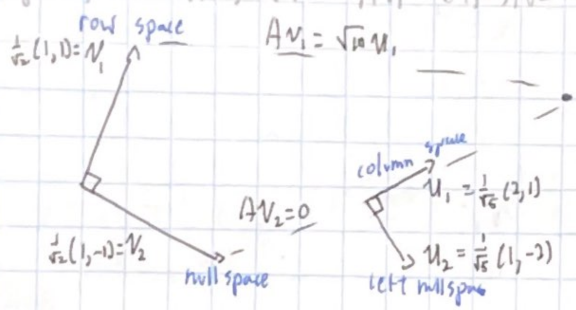


Fig 6.7.2: The SVD chooses orthonormal bases for 4 subspaces so that $Av_i=\sigma_i u_i$

  - The SVD could stop after the column and row space but it is customary for U and V to be square.
  - The vector v_2 is in the null space. It is perpendicular to v_1 in the row space, so we can chose $\frac{1}{\sqrt2}(1,-1)$. Likewise for u_2, which we can choose as $\frac{1}{\sqrt5}(1,-2)$ to make it perpendicular to u_1. We set σ to zero to get

    $A=U\Sigma V^T$ Full size $\quad \begin{bmatrix}2&2\\1&1\end{bmatrix}=\frac{1}{\sqrt5}\begin{bmatrix}2&1\\1&-2\end{bmatrix}\begin{bmatrix}\sqrt{10}&0\\0&0\end{bmatrix}\frac{1}{\sqrt2}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$

- The matrices U and V contain orthonormal bases for all 4 fundamental subspaces
  - first r columns of V: row space of A
  - last $n-r$ columns of V: nullspace of A
  - first r columns of U: column space of A
  - last $m-r$ columns of U: left nullspace of A
- As long as the last v's and u's in the nullspaces are orthonormal, the SVD will be correct.
- Proof of the SVD
  - Start from $A^TAv_i=\sigma_i^2 v_i$, which gives the v's and σ's. Multiplying by v_i^T leads to $v_i^TA^TAv_i=v_i^T\sigma_i^2v_i \Rightarrow (Av_i)^T(Av_i)=\sigma_i^2 v_i^Tv_i \Rightarrow \|Av_i\|^2=\sigma_i^2\|v_i\|^2=1$

    $\|Av_i\|^2=\sigma_i^2$ so $\|Av_i\|=\sigma_i$

    Multiply by A

    $AA^TAv_i=\sigma_i^2Av_i$ gives $(AA^T)(Av_i)=\sigma_i^2(Av_i) \Rightarrow Av_i$ as an eigenvector of AA^T.

    Then $u_i=Av_i/\sigma_i$ as a unit eigenvector of AA^T

Problem Set 6.7

1.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix}$$

$$\det(A^TA - \lambda I) = (10-\lambda)(40-\lambda) - (20)(20)$$
$$= \lambda^2 - 50\lambda + 400 - 400$$
$$= \lambda^2 - 50\lambda$$
$$= \lambda(\lambda - 50)$$

Set $\det(A^TA - \lambda I) = 0$

$\lambda = 0, 50$

case $\lambda = 0$

$(A^TA - 0I)x_1 = 0$

$$\begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix} x_1 = 0$$

$x_1 = (2, -1)$

$v_1 = \frac{1}{\sqrt{5}}(2, -1)$

case $\lambda = 50$:

$(A^TA - 50I)x_2 = 0$

$$\begin{bmatrix} -40 & 20 \\ 20 & -10 \end{bmatrix} x_2 = 0$$

$x_2 = (1, 2)$

$v_2 = \frac{1}{\sqrt{5}}(1, 2)$

$$Av_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \sigma_1 = 0,\ u_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Av_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 5 \\ 15 \end{bmatrix} = \sqrt{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$u_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \sigma_2 = \sqrt{50}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{10}}{10} & \boxed{\frac{3\sqrt{10}}{10}} \\ \frac{3\sqrt{10}}{10} & \boxed{-\frac{\sqrt{10}}{10}} \end{bmatrix}\begin{bmatrix} \sqrt{50} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \boxed{\frac{2\sqrt{5}}{5} \quad -\frac{\sqrt{5}}{5}} \end{bmatrix}$$

choose to be orthogonal to u_1 / v_1

2.

$C(A^T) = c\left(\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5}\right)$

$C(A) = c\left(\frac{\sqrt{10}}{10}, \frac{3\sqrt{10}}{10}\right)$

$N(A) = c\left(\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5}\right)$

$N(A^T) = c\left(\frac{3\sqrt{10}}{10}, -\frac{\sqrt{10}}{10}\right)$

7.

Largest $\sigma_i = \sqrt{3}$

$$u_1 \sigma_1 v_1^T = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)(\sqrt{3})\left[\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6}\right]$$

$$= \sqrt{3}\begin{bmatrix} \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{6} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix}$$

4.

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

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

$$\det(A^TA - \lambda I) = (2-\lambda)(1-\lambda) - (1)(1)$$
$$= \lambda^2 - 3\lambda + 2 - 1$$
$$= \lambda^2 - 3\lambda + 1$$

$\lambda_1 = \frac{3-\sqrt{5}}{2}, \lambda_2 = \frac{3+\sqrt{5}}{2}$

case $\lambda_1 = \frac{3-\sqrt{5}}{2}$

$$\begin{bmatrix} 2-\lambda_1 & 1 \\ 1 & 1-\lambda_1 \end{bmatrix} x_1 = 0$$

$x_1 = \left(\frac{1-\sqrt{5}}{2}, 1\right)$

$a_1 =$

6.

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

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

$$\det(A^TA - \lambda I) = (1-\lambda)(2-\lambda)(1-\lambda) - (1)(1)(1-\lambda) - (1-\lambda)(1)(1)$$
$$= (1-\lambda)^2(2-\lambda) - 2(1-\lambda)$$
$$= (\lambda^2 - 2\lambda + 1)(2-\lambda) - 2 + 2\lambda$$
$$= 2\lambda^2 - 4\lambda + 2 - \lambda^3 + 2\lambda^2 - \lambda - 2 + 2\lambda$$
$$= -\lambda^3 + 4\lambda^2 - 3\lambda$$
$$= -\lambda(\lambda^2 - 4\lambda + 3)$$
$$= -\lambda(\lambda - 3)(\lambda - 1)$$

Set $\det(A^TA - \lambda I) = 0$

$\lambda = 0, 1, 3$

case $\lambda = 0$:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} x_1 = 0$$

$x_1 = (1, -1, 1)$

$v_1 = \left(\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right)$

case $\lambda = 1$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} x_2 = 0$$

$x_2 = (1, 0, -1)$

$v_2 = \left(\frac{\sqrt{2}}{2}, 0, -\frac{\sqrt{2}}{2}\right)$

case $\lambda = 3$

$$\begin{bmatrix} -2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -2 \end{bmatrix} x_3 = 0$$

$x_3 = (1, 2, 1)$

$v_3 = \left(\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6}\right)$

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

$\det(AA^T-\lambda I) = (2-\lambda)(2-\lambda)-(1)(1)$
$= \lambda^2-4\lambda+4-1$
$= \lambda^2-4\lambda+3$
$= (\lambda-3)(\lambda-1)$

Set $\det(AA^T-\lambda I)=0$
$\lambda = 1, 3$

Case $\lambda=1$: $\begin{bmatrix}1&1\\1&1\end{bmatrix}x_1=0$ Case $\lambda=3$: $\begin{bmatrix}-1&1\\1&-1\end{bmatrix}x_2=0$

$x_1=(1,-1)$ $x_2=(1,1)$
$u_2=(\frac{\sqrt2}{2},-\frac{\sqrt2}{2})$ $u_1=(\frac{\sqrt2}{2},\frac{\sqrt2}{2})$

$Av_3 = \sigma_1 u_1 \Rightarrow \begin{bmatrix}1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}\frac{\sqrt6}{6}\\ \frac{\sqrt6}{3}\\ \frac{\sqrt6}{6}\end{bmatrix} = \begin{bmatrix}\frac{\sqrt6}{2}\\ \frac{\sqrt6}{2}\end{bmatrix} = \sqrt3\begin{bmatrix}\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2}\end{bmatrix}$

$\sigma_1=\sqrt3$

$Av_2 = \sigma_2 u_2 \Rightarrow \begin{bmatrix}1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}\frac{\sqrt2}{2}\\0\\-\frac{\sqrt2}{2}\end{bmatrix} = \begin{bmatrix}\frac{\sqrt2}{2}\\-\frac{\sqrt2}{2}\end{bmatrix}$

$\sigma_2=1$

$A=U\Sigma V^T \Rightarrow \begin{bmatrix}1&1&0\\0&1&1\end{bmatrix} = \begin{bmatrix}\frac{\sqrt2}{2}&\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2}&-\frac{\sqrt2}{2}\end{bmatrix}\begin{bmatrix}\sqrt3&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt6}{6}&\frac{\sqrt6}{3}&\frac{\sqrt6}{6}\\ \frac{\sqrt2}{2}&0&-\frac{\sqrt2}{2}\\ \frac{\sqrt3}{3}&-\frac{\sqrt3}{3}&\frac{\sqrt3}{3}\end{bmatrix}$

SVD of A:
$A=\begin{bmatrix}-1&3\\2&-6\end{bmatrix} = U\Sigma V^T = \begin{bmatrix}\frac{\sqrt5}{5}&\frac{2\sqrt5}{5}\\ \frac{2\sqrt5}{5}&\frac{\sqrt5}{5}\end{bmatrix}\begin{bmatrix}5\sqrt2&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{10}}{10}&-\frac{3\sqrt{10}}{10}\\ \frac{3\sqrt{10}}{10}&\frac{\sqrt{10}}{10}\end{bmatrix}$

$A^TA = \begin{bmatrix}-1&2\\3&-6\end{bmatrix}\begin{bmatrix}-1&3\\2&-6\end{bmatrix} = \begin{bmatrix}5&-15\\-15&45\end{bmatrix}$

$\det(A^TA-\lambda I) = (5-\lambda)(45-\lambda)-(-15)(-15)$
$=\lambda^2-50\lambda+225-225$
$=\lambda^2-50\lambda$
$=\lambda(\lambda-50)$

set $\det(A^TA-\lambda I)=0 \Rightarrow \lambda=0, 50$

case $\lambda=50$:
$(A^TA-50I)x_1=0$
$\begin{bmatrix}-45&-15\\-15&-5\end{bmatrix}x_1=0$
$x_1=(1,-3)$
$v_1=(\frac{\sqrt{10}}{10},-\frac{3\sqrt{10}}{10})$

case $\lambda=0$
$(A^TA-0I)x_2=0$
$\begin{bmatrix}5&-15\\-15&45\end{bmatrix}x_2=0$ $v_1^Tv_2=0$
$x_2=(3,1)$
$v_2=(\frac{3\sqrt{10}}{10},\frac{\sqrt{10}}{10})$

$Av_1=\sigma_1 u_1$:
$\begin{bmatrix}-1&3\\2&-6\end{bmatrix}\begin{bmatrix}\frac{\sqrt{10}}{10}\\-\frac{3\sqrt{10}}{10}\end{bmatrix} = \begin{bmatrix}-\sqrt{10}\\2\sqrt{10}\end{bmatrix} = -5\sqrt2\begin{bmatrix}-\frac{\sqrt5}{5}\\-\frac{2\sqrt5}{5}\end{bmatrix}$

$\sigma_1=5\sqrt2$, $u_1=(\frac{\sqrt5}{5},\frac{2\sqrt5}{5})$

v_2 is chosen to be orthogonal to v_1
u_2 is chosen to be orthogonal to u_1
$u_2=(-\frac{2\sqrt5}{5},\frac{\sqrt5}{5})$

SVD of A:
$A=\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix}$

Procedure:
Find v_i and then
u from Av_i;
easier to number

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

$\det(A^TA-\lambda I) = (2-\lambda)(2-\lambda)-(1)(1)$
$=\lambda^2-4\lambda+4-1$
$=\lambda^2-4\lambda+3$
$=(\lambda-3)(\lambda-1)$

Set $\det(A^TA-\lambda I)=0 \Rightarrow \lambda=1,3$

Case $\lambda=1$
$(A^TA-I)x_1=0$
$\begin{bmatrix}1&1\\1&1\end{bmatrix}x_1=0$
$x_1=(1,-1)$
$v_1=(\frac{\sqrt2}{2},-\frac{\sqrt2}{2})$

case $\lambda=3$
$(A^TA-3I)x_2=0$
$\begin{bmatrix}-1&1\\1&-1\end{bmatrix}x_2=0$
$x_2=(1,1)$
$v_2=(\frac{\sqrt2}{2},\frac{\sqrt2}{2})$

Find u and σ

$Av_1=\sigma_1 u_1$:
$\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix}\begin{bmatrix}\frac{\sqrt2}{2}\\-\frac{\sqrt2}{2}\end{bmatrix} = \begin{bmatrix}\frac{\sqrt2}{2}\\-\frac{\sqrt2}{2}\\0\end{bmatrix} \Rightarrow \sigma_1=1,\ u_1=\begin{bmatrix}\frac{\sqrt2}{2}\\-\frac{\sqrt2}{2}\\0\end{bmatrix}$

$Av_2=\sigma_2 u_2$:
$\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix}\begin{bmatrix}\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2}\end{bmatrix} = \begin{bmatrix}\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2}\\ \sqrt2\end{bmatrix} \Rightarrow \sigma_2=\sqrt3,\ u_2=\begin{bmatrix}\frac{\sqrt6}{6}\\ \frac{\sqrt6}{6}\\ \frac{\sqrt6}{3}\end{bmatrix}$

Choose u_3 to be orthonormal to u_1 and u_2
$u_3=\frac{1}{\sqrt3}(1,1,-1)=(\frac{\sqrt3}{3},\frac{\sqrt3}{3},-\frac{\sqrt3}{3})$

$A=U\Sigma V^T$:
$\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix} = \begin{bmatrix}\frac{\sqrt2}{2}&\frac{\sqrt6}{6}&\frac{\sqrt3}{3}\\-\frac{\sqrt2}{2}&\frac{\sqrt6}{6}&\frac{\sqrt3}{3}\\0&\frac{\sqrt6}{3}&-\frac{\sqrt3}{3}\end{bmatrix}\begin{bmatrix}1&0\\0&\sqrt3\\0&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt2}{2}&-\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2}&\frac{\sqrt2}{2}\end{bmatrix}$

Table of Eigenvalues and Eigenvectors

Symmetric $A^T = A$	real λ's	orthogonal $x_i^T x_j = 0$
Orthogonal $Q^T = Q^{-1}$	all $\lvert\lambda\rvert = 1$	orthogonal $\bar{x}_i^T x_j = 0$
Skew-Symmetric $A^T = -A$	imaginary λ's	orthogonal $\bar{x}_i^T x_j = 0$
Complex Hermitian $\bar{A}^T = A$	real λ's	orthogonal $\bar{x}_i^T x_j = 0$
Positive Definite $x^T A x > 0 \ \forall x \neq 0$	all $\lambda > 0$	orthogonal since $A^T = A$
Markov $m_{ij} > 0$, $\sum_{i=1}^{n} m_{ij} = 1$	$\lambda_{max} = 1$	Steady state $x > 0$
Similar $B = M^{-1} A M$	$\lambda(B) = \lambda(A)$	$x(B) = M^{-1} x(A)$
Projection $P = P^2 = P^T$	$\lambda = 1; 0$	column space; nullspace
Plane Rotation	$e^{i\theta}$ and $e^{-i\theta}$	$x = (1, i)$ and $(1, -i)$
Reflection $I - 2uu^T$	$\lambda = -1; 1, \ldots, 1$	u; whole plane $u_\perp$
Rank One uv^T	$\lambda = v^T u; 0, \ldots, 0$	u; whole plane $v^\perp$
Inverse A^{-1}	$1/\lambda(A)$	Same eigenvectors
Shift $A + cI$	$\lambda(A) + c$	Same eigenvectors
Stable powers $A^n \to 0$	all $\lvert\lambda\rvert < 1$	any eigenvectors
Stable exponential $e^{At} \to 0$	all $Re(\lambda) < 0$	any eigenvectors
Cyclic Permutation row 1 of I last	$\lambda_k = e^{2\pi i k/n}$	$x_k = (1, \lambda_k, \ldots, \lambda_k^{n-1})$
Tridiagonal $-1, 2, -1$ on diagonals	$\lambda_k = 2 - 2\cos\frac{k\pi}{n+1}$	$x_k = (\sin\frac{k\pi}{n+1}, \sin\frac{2k\pi}{n+1}, \ldots)$
Diagonalizable $A = S\Lambda S^{-1}$	diagonal of Λ	columns of S are independent
Symmetric $A = Q\Lambda Q^T$	diagonal of Λ (real)	columns of Q are orthonormal
Schur $A = QTQ^{-1}$	diagonal of T	columns of Q if $A^T A = AA^T$
Jordan $J = M^{-1} A M$	diagonal of J	each block gives $x = (0, \ldots, 1, \ldots, 0)$
Rectangular $A = U\Sigma V^T$	rank(A) = rank(Σ)	eigenvectors of A^TA, AA^T in V, U

Chapter 7: Linear Transformations

Chapter 7.1: The Idea of a Linear Transformation

- When the matrix A multiplies a vector v, it "transforms" into another vector Av. In goes v, out comes Av. $T(v) = Av$
  - A transformation is very much like a function. To find a specific output, we simply evaluate Av.
  - The deeper goal is to see all v's at once. We are transforming the entire space V when we multiply all vectors v.
- Start with a matrix A. It transforms vector v into Av and vector w into Aw. Then we know what it does to $u = v + w$; it equals $Av + Aw$. Matrix multiplication $T(v) = Av$ gives a linear transformation.
- A transformation T assigns an output $T(v)$ to each input v in V. It is linear if it meets the below requirements for all v and w

a) $T(v + w) = T(v) + T(w)$ b) $T(cv) = cT(v)$ for all c

- We combine these requirements into one.

  $T(cv + dw) = cT(v) + dT(w)$

- A linear transformation is very restrictive. A transformation $T(v) = v + u_0$ adding a constant vector u_0 is not linear:

  $T(v + w) = v + w + u_0 \neq T(v) + T(w)$

- $T(v) = v$ is the identity transformation ($A = I$). The input space V is the same as the output space W.
- The linear-plus-shift transformation $T(v) = Av + u_0$ is called "affine". Computers work with these because images must be able to move.
- Example
  - Choose a fixed vector $a = (1, 3, 4)$ and let $T(v)$ be the dot product $a \cdot v$

    The input is $v = (v_1, v_2, v_3)$ the output is $T(v) = v_1 + 3v_2 + 4v_3$

    This a linear transformation with $\mathbb{R}^3 \to \mathbb{R}^1$
- Example
  - The length transformation $T(v) = \|v\|$ is not linear.
  - Both requirements are false
    a) $T(v + w) \leq T(v) + T(w)$ (inequality, not equality)
    b) $T(cv) \neq cT(v)$ for negative c
- Example
  - T is the transformation that rotates every vector by $30°$.
    Its domain and range are both $\mathbb{R}^2$.
  - T is linear: The sum of rotations is the rotation of a sum.

Lines to Lines, Triangles to Triangles

- fig 7.1.1 shows the line from v to w in the input space and the line from $T(v)$ to $T(w)$ in the output space. Linearity tells us that every point on the input line goes on the output line. More than that, equally spaced points go to equally spaced points on the output.
- The second figure goes up a dimension. Now we have three points v_1, v_2, v_3 mapped to outputs $T(v_1), T(v_2), T(v_3)$. Equally space points along the triangle stay equally spaced. The middle point $u = \frac{1}{3}(v_1 + v_2 + v_3)$ goes to the middle point $T(u) = \frac{1}{3}(T(v_1) + T(v_2) + T(v_3))$.
- Linearity extends to n vectors

  $u = c_1v_1 + c_2v_2 + \dots + c_nv_n$

  Linearity $T(u) = c_1T(v_1) + c_2T(v_2) + \dots c_nT(v_3)$

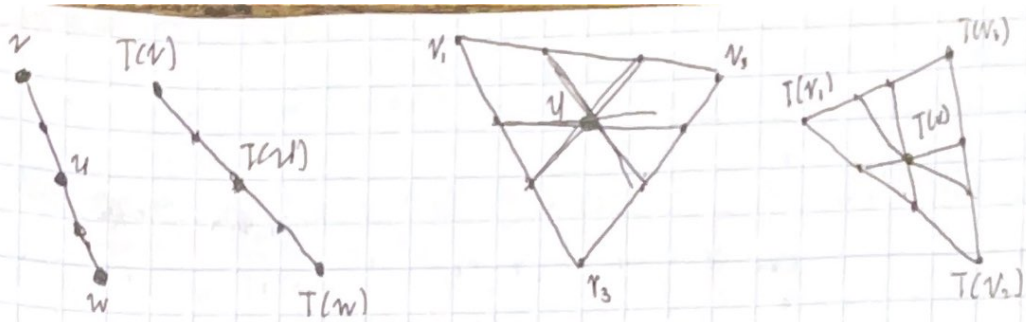


Fig 7.1.1

- Some terms for Linear Transformations
  - Range of T = set of all outputs $T(v)$; corresponds to the column space.
  - Kernel of T: set of all inputs for which $T(v)=0$; corresponds to nullspace.
- The range is in the output space W and the kernel is in the input space V.

Examples of Transformations (mostly linear)

- Example:
  - Project every 3-dimensional vector onto the xy space. Then $T(x,y,z)=(x,y,0)$.
  - The range is that plane. The kernel is the z-axis.
  - This projection is linear
- Example
  - Project every 3 dimensional vectors onto the horizontal plane $z=1$, $T(x,y,z)=(x,y,1)$, is not linear because $T(0)\neq 0$
- Example
  - Suppose A is an invertible matrix. The kernel of T is the zero vector; the range W equals the domain V. The inverse transformation T^{-1} multiplies the input by A^{-1}. $T^{-1}(T(v)) = v$ matches the matrix multiple $A^{-1}(Av)$
- All linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ are produced by matrices

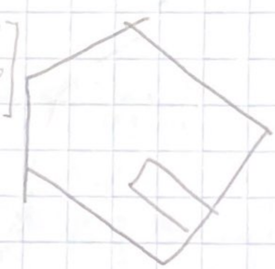
Linear Transformations of the Plane

- House Matrix

$$H=\begin{bmatrix} -6 & -6 & -7 & 0 & 7 & 6 & 6 & -3 & -3 & 0 & 0 & -6 \\ -7 & 2 & 1 & 8 & 1 & 2 & -7 & -7 & -2 & -2 & -7 & -7 \end{bmatrix}$$

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

$$A=\begin{bmatrix} \cos 35^\circ & -\sin 35^\circ \\ \sin 35^\circ & \cos 35^\circ \end{bmatrix}$$



$$A=\begin{bmatrix} .7 & .3 \\ .3 & .7 \end{bmatrix}$$

Problem set 7.1

1.
$T(v+0) = T(v) + T(0)$
$T(v) = T(v) + T(0)$
$T(0) = 0$

$T(0v) = 0T(v)$
$T(0) = 0T(v)$
$T(0) = 0$

4.
a) $S(T(v)) = S(v) = v$
b)
$S(T(v_1) + T(v_2))$
$= S(T(v_1)) + S(T(v_2))$

10
11.
a), c) $\mathbb{R}^n \to \mathbb{R}^m$

3.
a)
$T(v+w) = T(v) + T(w)$
$(v_2 + w_2, v_1 + w_1) = (v_2 + w_2, v_1 + w_1)$
d)
$T(0) \neq 0$, no

7.
a) $T(T(v)) = T(-v) = -(-v) = v$
b) $T(T(v)) = T(v + (1,1)) = v + (2,2)$

9.
$T(T(v)) = T(v_2, v_3, v_1) = (v_3, v_1, v_2)$
$T^{100}(v) = T(v) = (v_2, v_3, v_1)$

13.
$T(M) = AM$
$T(cM) = A(cM) = cAM$
$T(M_1 + M_2) = A(M_1 + M_2) = AM_1 + AM_2$

e)
$T(cv + dw) = cT(v) + dT(w)$
$T(c(v_1 + dw_1, cv_2 + dw_2) = cT(v_1, v_2) + dT(w_1, w_2)$
$(v_1 + dw_1, -cv_2 - dw_2) = c(v_1, -v_2) + d(w_1, -w_2)$

Yes

Chapter 7.2: The Matrix of a Linear Transformation

- The next pages assign a matrix to every linear transformation T. For column vectors, the input v is in $V = \mathbb{R}^n$ and the output $T(v)$ is in $W = \mathbb{R}^m$. The matrix A for the transformation will be m by n. Our choice of basis in V and W will decide A.
- The standard basis vectors for $\mathbb{R}^n$ and $\mathbb{R}^m$ are the columns of I.
- When V and W are not $\mathbb{R}^n$ and $\mathbb{R}^m$, they still have bases. Each choice of basis leads to a matrix for T. When the input basis is different from the output basis, the matrix for $T(v) = v$ will not be I - It will be a "change of basis" matrix.
- Suppose you know $T(v_1), \dots, T(v_n)$ for basis vectors $v_1, \dots, v_n$. Then linearity produces $T(v)$ for every other input vector v.
- Example 1
  - Suppose T transforms $v_1 = (1,0)$ to $T(v_1) = (2,3,4)$ and $v_2 = (0,1)$ goes to $T(v_2) = (5,5,5)$. If T is linear from $\mathbb{R}^2$ to $\mathbb{R}^3$ then its "standard matrix" is 3 by 2. These outputs $T(v_1)$ and $T(v_2)$ go into its columns.

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix} \quad T(v_1 + v_2) = T(v_1) + T(v_2) \quad \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

- Example 2
  - If we have the transformation $T(v) = dv/dt$, it is a linear transformation.
    $T(cv + dw) = c\frac{dv}{dt} + d\frac{dw}{dx}$
  - This linearity and the derivative on the basis vectors $1, x, x^2, x^3$ allow us to find the derivative of polynomials like $5 + 2x - 2x^2 + 2x^3$

- Here the input space contains all linear combinations of $1, x, x^2$, and x^3. These specific functions are the basis for the space of cubic polynomials
- To find the kernel, we solve $\frac{dv}{dx} = 0$, which is $v = C$. This gives us a nullspace of T that is one-dimensional.
- To find the range, we look at the outputs $T(v)$, which are all quadratic polynomials
- The input space has dimension 4 and the output space 3, so the "derivative matrix" will be 3 by 4.
- The range of T is a three dimensional subspace, so the matrix will have rank 3. The kernel is of dimension one, so $3+1=4$ is the dimension of the input space

- Example 3
  - The integral is the inverse of the derivative, so the transformation T^{-1} represents this
    $$\int_0^x dt = x, \quad \int_0^x t\,dt = \frac{1}{2}x^2, \quad \int_0^x t^2\,dt = \frac{1}{3}x^3$$
  - By linearity, the integral of $B + Cx + Dx^2$ is $T^{-1}(w) = Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$. Input space = quadratics, Output space = cubics
  - Integration takes W back to V. Its matrix will be 4 by 3
  - Range of T^{-1}: The outputs $Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$ are cubics with no constant term
  - Kernel of T^{-1}: The output is 0 iff $B=C=D=0$. The nullspace is $Z = \{0\}$
  - Fundamental theorem: $3+0$ is the dimension of the input space W for T^{-1}.

Matrices for the Derivative and Integral

- The matrix form of derivative T:
  $$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$
  - Why is A the correct matrix? Because multiplying by A agrees with transforming by T. The derivative of $v = a + bx + cx^2 + dx^3$ is $T(v) = b + 2cx + 3dx^3$
    $$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}$$
- Similarly, the matrix form of the integral T^{-1}
  $$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ B \\ \frac{1}{2}C \\ \frac{1}{3}D \end{bmatrix}$$
  - We'll call this matrix A^{-1}, although you'll notice A is rectangular. So this integral matrix is a one-sided inverse of the derivative matrix.
  - Integrating then differentiating gives us $AA^{-1} = I$. The opposite direction would lose the constant term before integrating. The integral of the derivative of 1 is 0

Construction of the Matrix

- Suppose T transforms n-dimensional space V to m-dimensional space W. We choose a basis $v_1, \dots, v_n$ for V and $w_1, \dots, w_m$ for W. The matrix A will be m by n. To find the first column of A, apply T to the first basis vector v_1: $T(v_1)$. $T(v_1)$ will be in W.

  $T(v_1)$ is a combination $a_{11}w_1 + \dots + a_{m1}w_m$ of the output basis for W

  The numbers $a_{11}, \dots, a_{m1}$ go into the first column of A. Transforming v_1 to $T(v_1)$ matches multiplying $(1, 0, \dots, 0)$ by A. It yields that first column of the matrix.
- When T is the derivative and the first basis vector is 1, its derivative $T(v_1) = 0$, so the first column of the derivative matrix is all zeros.
- For the integral, the first basis function is again 1. Its integral is the second basis function x. So the first column of A^{-1} was $(0, 1, 0, 0)$
- The jth column of A is found by applying T to the jth basis vector v_j

  $T(v_j)$ = combination of basis vectors of $W = a_{1j}w_1 + \dots a_{mj}w_m$

  - These numbers $a_{1j}, \dots, a_{mj}$ go into column j of A. The matrix is constructed to get the basis vectors right.
- Example 4
  - If the bases change, T is the same but the matrix A is different. Suppose we reorder the basis to $x, x^2, x^3, 1$ for the cubics in V. Keep the original basis $1, x, x^2$ for the quadratics in W. The derivative of the first basis vector $v_1 = x$ is $w_1 = 1$, so we get

$$A_{new} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}$$

= matrix for the derivative T when the bases change to $x, x^2, x^3, 1$ and $1, x, x^2$

  - Reordering the basis vectors of V, we reorder the columns of A.

Products AB match Transformations TS

- Example
  - T rotates every vector by the angle θ. Here $V = W = \mathbb{R}^2$. Find A
    The standard basis for $\mathbb{R}^2$ are $\hat{i} = (1, 0)$ and $\hat{j} = (0, 1)$. Apply T to $\hat{i}$ gets us $(\cos\theta, \sin\theta)$ and doing the same for $\hat{j}$ gets us $(-\sin\theta, \cos\theta)$. This gives A:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = A$$

- Example
  - Suppose T projects every plane vector onto the 45° line. Find its matrix for two different choices of the basis.

- We start with the basis v_1, v_2 which point in the $45°$ and $135°$ directions respectively.
  $T(v_1)$ projects onto itself: $T(v_1) = v_1$, so we get $a_1 = (1, 0)$
  $T(v_2)$ projects to 0: $T(v_2)$, so we get $a_2 = (0, 0)$
  $P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ~~when V and W have basis vectors at 45° and 135°~~
- If we choose the standard basis vectors $(1,0)$ and $(0,1)$, Both vectors project to $(\frac{1}{2}, \frac{1}{2})$. So we get
  $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

- Matrix multiplication is defined as such because it gives the correct output which represents the composition of transformations T and S as the product of their matrices A and B.
- The linear transformation TS starts with any vector u in U, goes to $S(u)$ in V and then to $T(S(u))$ in W. The matrix AB starts with any x in $\mathbb{R}^p$, goes to Bx in $\mathbb{R}^n$ and then to ABx in $\mathbb{R}^m$

  $TS: U \to V \to W$ $\quad\quad$ $AB: (m \text{ by } n)(n \text{ by } p) = (m \text{ by } p)$

- Example
  - S rotates the plane by θ and so does T. Then TS rotates by 2θ.
    We have $S = T$ and $A = B$
    $$T^2 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^2 = \begin{bmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}$$

The Identity Transformation and the Change of Basis Matrix

- When each output $T(v_j) = v_j$ is the same as w_j, the matrix is just I.
- Suppose the basis is different. Then $T(v_1) = v_1$ is a combination of w's. The combinations $m_{11}w_1 + \dots + m_{n1}w_n$ gives us the first column of the matrix M.
- Identity transformation:
  - When the outputs $T(v_j) = v_j$ are combinations $\sum_{i=1}^{n} m_{ij} w_i$, the "change of base" matrix is M
  - The basis is changing but the vectors are not: $T(v) = v$
- Example
  - The input basis is $v_1 = (3,7)$, $v_2 = (2,5)$. The output basis is $w_1 = (1,0)$, $w_2 = (0,1)$.
    Then the matrix M is easy to compute:
    $T(v_1) = (3,7) = 3w_1 + 7w_2$
    $T(v_2) = (2,5) = 2w_1 + 5w_2$ $\Rightarrow M = \begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}$
- Example
  - The input basis is $\{(1,0), (0,1)\}$ and the output basis is $\{(3,7), (2,5)\}$
    $T(v_1) = (1,0) = 5w_1 - 7w_2$
    $T(v_2) = (0,1) = -2w_1 + 3w_2$ $\Rightarrow M = \begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -2 \\ -7 & 3 \end{bmatrix}$
    $[w_1 \; w_2]\begin{bmatrix} 5 & -2 \\ -7 & 3 \end{bmatrix} = [v_1 \; v_2]$ is $MM^{-1} = I$

Wavelet Transform = Change to Wavelet Basis

- Wavelets are little waves. They have different lengths and they are localized at different places. The first basis vector is actually not a wavelet, but the very useful matrix of all 1s. This example show "Haar Wavelets"

$$\text{Haar Basis} \quad w_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad w_2 = \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix} \quad w_3 = \begin{bmatrix}1\\-1\\0\\0\end{bmatrix} \quad w_4 = \begin{bmatrix}0\\0\\1\\-1\end{bmatrix}$$

- These vectors are orthogonal. w_3 is localized in the second half and w_4 is localized in the second half. The wavelet transform finds the coefficients c_1, c_2, c_3, c_4 when the input signal $v = (v_1, v_2, v_3, v_4)$ is expressed in the wavelet basis:

$$\text{Transform } v \text{ to } c \qquad v = c_1 w_1 + c_2 w_2 + c_3 w_3 + c_4 w_4 = Wc$$

- The coefficients c_3 and c_4 tell us about details in the first and second half of v. The coefficient c_1 is the average.
- Why do we want to change the basis? Think of v_1, v_2, v_3, and v_4 as the intensities of a signal. We want to compress this signal but if we keep only 5% of the standard basis coefficients, we lose 95% of the signal. But if we choose a better basis of w's, 5% of the basis vectors can combine to form a signal very close to the original.
  - One good basis vector would be $(1,1,1,1)$. This can represent the constant background of an image. A short wave like $(0,0,1,-1)$ represents a detail at the end of our signal.

$$\text{Input } v \underset{[\text{lossless}]}{\longrightarrow} \text{coefficients } c \underset{[\text{lossy}]}{\longrightarrow} \text{compressed } \hat{c} \underset{[\text{reconstruct}]}{\longrightarrow} \text{compressed } \hat{v}$$

  - Ideally, we don't need the compression step and transform gives us $c = W^{-1}v$ and the reconstruction gives us $v = Wc$. In true signal processing, we have $\hat{v} = W\hat{c}$
- Example
  - For a vector $(6,4,5,1) = v$, its wavelet coefficients are $c = (4,1,1,2)$

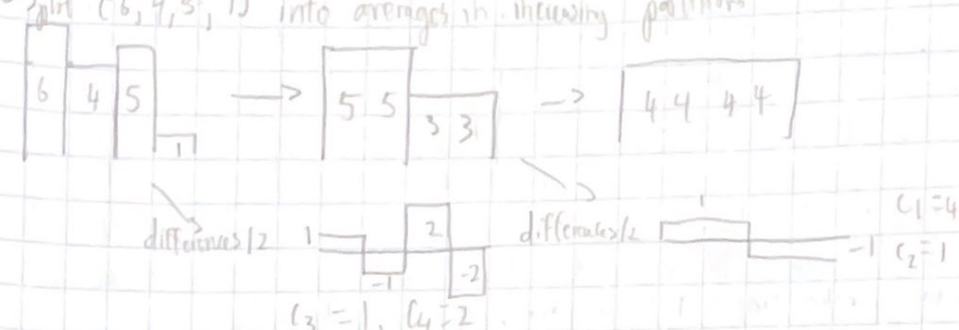
$$\begin{bmatrix}6\\4\\5\\1\end{bmatrix} = 4\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix} + \begin{bmatrix}1\\-1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\0\\1\\-1\end{bmatrix} = \begin{bmatrix}1&1&1&0\\1&1&-1&0\\1&-1&0&1\\1&-1&0&-1\end{bmatrix}\begin{bmatrix}4\\1\\1\\2\end{bmatrix}$$

  - The coefficients are $c = W^{-1}v$. W^{-1} is easy to find because w's are orthogonal. They are not orthonormal, so we need to rescale

$$W^{-1} = \begin{bmatrix}\frac{1}{4}&&&\\&\frac{1}{4}&&\\&&\frac{1}{2}&\\&&&\frac{1}{2}\end{bmatrix}\begin{bmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&0&0\\0&0&1&-1\end{bmatrix}$$

  - The $\frac{1}{4}$'s in the first row of $c = W^{-1}v$ means that $c_1 = 4$ is the average of $6,4,5,1$

- Example
  - You can pick off the details in c_3 and c_4, and the coarse details in c_2 and the average in c_1
  - Split $(6, 4, 5, 1)$ into averages in the wavy partitions



Fourier Transform (DFT) = Change to Fourier Basis

- The DFT involves complex numbers (powers of $e^{2\pi i/n}$). But if we choose $n=4$, the matrices are small and the only complex numbers are i and $i^3 = -i$

Fourier basis w_1 to w_4 in the columns of F

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix}$$

- The first column is the flat wave $(1,1,1,1)$. It represents the average signal or the direct current. It is a wave of zero frequency.
- The third column is $(1,-1,1,-1)$, which alternates at the highest frequency
- The Fourier transform decomposes the signal into waves at equally spaced frequencies

Problem set 7.2

1.

$S v_1 = 0 \Rightarrow b_1 = (0,0,0,0)$

$S v_2 = 0 \Rightarrow b_2 = (0,0,0,0)$

$S v_3 = 2 \Rightarrow b_3 = (2,0,0,0)$

$S v_4 = 6x \Rightarrow b_4 = (0,6,0,0)$

$$B = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.

$y = Ax + B$

3.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow A^2 = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5.

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

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

7.

$$\begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$$

10.

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

$T(v_1 - v_2) = w_1$

11.

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

$T^{-1}(w_1) = v_1$

$T^{-1}(w_2) = v_2 - v_1$

$T^{-1}(w_3) = v_3 - v_2$

13.

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

b)

14.

a) $\begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$

b) $\begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$

20.

a) $w_2 = -(x^2 - 1) = (0, 1, 0)$ $\quad (0,0,1)$

b) $w_3 = \frac{1}{2}(x^2 - x) = (1, 0, 0)$

$y = 4w_1 + 5w_2 + 6w_3$

21.

$T(v) = v$

$T(1) = 1 = w_1 - w_3 = (1, 0, -1)$

$T(x) = x =$

22.

$4 = A + aB + a^2C$

$6 = A + bB + b^2C$

$6 = A + cB + c^2C$

$$\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix} = A$$

$\det A = bc^2 + ab^2 + a^2c - a^2b - ac^2 - b^2c$

24.

$T(v) = aT(v_1) + bT(v_2) + cT(v_3)$

$= a\lambda_1 v_1 + b\lambda_2 v_2 + c\lambda_3 v_3$

$T(v_1) = \lambda_1 v_1$

$$A = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

34.

$(7,5,3,1) = (6,0,2,2) + (1,-1,1,-1)$

$= (4,4,4,4) + (2,2,-2,-2) + (1,-1,1,-1)$

$= 4(1,1,1,1) + 2(1,1,-1,-1) + (1,-1,1,-1)$

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

Chapter 7.3: Diagonalization and the Pseudoinverse

- This section produces better matrices by choosing better bases.
  - When the goal is a diagonal matrix, one way is a basis of eigenvectors
  - The other way is two bases (the input and output bases are different). The left and right singular vectors are orthonormal basis vectors for the 4 fundamental subspaces. They come from SVD.
- By reversing those input and output bases we find the "pseudoinverse" of A. This matrix A^+ sends $\mathbb{R}^m$ back to $\mathbb{R}^n$ and the column space back to the row space
- All our great factorizations of A can be viewed as a change of basis; we'll focus on two:
  1. $S^{-1}AS = \Lambda$ when the input and output bases are the eigenvectors of A
  2. $U^{-1}AV = \Sigma$ when those bases are eigenvectors of A^TA and AA^T
- In Λ, the bases are the same. Then $m = n$ and the matrix A must be square. Some square matrices cannot be diagonalized by any S because they don't have n independent eigenvectors
- In Σ, the input and output bases are different. The matrix A can be rectangular. The bases are orthonormal because A^TA and AA^T are symmetric. Then $U^{-1} = U^T$ and $V^{-1} = V^T$. Every matrix A is allowed and A has the diagonal form Σ
- The eigenvector basis is orthonormal only when $A^TA = AA^T$ (a "normal" matrix). That includes symmetric, antisymmetric and orthogonal matrices. In this case the singular values of Σ are the absolute values $\sigma_i = |\lambda_i|$. The two diagonalizations are the same when $A^TA = AA^T$ except for possible factors -1 and $e^{i\theta}$.
- The factorization $A = QR$ chooses only one new basis. That is the orthogonal output basis given by Q. The input uses the standard basis given by I. The output basis matrix appears on the left and the input basis appears on the right, in $A = QRI$. We start with input basis = output basis, that will produce S and S^{-1}

Similar Matrices: A and $S^{-1}AS$ and $W^{-1}AW$

- Begin with a square matrix and one basis. The input space V is $\mathbb{R}^n$ and the output space W is also $\mathbb{R}^n$. The standard basis vectors are the columns of I.
  The linear transformation T is "multiplication by A".
- The change from A to Λ comes from a change of basis: Eigenvalue matrix from eigenvector basis.
- When you change the basis for V, the matrix changes from A to AM. Because V is the input space, M goes on the right (to come first). When you change the basis for W, the new matrix is $M^{-1}A$. We are working with the output space so M^{-1} is on the left.
- If you change both bases in the same way, the new matrix is $M^{-1}AM$. The good basis vectors are the eigenvectors of A, when the matrix becomes $S^{-1}AS = \Lambda$

  When the basis contains the eigenvectors $x_1, \ldots, x_n$, the matrix for T is Λ.

  - To find column 1 of the matrix, input the first basis vector x_1. The transformation multiplies by A. The output is $Ax_1 = \lambda x_1$, giving us a basis vector $(\lambda_1, 0, \ldots, 0)$.
- Example
  - Project onto the line $y = -x$. $(1,0)$ projects to $(.5, -.5)$. $(0,1)$ projects to $(-.5, .5)$
    1. Standard matrix: project standard basis $A = \begin{bmatrix} .5 & -.5 \\ -.5 & .5 \end{bmatrix}$
    2. Find the diagonal matrix Λ in the eigenvector basis
       eigenvectors are $x_1 = (1,-1)$ and $x_2 = (1,1)$
       Projection of $x_1 = x_1 \Rightarrow$ basis of $(1,0)$ $\quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
       Projection of $x_2 = 0 \Rightarrow$ basis of $(0,0)$
    3. Find a third matrix B using another basis $v_1 = w_1 = (2,0)$ and $v_2 = w_2 = (1,1)$
       w_1 is not an eigenvector, so B will not be diagonal
       Projection of $v_1 = (1,-1) = w_1 - w_2 \Rightarrow$ basis of $(1,-1)$ $\quad B = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$
       Projection of $v_2 = (0,0) \Rightarrow$ basis of $(0,0)$
  - Another way to find B: use W^{-1} and W to change between the standard basis and the w's. Those change of basis matrices are representing the identity transformation. The product of transformations is just ITI. The product of matrices is $W^{-1}AW$. B is similar to A.
- For any basis $w_1, \ldots, w_n$, we find B in 3 steps. Change the input basis to the standard basis with W. The matrix in the standard matrix is A. Change the output basis back to the w's with W^{-1}. Then $B = W^{-1}AW$ represents ITI
- A change of basis produces a similarity transformation to $W^{-1}AW$ in the matrix.

  $$B_{w\text{'s to }w\text{'s}} = W^{-1}_{\text{standard to }w\text{'s}} \, A_{\text{standard}} \, W_{w\text{'s to standard}}$$

- Example
  - Continuing with the previous example

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

Rewrites A (in standard basis) with W as the basis

$$B=W^{-1}AW=\begin{bmatrix}\frac{1}{2}&-\frac{1}{2}\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}\end{bmatrix}\begin{bmatrix}2&1\\0&1\end{bmatrix}=\begin{bmatrix}1&0\\-1&0\end{bmatrix}$$

The Singular Value Decomposition (SVD)

- Now the input basis can differ from the output basis $u_1, \dots, u_m$.
- Again, the best matrix is diagonal (now m by n). To achieve this diagonal Σ_1, each input vector v_j must transform into a multiple of the output vector u_j. That multiple is the singular value σ_j.

SVD $\quad Av_j = \begin{cases}\sigma_j u_j & \text{for } j \le r\\ 0 & \text{for } j > r\end{cases}$ with orthonormal bases

- A and Σ_1 represent the same transformation. $A = U\Sigma_1 V^T$ uses the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$. The diagonal Σ_1 uses the input basis of v's and the output bases of u's. The orthogonal matrices V and U give the basis changes. They represent the identity transformations (in $\mathbb{R}^n$ and $\mathbb{R}^m$)

$$\Sigma_{1\,v\text{'s to }u\text{'s}} = U^{-1}_{\text{standard to }u\text{'s}}\, A_{\text{standard}}\, V_{v\text{'s to standard}}$$

Polar Decomposition

- Every complex number has the polar form $re^{i\theta}$. A non-negative number r multiplies a number on the unit circle. Think of these numbers as 1 by 1 matrices: $r \ge 0$ corresponds to a positive semidefinite matrix, call it H, and $e^{i\theta}$ corresponds to an orthogonal matrix Q. The polar decomposition extends this factorization to matrices: orthogonal times semidefinite. $A = QH$
- Every real square matrix can be factored into $A = QH$, where Q is orthogonal and H is symmetric positive semidefinite.

$$A = U\Sigma_1 V^T = UV^TV\Sigma V^T = (UV^T)(V\Sigma_1 V^T) = QH$$

- UV^T: the product of orthogonal matrices is also orthogonal
- $V\Sigma_1 V^T$: It is positive semidefinite because its eigenvalues are in Σ_1. If A is invertible, so are Σ_1 and H. H is the symmetric positive definite square root of A^TA ($H^2 = V\Sigma_1^2 V^T = A^TA$)

- There is also a $A = KQ$ in the reverse direction. Q is the same but $K = U\Sigma_1 U^T$. This is the symmetric positive definite square root of AA^T

- Example

$$A=\begin{bmatrix}2&2\\-1&1\end{bmatrix}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}\sqrt{2}&\\&2\sqrt{2}\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}=U\Sigma V^T$$

$$Q=UV^T,\quad H=V\Sigma V^T \text{ or } H=Q^{-1}A=Q^TA$$

$$Q=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix},\quad H=\begin{bmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}2&2\\-1&1\end{bmatrix}=\begin{bmatrix}\frac{3}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{3}{\sqrt{2}}\end{bmatrix}$$

- In mechanics, Q is a rotation and H is the stretching factor
  - The eigenvalues of H are the singular values of A; they give the stretching factors
  - The eigenvectors of H are the eigenvectors of A^TA. They give the stretching directions.
  - Then Q rotates those axes
- $A=QH$ splits $Av_i=\sigma_i u_i$ into 2 steps; H multiplies v_i by σ_i. Q rotates v_i into u_i

The Pseudo Inverse

- By choosing good bases, A multiplies v_i in the row space to give $\sigma_i u_i$ in the column space. A^{-1} must do the opposite; $A^{-1}u=v/\sigma$. The singular values of A^{-1} are $1/\sigma$, just as the eigenvalues of A^{-1} are $1/\lambda$. The u's are in the row space of A^{-1} and the v's are in the column space
- A matrix that multiplies u_i to produce v_i/σ_i does exist and is the pseudoinverse A^+.

$$\text{Pseudoinverse: } A^+=V\Sigma^+U^T=\underset{n \text{ by } n}{\begin{bmatrix}v_1&\cdots&v_r&\cdots&v_n\end{bmatrix}}\underset{n \text{ by } m}{\begin{bmatrix}\sigma_1^{-1}&&\\&\ddots&\\&&\sigma_r^{-1}\end{bmatrix}}\underset{m \text{ by } m}{\begin{bmatrix}u_1&\cdots&u_r&\cdots&u_m\end{bmatrix}^T}$$

- The pseudoinverse A^+ is an n by m matrix. If A^{-1} exists, then A^+ is the same as A^{-1}.
  - In that case $m=n=r$ and we are inverting $U\Sigma V^T$ to get $V\Sigma^{-1}U^T$.
- If $r<m$ or $r<n$, then A has no two-sided inverse, but a pseudoinverse A^+ with the same rank r

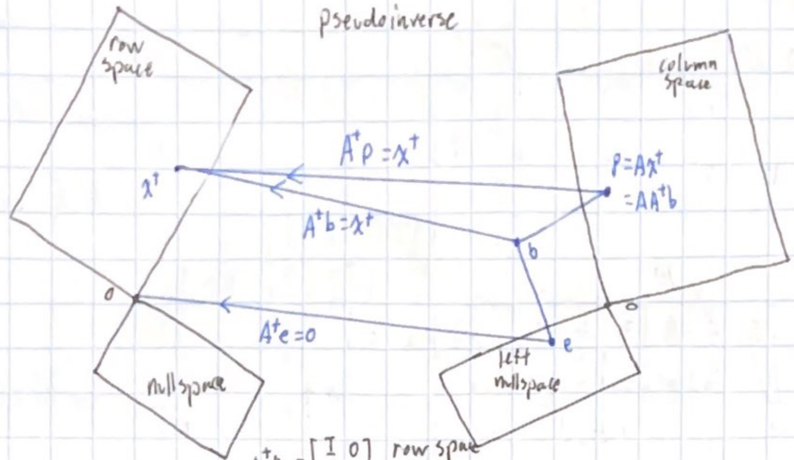
$$A^+u_i=\frac{1}{\sigma_i}v_i \text{ for } i\le r \text{ and } A^+u_i=0 \text{ for } i>r$$

- The vectors $u_1,\dots,u_r$ in the column space of A go back to $v_1,\dots,v_r$ in the row space.
- The other vectors $u_{r+1},\dots,u_m$ are in the left nullspace and A^+ sends them to 0
- Notice the pseudoinverse Σ^+ of the diagonal matrix. Each σ is replaced by σ^{-1}. The product $\Sigma^+\Sigma$ is as near to identity as we can get it (It is partly I and partly 0). We get r 1's. We take the reciprocal of Σ and transpose it.

$$\Sigma^+\Sigma=\begin{bmatrix}\frac{1}{2}&0&0\\0&\frac{1}{3}&0\\0&0&0\end{bmatrix}\begin{bmatrix}2&0&0\\0&3&0\\0&0&0\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}\quad \Sigma^+\Sigma \text{ is a projection matrix}$$

- The pseudoinverse is the n by m matrix that makes AA^+ and A^+A into projections

pseudoinverse



$A^+A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ row space nullspace

fig 7.3.1: Ax^+ in the column space goes back to $A^+Ax^+ = x^+$

- Trying for $AA^{-1} = A^{-1}A = I$
  - AA^+ = Projection matrix onto the column space of A
  - A^+A = Projection matrix onto the row space of A
- Example
  - Find the pseudoinverse of $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$. A has 1 singular value $\sqrt{10}$

    $A^+ = V\Sigma_1^{+}U^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \sqrt{10}/10 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$
  - A^+ also has rank 1. Its column space is the row space of A. When A takes (1, 1) in the row space to (4, 2) in the column space, A^+ does the reverse: $A^+(4,2) = (1,1)$.
  - Every rank 1 matrix is a columns times a row. With unit vectors u and v, that is $A = \sigma uv^T$. Then the best inverse is $A^+ = vu^T/\sigma$. The product AA^+ is uu^T, the projection onto the line through u. The product A^+A is vv^T.
- Application to least squares
  - The equation $A^TA\hat{x} = A^Tb$ assumes A^TA is invertible
  - A may have dependent columns (rank < n) and there are many solutions to $A^TA\hat{x} = A^Tb$. One solution is $x^+ = A^+b$ from the pseudoinverse
  - We can check that A^TAA^+b is A^Tb because fig 7.3.1 shows that $e = b - AA^+b$ is the part of b in the left nullspace. Any vector in the nullspace of A can be added to give another solution, but x^+ will be the shortest

The shortest least squares solution to $Ax = b$ is $x^+ = A^+b$

Problem Set 7.3

1.

a) $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

$A^TA = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix}$

$\lambda = 0, 50$

case $\lambda = 0$

$(A^TA - 0I)x_1 = 0$

$\begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix} x_1 = 0$

$x_1 = (2, -1)$

$v_1 = \left(\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5}\right)$

case $\lambda = 50$:

$(A^TA - 50I)x_2 = 0$

$\begin{bmatrix} -40 & 20 \\ 20 & -10 \end{bmatrix} x_2 = 0$

$x_2 = (1, -2)$

$v_2 = \left(\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5}\right)$

$Av_1 = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} \frac{2\sqrt{5}}{5} \\ -\frac{\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\sigma_1 = 0 \quad u_1 = (0, 0)$

Choose v_1 so $v_1 \perp v_2$

$Av_2 = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} \sqrt{5} \\ 3\sqrt{5} \end{bmatrix}$

$\sigma_2 = 5\sqrt{2}$, $u_2 = \left(\frac{\sqrt{10}}{10}, \frac{3\sqrt{10}}{10}\right)$

$u_1 = \left(-\frac{3\sqrt{10}}{10}, \frac{\sqrt{10}}{10}\right)$

$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{10}}{10} & -\frac{3\sqrt{10}}{10} \\ \frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \end{bmatrix}\begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix}$

2.

a)

basis $C(A^T) = \{(\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5})\}$

basis $N(A) = \{(\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5})\}$

basis $C(A) = \{(\frac{\sqrt{10}}{10}, \frac{3\sqrt{10}}{10})\}$

basis $N(A^T) = \{(-\frac{3\sqrt{10}}{10}, \frac{\sqrt{10}}{10})\}$

4.

$$A^+ = \begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{10} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{10}}{10} & \frac{3\sqrt{10}}{10} \\ -\frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \end{bmatrix} = \begin{bmatrix} \frac{1}{50} & \frac{3}{50} \\ \frac{1}{25} & \frac{3}{25} \end{bmatrix}$$

$$P_{row} = A^+A = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{bmatrix}$$

$$P_{col} = AA^+ = \begin{bmatrix} \frac{1}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{9}{10} \end{bmatrix}$$

3.

$$Q = UV^T = \begin{bmatrix} \frac{\sqrt{10}}{10} & -\frac{3\sqrt{10}}{10} \\ \frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$H = V\Sigma V^T = \begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 2\sqrt{2} & 4\sqrt{2} \end{bmatrix}$$

16.

$A = [3 \quad 4 \quad 0]$

$$A^TA = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} [3 \quad 4 \quad 0] = \begin{bmatrix} 9 & 12 & 0 \\ 12 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\lambda = 0, 0, 25$

$v_1 = \frac{\sqrt{5}}{5}(-\frac{4}{3}, 1, 0)$

$v_2 = (0, 0, 1)$

$v_3 = \frac{4}{5}(\frac{3}{4}, 1, 0)$

15.

$A^TA\hat{x} = A^Tb$

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \hat{x} = \begin{bmatrix} \frac{1}{5} & \frac{3}{10} \\ \frac{1}{5} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \hat{x} = \begin{bmatrix} \frac{1}{5}b_1 + \frac{1}{10}b_2 \\ 0 \end{bmatrix}$$

A^TA is singular

17.

$AA^+p = p$

$AA^+e = 0$

$A^+Ax_r = x_r$

$A^+Ax_n = 0$

18.

$A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$A^TA = [3 \quad 4] \begin{bmatrix} 3 \\ 4 \end{bmatrix} = [25]$

$\lambda = 25$

Case $\lambda = 25$

$(A - 25I)x_1 = 0$

$x_1 = (1)$

$v_1 = (1)$

$Av_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}[1] = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$\sigma_1 = 5, \; u_1 = (\frac{3}{5}, \frac{4}{5})$

choose $u_2 = (\frac{4}{5}, -\frac{3}{5})$

$$A = U\Sigma V^T \Rightarrow \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} [1]$$

$$A^+ = V\Sigma^+U^T = [1][\tfrac{1}{5} \quad 0] \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} = [\tfrac{3}{25} \quad \tfrac{4}{25}]$$

$A^+A = [1]$

$$AA^+ = \begin{bmatrix} \frac{9}{25} & \frac{12}{25} \\ \frac{12}{25} & \frac{16}{25} \end{bmatrix}$$

$x^+ = A^+b = [\tfrac{3}{25} \quad \tfrac{4}{25}] \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 1$

$x^+ = A^+b = [\tfrac{3}{25} \quad \tfrac{4}{25}] \begin{bmatrix} -4 \\ 3 \end{bmatrix} = 0$

Chapter 8: Applications

Chapter 8.1: Matrices in Engineering

- Engineering problems produce symmetric matrices K, which is often positive definite. This because a) $K = A^TA$ and $K = A^TCA$, and b) the expression $\frac{1}{2}u^TKu$ represents energy, which is never negative. The matrix C, often diagonal, contains positive physical constants
- Our first example: K is the stiffness matrix, and $K^{-1}f$ is the structure's responses to forces f from outside
- Engineering leads to linear algebra in two ways:
  - Direct way: the laws connecting position or velocity are linear. Expressed as matrix eqs.
  - Indirect way: the physical system is continuous; expressed as differential equations. To find accurate solutions, we approximate by finite difference or finite element equations

- We use the equilibrium equations $Ku = f$. With motion $M\frac{d^2u}{dt^2} + Ku = f$ becomes dynamic. Then we use eigenvalues from $Kx = \lambda Mx$, or finite differences.
- The matrices

$$K_0 = A_0^T A_0 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad A_0^T C_0 A_0 = \begin{bmatrix} c_1+c_2 & -c_2 & 0 \\ -c_2 & c_2+c_3 & -c_3 \\ 0 & -c_3 & c_3+c_4 \end{bmatrix}$$

Fixed-fixed — Spring constants included

$$K_1 = A_1^T A_1 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \qquad A_1^T C_1 A_1 = \begin{bmatrix} c_1+c_2 & -c_2 & 0 \\ -c_2 & c_2+c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{bmatrix}$$

Fixed-free — Spring constants included

$$K_{singular} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \qquad K_{circular} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Free-free

  - The matrices $K_0, K_1, K_{singular}$, and $K_{circular}$ have $C = I$ for simplicity. This means all spring constants are $c_i = 1$

A Line of Springs

- Fig 8.1.1 shows 3 masses m_1, m_2, m_3 connected by a line of springs. One case has 4 springs, with top and bottom fixed. The fixed-free case has 3 springs: the lowest mass hangs freely. The fixed-fixed problem will lead to K_0 and $A_0^T C_0 A_0$. The fixed-free problem will lead to K_1 and $A_1^T C_1 A_1$. A free-free problem produces $K_{singular}$.
- We want equations for the mass movements u and the tensions y:
  $u = (u_1, u_2, u_3)$ = movement of the masses
  $y = (y_1, y_2, y_3, y_4)$ or (y_1, y_2, y_3) = tensions in the springs
- When a mass moves downward, its displacement is positive ($u_i > 0$). For the springs, tension is positive. Hooke's Law $y = ce$.
- We want to link these equations into $Ku = f$. f comes from gravity, so we get $f = (m_1 g, m_2 g, m_3 g)$

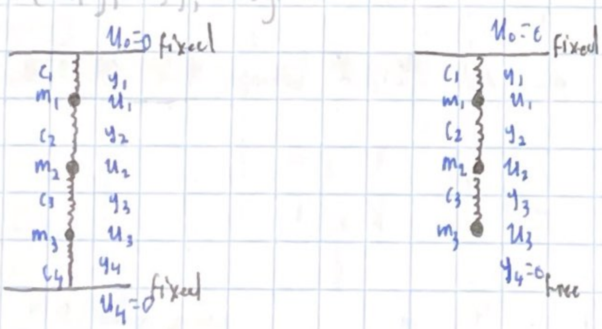


Fig 8.1.1: Fixed-fixed and fixed-free spring lines.

- We have the following quantities:

  $u = (u_1, u_2, \ldots, u_n)$ Movements of n masses — $y = (y_1, y_2, \ldots y_m)$ Internal forces in m springs

  $e = (e_1, e_2, \ldots, e_m)$ Elongation of m springs — $f = (f_1, f_2, \ldots, f_n)$ External forces on n masses

- The framework that connects u to e to y to f looks like this:

$$\begin{array}{ccc} \boxed{u} & & \boxed{f} \\ A\downarrow & & \uparrow A^T \\ \boxed{e} & \xrightarrow{C} & \boxed{y} \end{array}$$

$e = Au$ — A is m by n

$y = Ce$ — C is m by m

$f = A^T y$ — A^T is n by m

- The elongations e is the displacements of each spring. When the system is vertical, the masses move by distances u_1, u_2, u_3. Each spring is displaced by $e_i = u_i - u_{i-1}$, the difference in displac

Stretching of each spring

$e_1 = u_1 \quad (u_0 = 0)$

$e_2 = u_2 - u_1$

$e_3 = u_3 - u_2$

$e_4 = \quad - u_3 \quad (u_4 = 0)$

- This is a 4 by 3 (m by n) difference matrix

$$e = Au \text{ is } \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \\ & & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

- The next equation $y = Ce$ is just Hooke's law

$$\begin{aligned} y_1 &= c_1 e_1 \\ y_2 &= c_2 e_2 \\ y_3 &= c_3 e_3 \\ y_4 &= c_4 e_4 \end{aligned} \quad \text{is} \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} c_1 & & & \\ & c_2 & & \\ & & c_3 & \\ & & & c_4 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

- Combining $e = Au$ with $y = Ce$, we get $y = CAu$
- The internal spring forces balance the gravitational forces
  - Each mass is pushed or pulled by the spring force above it (y_j). From below it feels the spring force y_{j+1} and f_j from gravity. Thus $y_j = y_{j+1} + f_j$ or $f_j = y_j - y_{j+1}$

$$\begin{aligned} f_1 &= y_1 - y_2 \\ f_2 &= y_2 - y_3 \\ f_3 &= y_3 - y_4 \end{aligned} \quad \text{is} \quad \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

- This is $f = A^T y$. These combine into $Ku = f$, where the stiffness matrix is $K = A^T C A$

$$\left.\begin{aligned} e &= Au \\ y &= Ce \\ f &= A^T y \end{aligned}\right\} \text{ combine into } A^T C A u = f \text{ or } Ku = f$$

- In the language of elasticity, $e = Au$ is the kinematic equation (for displacement). The force balance $f = A^T y$ is the static equation (for equilibrium). The constitutive law is $y = Ce$

- With $C = I$, we get

$$K_0 = A_0^T A_0 = \begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix}$$

- Properties:
  1. K is <u>tridiagonal</u>, because mass 3 is not connected to mass 1.
  2. K is <u>symmetric</u> because C is symmetric and A^T comes with A.
  3. K is <u>positive semidefinite</u> because $c_i > 0$ and A has independent columns
  4. K^{-1} is a full matrix with <u>all positive entries</u>
- Example
  - Suppose $c_i = c$ and $m_j = m$. Find the movements u and tensions y.

$$u = K^{-1} f = \frac{1}{4c}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} mg \\ mg \\ mg \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} \frac{3}{2} \\ 2 \\ \frac{3}{2} \end{bmatrix}$$

  - The displacement u_2 is larger than the others

$$e = Au \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \frac{mg}{c}\begin{bmatrix} \frac{3}{2} \\ 2 \\ \frac{3}{2} \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{3}{2} \end{bmatrix} \Big\} \text{ add to } 0$$

  - To find the spring force, multiply e by c; $y = (\frac{3}{2}mg, \frac{1}{2}mg, -\frac{1}{2}mg, -\frac{3}{2}mg$
- Warning: Normally you cannot write $K^{-1} = A^{-1} C^{-1} (A^T)^{-1}$ bc A is rectangular

Fixed End and Free End

- Remove the fourth spring. All matrices becomes 3 by 3; the pattern does not change.

$$A_1^T C_1 A_1 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} c_1 & & \\ & c_2 & \\ & & c_3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

and (set $c_4 = 0$)

$$K_1 = A_1^T C_1 A_1 = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{bmatrix}$$

- Example
  - All $c_i = c$ and all $m_j = m$

$$K_1 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \text{ and } K_1^{-1} = \frac{1}{c}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$u = K_1^{-1} f = \frac{1}{c}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} mg \\ mg \\ mg \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}$$

3+2

3+2+1

$$e = Au \quad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \frac{mg}{c}\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix} = \frac{mg}{c}\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

Two Free Ends: K is Singular

- Now the matrix A is 2 by 3:

$$e = Au \Rightarrow \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} u_2 - u_1 \\ u_3 - u_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

- Now there is a nonzero to $Au = 0$. The masses can move without stretching the springs. The whole line can shift by $u = (1,1,1)$ and the leaves $u = (0,0)$. A has dependent columns and the vector $(1,1,1)$ is in its nullspace:

$$Au = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \text{no stretching}$$

- $Au = 0$ leads $A^TCAu = 0$, so A^TCA is only positive semidefinite, without C_1 and C_4. The pivots will be C_2 and C_3 and no third pivot. The rank is only 2

$$\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C_2 & 0 \\ 0 & C_3 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} C_2 & -C_2 & 0 \\ -C_2 & C_2 + C_3 & -C_3 \\ 0 & -C_3 & C_3 \end{bmatrix}$$

- Two eigenvalues will be positive but $x = (1,1,1)$ is an eigenvector for $\lambda = 0$. We can solve $A^TCAu = f$ only for special vectors f. The forces need to add to $f_1 + f_2 + f_3 = 0$.

Circle of Springs

- A third spring will complete the circle from mass 3 to mass 1. K is still singular

$$K_{circular} = A^T_{circular} A_{circular} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

- $K_{circular}$ is always symmetric and semidefinite
- We have pivots 2 and $\frac{3}{2}$ and eigenvalues are 3 and 3 and 0. The nullspace still contains $x = (1,1,1)$

Continuous Instead of Discrete

- Matrix equations are discrete. Differential equations are continuous. We'll see the differential equations that corresponds to the tridiagonal $-1, 2, -1$ matrix A^TA.
- The matrices A and A^T correspond to the derivatives $\frac{d}{dx}$ and $-\frac{d}{dx}$. Remember that $e = Au$ takes differences $u_i - u_{i-1}$ and $f = A^Ty$ takes differences $y_i - y_{i+1}$

$$\frac{u_i - u_{i-1}}{\Delta x} \text{ is like } \frac{du}{dx} \qquad \frac{y_i - y_{i+1}}{\Delta x} = -\frac{dy}{dx}$$

- Δx didn't appear earlier, we imagine the distance between masses was 1.

$$e(x) = Au = \frac{du}{dx} \quad y(x) = c(x)e(x) \quad A^Ty = -\frac{dy}{dx} = f(x)$$

- Combining the equations, we get $A^TCAu(x) = f(x)$, we have a differential equation.
- The line of springs becomes an elastic bar

$$A^TCAu(x) = f(x) \text{ is } -\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x)$$

- A^TA corresponds to a second derivative. A is a difference matrix and A^TA is a second derivative matrix. The matrix has $-1, 2, -1$ and the equation has $-d^2u/dx^2$

  $-u_{i+1} + 2u_i - u_{i-1}$ is a second difference $-\frac{d^2u}{dx^2}$ is a second derivative

- We usually have 3 choices for a derivative (forward, backward, centered difference)

  $$\frac{du}{dx} \approx \frac{u(x+\Delta x)-u(x)}{\Delta x} \quad \text{or} \quad \frac{u(x)-u(x-\Delta x)}{\Delta x} \quad \text{or} \quad \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}$$

- For $\frac{d^2u}{dx^2}$, we have $[u(x+\Delta x) - 2u(x) + u(x-\Delta x)]/(\Delta x)^2$. The signs are $-1, 2, -1$ because first derivative is antisymmetric, so the second differences are negative definite and we change to $-d^2u/dx^2$
- In scientific computing, we typically create the discrete matrix K by approximately the continuous problem

Fixed
Fixed
$$Au = \frac{1}{\Delta x}\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\\0&0&-1\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix} \approx \frac{du}{dx} \quad \text{with } u_0 = u_4 = 0$$

Fixed
Free
$$A^Ty = \frac{1}{\Delta x}\begin{bmatrix}1&-1&0\\0&1&-1\\0&0&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} \approx -\frac{dy}{dx} \quad \text{with } u_0 = y_4 = 0$$

- Process
  1. Model the problem with a differential equation
  2. Discretize the differential equation to a difference equation
  3. Understand and solve the difference equation (and boundary conditions)
  4. Interpret the solution; visualize it; redesign if needed

Problem Set 8.1

1.

$$\det\begin{bmatrix}c_1+c_2 & -c_2 & 0\\ -c_2 & c_2+c_3 & -c_3\\ 0 & -c_3 & c_3+c_4\end{bmatrix} = (c_1+c_2)(c_2+c_3)(c_3+c_4) - (-c_2)(-c_2)(c_3+c_4) - (-c_3)(-c_3)(c_1+c_2)$$
$$= (c_1c_2 + c_1c_3 + c_2^2 + c_2c_3)(c_3+c_4) - c_2^2c_3 - c_2^2c_4 - c_1c_3^2 - c_2c_3^2$$
$$= c_1c_2c_3 + c_1c_3^2 + c_2^2c_3 + c_2c_3^2 + c_1c_2c_4 + c_1c_3c_4 + c_2^2c_4 + c_2c_3c_4 - c_2^2c_3 - c_2^2c_4 - c_1c_3^2 - c_2c_3^2$$
$$= c_1c_2c_3 + c_1c_2c_4 + c_1c_3c_4 + c_2c_3c_4$$

4.

$$-\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x)$$

$\neq 0$

$$-c(x)\frac{du}{dx} = \int_{x_{end}}^{x} f(t)\,dt$$

this is zero at both endpoints

$$\int_{x_{end}}^{x_{end}} f(x)\,dx = 0$$

$$\int_{x_{end}}^{x_{start}} f(x)\,dx = 0$$

5.

$$\frac{dy}{dx} = f(x) \qquad y = \int_1^x f(t)\,dt$$

$$-dy = f(x)\,dx \qquad y = \int_1^x dt = [t]_1^x$$

$$= x-1$$

$$-\int dy = \int f(x)\,dx$$

$$-y + c = \int f(x)\,dx$$

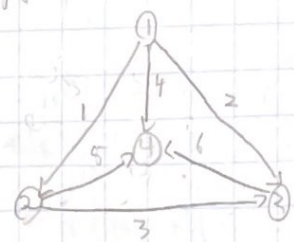
$$y = c - \int f(x)\,dx$$

$$y|_{x=1} \quad 0 = c - \int_1^1 f(t)\,dt \quad c = 0$$

Chapter 8.2: Graphs and Networks

- Graphs are entities of nodes connected by edges
- This section focuses on incidence matrices of graphs, which tell us how the n nodes are connected by the m edges. Normally $m > n$.
- For any m by n matrix, there are two fundamental subspaces in $\mathbb{R}^n$ and two in $\mathbb{R}^m$
  - We review the four subspaces for any matrix and constructed a directed graph and its incidence matrix
  - By specializing to incidence matrices, the laws of linear algebra become Kirchhoff's Laws
- Every entry in an incidence matrix is -1, 0, or 1. This holds during elimination. All pivots and multipliers are ± 1. Therefore both factors in $A = LU$ also had 0 or ± 1.
- Example: Differences in voltage across 6 edges of a graph. The columns are voltages at each node

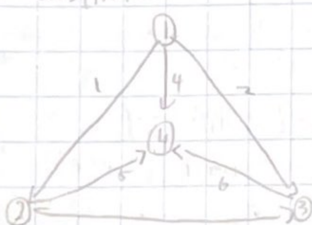
$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \text{ reduces to } U = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



- The nullspace of A and U is line through $x = (1, 1, 1, 1)$. The column space of A and U have dimension $r = 3$. The pivot rows are a basis for the row space. $x = (1, 1, 1, 1)$ in the nullspace is perpendicular to all the basis rows.
  - Equal voltages produce no current.

Directed Graphs and Incidence Matrices

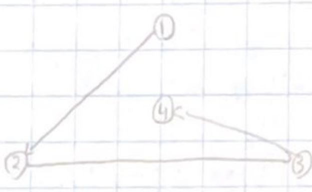
- Looking at the graph row by row, we can establish the directed edges: -1 = source, 1 = destination



$$\begin{array}{c} \\ A = \end{array} \begin{array}{cccc|l} 1 & 2 & 3 & 4 & \text{node numbers} \\ \end{array}$$

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \text{ edge numbers}$$

- Complete graph: every pair of nodes is connected by an edge
- Max edges: $\frac{1}{2} n(n-1)$



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

- Tree: no closed loops
- min edges: $m = n - 1$

- The rows of B match the non zero rows of U. Elimination reduces every graph to a tree.
- The loops produce zero rows in U

$$\text{edges } 1,2,3 \left\{ \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \right. \longrightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- These steps are typical. When two edges share a node, elimination gives the shortcut edge that skips the shared node
- A Idea: Rows are dependent when edges form a loop. Independent rows come from trees. This is the key for the row space
- For the column space, Ax is a vector of differences

$$Ax = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_1 \\ x_4 - x_2 \\ x_4 - x_3 \end{bmatrix}$$

  - The unknowns represent potentials/voltage.
  - The vector Ax is then a vector of potential or electric differences
- The nullspace contains all solution to $Ax = 0$ — all six potentials differences are 0, so all 4 potentials are the same. We can raise or increase all potentials by an arbitrary constant (like Calculus $+C$) without changing the differences or If we set $x_4 = 0$ ("grounding it"), we can get absolute values without a constant
- An arbitrary vector is in the row space if it is orthogonal to $(1,1,1,1)$
- How can we tell if a particular b is in the column space of an incidence matrix
  - Ax is a vector of differences. Adding differences across a closed loop must sum to 0.
  - The components of Ax add to zero around every loop. When b is in the column space, its components in each loop sum to 0.
  - This is Kirchhoff's loop law for voltage:
    The directed sum of the potential differences around any closed loop is 0
- Looking at the left nullspace:

$$A^Ty = \begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ y_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

  - We have 3 equations ($r = 3$)
  - The first equation is $-y_1 - y_2 - y_4 = 0$. The net flow into node 1 is 0.
  - The fourth equation says $y_4 + y_5 + y_6 = 0$. Flow into the node - Flow out is 0.
  - $A^Ty = 0$ is Kirchhoff's Current Law:
    Flow in equals flow out at each node
- What's the solution to $A^Ty = 0$, the currents must balance each other. The easiest way is to flow around a loop: If a unit of current goes around the big triangle (forward on edge 1 and 3, backward on edge 3), we get $y = (1, -1, 1, 0, 0, 0)$. This satisfies $A^Ty = 0$. Every loop current is a solution to the current law.
- We get 3 basis solutions, one for each small loop. The big loop is the sum of these basis vectors
- For every graph in a plane, linear algebra gives Euler's formula:
  (number of nodes) − (number of edges) + (number of small loops) = 1

Networks and A^TCA

- In a real network, the current y along an edge is the product of potential difference (voltage) and conductance c. c is decided by the material.
- The graph is known from its "connectivity matrix", which tells us the connections between nodes and edges. A network goes further and assigns a conductance c_i to each. These constants go into the diagonal matrix C.
- For a network of resistors, the conductance c is $1/(\text{resistance})$
  - From this we have ohms law: $I = \frac{V}{R}$
- Ohm's Law for all m currents is $y = -CAx$. Ax gives potential differences and C multiplies by conductance. Combining with Kirchhoff's Law, we get $A^TCAx = 0$. 0 shouldn't be on the right side. We want external power. Ex. $A^TCAx = (5, 0, 0)$ 5 volts from node 1.
- In circuit theory, we change Ax to $-Av$. The flow is from higher to lower potential.

Chapter 8.3: Markov Matrices, Population, and Economics

- This section is about positive matrices: every $a_{ij} > 0$. The largest eigenvalue is real and positive and so is its eigenvector.
- In economics, ecology, and population dynamics, this facts leads a long way.

  Markov $\lambda_{max} = 1$ population $\lambda_{max} > 1$ consumption $\lambda_{max} < 1$

  - λ_{max} controls the powers of A

Markov Matrices

- Suppose we multiply a vector $u_0 = (a, 1-a)$ again and again by this A

  Markov Matrix $A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$ $u_1 = Au_0$ $u_2 = A^2u_0$ ... $u_n = A^nu_0$
- If we take the limit, we reach a steady state

  $\lim_{k\to\infty} A^k u_0 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$ regardless of the "a" we choose for u_0
  - Doing $\lim_{k\to\infty} Au_k = \lim_{k\to\infty} u_k$ gives that $\lim_{k\to\infty} u_k$ is an eigenvector with eigenvalue 1.
- A Markov Matrix:
  1. Every entry of A is non-negative
  2. Every column of A sums to 1.
- A Markov Matrix has the following properties
  1. Multiplying a nonnegative u_0 by A produces a nonnegative $u_1 = Au_0$
  2. If the components of u_0 sum to 1, so do the components of $u_1 = Au_0$
     - The components of u_0 sum to 1 when $[1 \dots 1]u_0 = 1$. This is true for each column of A by property 2. Then by matrix multiplication $[1 \dots 1]A = [1 \dots 1]$

       components of Au_0 add to 1 $[1 \dots 1]Au_0 = [1 \dots 1]u_0 = 1$
- Every vector A^ku_0 is nonnegative with components that sum to 1.

- Example 1
  - The fraction of rental cars in Denver starts at $\frac{1}{50} = 0.02$. The fraction outside is $\frac{49}{50} = 0.98$. Every month, 80% of the Denver cars stay in Denver, and 20% leave. Also 5% of outside cars come in (95% stay out). Then we get the Markov Matrix

$$A = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix} \begin{matrix} \leftarrow \text{cars in} \\ \leftarrow \text{cars out} \end{matrix} \quad \text{leads to} \quad u_1 = Au_0 = A\begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .065 \\ .935 \end{bmatrix}$$

  - Notice that $0.065 + 0.935 = 1$
  - We calculate powers with the factorization $A = S\Lambda S^{-1} \Rightarrow A^k = S\Lambda^k S^{-1}$
  - Diagonalizing, we get $\lambda = 1$ and $.75$ and $x = (.2, .8)$ and $x = (-1, 1)$
    We write u_0 as a combination of eigenvectors

$$u_0 = \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$A^k u_0 = A^k\begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18A^k\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \underbrace{(1)^k}_{\text{steady state}}\begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18\underbrace{(.75)^k}_{1<|\lambda| \Rightarrow \text{decay}}\begin{bmatrix} -1 \\ 1 \end{bmatrix} \rightarrow \lim_{k\to\infty} A^k u = (1)\begin{bmatrix} .2 \\ .8 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix}$$

  - Eventually we get that 20% of cars are in Denver and 80% outside, no matter u_0, as long as its components sum to 1.
- If A is a positive markov matrix, then $\lambda_1 = 1$ is the largest eigenvalue and eigenvector x_1 is the steady state

$$u_k = x_1 + c_2(\lambda_2)^k x_2 + \dots + c_n(\lambda_n)^k x_n \text{ always approaches } u_\infty = x_1$$

- No eigenvalue can have $|\lambda| \geq 1$, but you should watch out that another eigenvalue has $|\lambda| = 1$
- Example
  - $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has no steady state because $\lambda_2 = -1$.
    This matrix sends all cars inside Denver outside and vice versa, so the states alternate without reaching a steady state
  - If A had all entries strictly positive, $\lambda = 1$ is strictly larger than any other eigenvalue

Perron-Frobenius Theorem
- Applies to matrices with $a_{ij} \geq 0$
- For the strict equality $a_{ij} > 0$, all numbers in $Ax = \lambda_{max} x$ are positive

Population Growth

- Divide the population into 3 groups: $\underset{n_1}{<20}$, $\underset{n_2}{20 \le age < 40}$, $\underset{n_3}{40 \le age < 60}$.
  Twenty years later, the sizes have changed for 2 reasons:
  1. Reproduction: $n_1^{new} = F_1 n_1 + F_2 n_2 + F_3 n_3$ gives a new generation
  2. Survival: $n_2^{new} = P_1 n_1$ and $n_3^{new} = P_2 n_2$ gives the older generations
- The fertility rates are F_1, F_2, F_3 (F_2 largest). The Leslie Matrix A might look like:

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}^{new} = \begin{bmatrix} F_1 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} .04 & 1.1 & .01 \\ .98 & 0 & 0 \\ 0 & .92 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$$

  - This is population projection at its simplest, with a constant A. More advanced projections may have changing A's
- The matrix has $A \ge 0$ but not $A > 0$. The Perron-Frobenius theorem still applies because $A^3 > 0$. The largest eigenvalue is $\lambda_{max} \approx 1.06$. You can see the generations move

$$A^2 = \begin{bmatrix} 1.08 & 0.05 & .00 \\ 0.04 & 1.08 & .01 \\ 0.90 & 0 & 0 \end{bmatrix} \quad A^3 = \begin{bmatrix} 0.10 & 1.19 & .01 \\ 0.06 & 0.05 & .00 \\ .04 & .99 & .01 \end{bmatrix}$$

- λ_{max} is the asymptotic growth rate

Linear Algebra in Economics: The Consumption Matrix

- The consumption matrix tells how much of each input goes into a unit of output. This describes the manufacturing side of the economy.
- We have n industries like chemicals, food, and oil. To produce a unit of chemicals, we may require .2 units of chemicals, .3 units of food, and .4 units of oil. Those numbers go into row 1 of the consumption matrix.

$$\begin{bmatrix} \text{chemical output} \\ \text{food output} \\ \text{oil output} \end{bmatrix} = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix} \begin{bmatrix} \text{chemical input} \\ \text{food input} \\ \text{oil input} \end{bmatrix}$$

- Similarly, row 2 shows the inputs to produce food and row 3 shows the inputs to produce oil.
- The question here is: Can the economy meet demands y_1, y_2, y_3 for chemicals, food, and oil. To do that, the inputs p_1, p_2, p_3 will have to be higher because part of p is consumed to produce y. The input is p and the consumption of Ap, which leaves the output $p - Ap$. This net production meets the demand y.
  - Find the vector p such that $p - Ap = y$ or $p = (I-A)^{-1} y$.
    So when is $(I-A)^{-1}$ a nonnegative matrix
- If A is small compared to I, then Ap is small compared to p. There is plenty of output. If A is too large, then production consumes more than it yields. In this case the external demand y cannot be met

- "Small" or "large" is decided by the largest eigenvalue λ_1 of A (which is positive)
  - If $\lambda_1 > 1$ then $(I-A)^{-1}$ has negative entries
  - If $\lambda_1 = 1$ then $(I-A)^{-1}$ fails to exist
  - If $\lambda_1 < 1$ then $(I-A)^{-1}$ is nonnegative as desired
- We have an infinite series $1 + x + x^2 + \dots = 1/(1-x)$ if x lies between -1 and 1. Similarly, for matrices, we have $(I-A)^{-1} = I + A^2 + A^3 + \dots$
  - If we multiply $S = I + A + A^2 + \dots$ by A, we get the series except for I. Therefore $S - AS = I \Rightarrow (I-A)S = I$. Therefore $S = (I-A)^{-1}$ if it converges. It converges if all λ of A have $|\lambda| < 1$
- Example:

$$A = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix} \text{ has } \lambda_{max} = .9 \text{ and } (I-A)^{-1} = \frac{1}{93}\begin{bmatrix} 41 & 25 & 27 \\ 33 & 36 & 24 \\ 34 & 23 & 36 \end{bmatrix}$$

  - This economy is productive. A is small compared to I, because $\lambda_{max} = .9$. To meet the demand y, start from $p = (I-A)^{-1}y$. Then Ap is consumed in production, leaving $p - Ap$. This is $(I-A)p = y$ and the demand is met.
- Example

$$A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix} \text{ has } \lambda_{max} = 2 \text{ and } (I-A)^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$$

  - This consumption matrix A is too large. Demands cannot be met, because the production consumes more than it yields.

Problem Set 8.3

1.

$$A = \begin{bmatrix} .90 & .15 \\ .10 & .85 \end{bmatrix}$$

$$(A - I)x_1 = \begin{bmatrix} -.10 & .15 \\ .10 & -.15 \end{bmatrix} x_1 = 0$$

$$x_1 = (3, 2)$$

$$(A - 0.75I)x_2 = \begin{bmatrix} .15 & .15 \\ .10 & .10 \end{bmatrix} x_2 = 0$$

$$x_2 = (1, -1)$$

$$A = S\Lambda S^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & .75 \end{bmatrix}\begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{bmatrix}$$

15.

$$(I-A)^{-1}, \quad (I-A)^{-1}$$

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

$$= \begin{bmatrix} 1 & -4 \\ -.2 & 1 \end{bmatrix}^{-1} = \frac{1}{.8}\begin{bmatrix} 1 & 4 \\ .2 & 1 \end{bmatrix}$$

Chapter 8.4: Linear Programming

- Linear programming = linear algebra + inequalities + minimization
- We start at $Ax=b$, but the only acceptable solutions are nonnegative. The matrix has $n \geq m$; more unknowns than equations. If there are solutions, there are infinite. Linear programming picks the solution $x^* \geq 0$ that minimizes cost.
  - The cost is $c_1x_1 + \dots + c_2x_n$.
- Thus, a linear programming problem starts with a matrix A and two vectors b and c:
  - i) A has $n > m$; for example, $A = [1 \ 1 \ 2]$
  - ii) b has m components for m equations $Ax=b$; for example $b=[4]$
  - iii) The cost vector c has n components; for example $c = [5 \ 3 \ 8]$.
- We want the solution to $Ax=b$ that is $x \geq 0$ and minimizes $c \cdot x$
- Example:
  - Minimize $5x_1 + 3x_2 + 8x_3$ subject to $x_1 + x_2 + 2x_3 = 4$ and $x \geq 0$

$$Ax=b \Rightarrow \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [4], \quad c = \begin{bmatrix} 5 \\ 3 \\ 8 \end{bmatrix}$$

  - The equation $x_1 + x_2 + 2x_3 = 4$ gives a plane in three dimensions. The constraints $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ chops the plane into a triangle. The solution x^* must lie in such triangle, illustrated in fig 8.4.1
  - Inside that triangle, all components of x are positive. On the edges of PQR, one component is zero. At the corners P, Q, R, two components are zero. The optimal solution x^* will be in one of these corners
  - These possible x's are the <u>feasible points</u>, and the triangle is the <u>feasible set</u>
  - The vectors that have zero cost are on the plane $5x_1 + 3x_2 + 8x_3 = 0$. This plane does not meet the triangle. So we increase the cost C until the plane $5x_1 + 3x_2 + 8x_3 = C$ meets the triangle

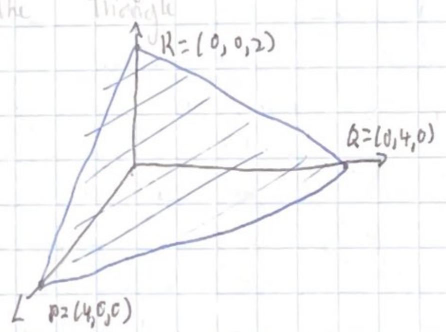


fig 8.4.1
The triangle contains all nonnegative solutions. The lowest cost solution is a corner P, Q, or R

- The first plane $5x_1 + 3x_2 + 8x_3 = C$ to touch the triangle has minimum cost C. The intersection point is the solution x^*. This point must be a corner P, Q, R. A moving plane cannot touch the interior without touching the corners first.
  - Check the cost of each corner

    $P = (4,0,0)$ costs 20 | $Q = (0,4,0)$ costs 12 | $R = (0,0,2)$ costs 16

  - Then $x^* = (0,4,0)$

- Note 1: Some linear programs maximize profit instead of minimizing cost. The mathematics are very similar. We start with a large value of C and decrease it until we have a solution.
- Note 2: It may happen that $Ax=b$ and $x \geq 0$ is impossible to satisfy. Then our feasible set is empty.
- Note 3: It may also occur that the feasible set is unbounded. The requirement $x_1+x_2-2x_3=4$ has P and Q as candidates, but R has moved to infinity
- Note 4: It is possible for the minimum cost to be $-\infty$ if we are dealing with an unbounded feasible set.

The Primal and Dual Problems

- Example:
  - The unknowns x_1, x_2, x_3 are hours of work for a Ph.D., a student, and a machine. The costs per hour are \$5, \$3, and \$8. The hours worked cannot be negative: $x \geq 0$
  - The Ph.D and student can do 1 problem per hour. The machine does 2 per hour. They collaborate on their homework, which has 4 problems $x_1+x_2+2x_3=4$. The cost is $5+3+8=16$ if all of them get in 1 hour. However, the Ph.D. should be put out of work by the student, who works as fast and costs less.
  - Setting x_1 to 0, we get a minimum cost of $2 \times 3+8=14$. However, the best cost is achieved by getting the student to do everything $4 \times 3=12$
- When $Ax=b$ has m equations, corners have m nonzeros. We solve $Ax=b$ for those m variables, with the other $n-m$ free variables set to zero. However, we don't know which m variables to choose. We need to choose all combinations. That is m choose n, or $n!/m!(n-m)!$. If we had $n=20$ and $m=8$, that's 5 billion combinations. Not good.
- The Dual Problem
  - In linear programming, problems come in pairs. The original problem and its dual. The original problem has A and two vectors b and c. The dual problem transposes A and swaps b and c. Maximize $b \cdot y$. Here's the dual to our example:
  - A cheater offers to solve homework problems by selling the answers. The charge is y dollars per problem, or $4y$ altogether ($b \cdot y$). The cheater must be as cheap as the Ph.D, student, and machine: $y \leq 5$, $y \leq 3$, $2y \leq 4$.
  - Maximize $b \cdot y$ subject to $A^Ty \leq c$. The maximum is $y=3$, maximizing profit at \$12, which matches the minimum in the original. This is the duality
- If either problem has a best vector x^* or y^*, then so does the other. Minimum cost $c \cdot x^*$ equals maximum income $b \cdot y^*$. This is the <u>strong duality theorem.</u>
- The cheater's income b^Ty cannot exceed the honest cost
  If $Ax=b$, $x \geq 0$, $A^Ty \leq c$ then $b^Ty = (Ax)^Ty = x^T(A^Ty) \leq x^Tc$
- The full theorem states that b^Ty at its max equals x^Tc at its minimum. The dot product of $x \geq 0$ and $s = c - A^Ty \geq 0$ gives $x^Ts \geq 0$. This is $x^TA^Ty \leq x^Tc$
- Equality needs $x^Ts=0$ so the optimal solution has $x_j^*=0$ or $s_j^*=0$ for each j

The Simplex Method

- The Simplex Method goes from one corner to a neighboring corner of lower cost
  - A corner is a vector $x \geq 0$ that satisfies the m equations $Ax = b$ with at most m positive components. The other $n-m$ components are 0. Those are the free variables, which we set to 0 to solve the m components. All the m component must be non-negative or it is a false corner.
  - A neighboring corner has a zero component of x become positive and one positive component become zero
  - The simplex method chooses which component "leaves" (becomes zero) and which "enters" (becomes positive) such that the total cost is lowered
- Here's the overall algorithm:
  - Look at every zero component of the current corner. If it changes from 0 to 1, the other nonzeroes need to adjust to keep $Ax = b$. Find the new x by back substitution and compute the change in cost $c \cdot x$. The change is the reduced cost "r" of the new component.
  - The <u>entering variable</u> is the one that gives the most negative r.
- Example
  - Suppose the current corner is $P = (4, 0, 0)$ with the Ph.D doing all the work. The current cost is \$20. Now suppose the student does 1 hour of work: $x = (3, 1, 0)$. Then $c \cdot x = \$18$. If the machine works one hour, we get $x = (3, 0, 1)$. Then $c \cdot x = \$18$. Both reduced costs are $r = -2$. The simplex method may choose either as the entering variable.
  - Now we move to selecting the leaving variable, the Ph.D, setting it to 0. Then we get $x = (0, 4, 0)$, a neighboring corner. We repeat again until all reduced costs are positive, meaning we are at the optimal corner, x^*.
- Generally, the Simplex Method takes $O(n)$ steps, but can take $O(2^n)$ in certain cases. A new approach is more complex but takes fewer steps.
- Example
  - Minimize the cost $c \cdot x = 3x_1 + x_2 + 9x_3 + x_4$ with $x \geq 0$ and

$$Ax = b \Rightarrow \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$m = 2$ equations

$n = 4$ unknowns

  - Starting at corner $x = (4, 2, 0, 0)$ which costs $c \cdot x = 14$. It has $m = 2$ nonzeroes and $n - m = 2$ zeroes. Try one unit each of them
    - $x_3 = 1,\ x_4 = 0$, then $x = (2, 1, 1, 0)$ costs 16 $\quad r = 2$
    - $x_4 = 1,\ x_3 = 0$, then $x = (3, 3, 0, 1)$ costs 13 $\quad r = -1$ Entering
  - x_1 is leaving, giving us $x = (0, 6, 0, 4)$. Trying more steps

$$\left.\begin{array}{l} x_1 = 1,\ x_3 = 0, \text{ then } x = (1, 5, 0, 3) \text{ costs } 11 \\ x_3 = 1,\ x_1 = 0, \text{ then } x = (0, 3, 1, 2) \text{ costs } 14 \end{array}\right\} \text{both } r\text{'s are positive} \Rightarrow x^* = (0, 6, 0, 4)$$

Interior Point Methods

- Interior point methods inside the feasible set, where $x > 0$, hoping to reach x^*
- To stay inside, we put a barrier at the boundary. Add an extra cost as a logarithm that blows up when any variable x_j touches 0.
  The number θ is a small parameter that we move towards zero

Barrier Problem: Minimize $c^T x - \theta(\log x_1 + \dots + \log x_n)$ with $Ax = b$

- The cost is non-linear. The constraints $x_j \geq 0$ are not needed.
- The barrier gives an approximate problem for θ. The m constraints $Ax = b$ have Lagrange multipliers $y_1, \dots, y_m$. This is the good way to deal with constraints

y from Lagrange: $L(x, y, \theta) = c^T x - \theta(\sum \log x_i) - y^T(Ax - b)$

- $\partial L / \partial y = 0$ brings back $Ax = b$.

Optimality in barrier prbm $\frac{\partial L}{\partial x_j} = c_j - \frac{\theta}{x_j} - (A^T y)_j = 0$ which is $x_j s_j = \theta$

- The true problem is $x_j s_j = 0$. The barrier problem has $x_j s_j = \theta$. The solutions $x^*(\theta)$ lie on the central path to $x^*(0)$. We can solve the n optimality equations $x_j s_j = \theta$ can be solved iteratively via Newtons Method

Chapter 8.5: Fourier Series: Linear Algebra for Functions

- We will transition to infinite dimensional vector spaces
- What does it mean for a vector to have infinite components?
  - The vector becomes $v = (v_1, v_2, \dots)$
  - The vector becomes a function $f(x)$
- We will go both ways, and they will be connected by the idea of a Fourier Series
- Dot products would be $v \cdot w = v_1 w_1 + v_2 w_2 + \dots$, an infinite series.
  - That poses the question, does $v \cdot w$ converge.
- We include all vectors $v = (v_1, v_2, v_3, \dots)$ in our infinite-dimensional Hilbert Space if and only if its length $\|v\|$ is finite:
  $\|v\|^2 = v \cdot v = v_1^2 + v_2^2 + v_3^2 + \dots$ must add to a finite number
- Example:
  - The vector $v = (1, \frac{1}{2}, \frac{1}{4}, \dots)$ is included in the Hilbert Space. Its length is
    $\sqrt{v_1^2 + v_2^2 + \dots} = \sqrt{\frac{1}{1 - \frac{1}{4}}} = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{3}$
- If v and w have finite length, how large can their dot product be?
  - The sum $v \cdot w = v_1 w_1 + v_2 w_2 + \dots$ also adds to a finite number. The Swartz inequality still holds

    Swartz inequality $|v \cdot w| \leq \|v\| \|w\|$
  - Even in infinite dimensional space, the ratio of $v \cdot w$ to $\|v\|\|w\|$ is $\cos\theta$

- Now change to functions: The space of functions defined for $0 \le x \le 2\pi$: It must somehow be bigger than $\mathbb{R}^n$.
- We redefine the inner product and length as integrals

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx \quad \text{and} \quad \|f\|^2 = \int_0^{2\pi} [f(x)]^2 dx$$

- Example:
  - $f(x) = \sin x$

$$\|f\|^2 = \langle f, f \rangle = \int_0^{2\pi} \sin^2 x dx = \left[\frac{1}{2}\sin x \cos x - \frac{1}{2}\int dx\right]_0^{2\pi} = \left(\frac{1}{2}\sin(2\pi)\cos(2\pi) - \frac{1}{2}(2\pi)\right) - \left(\frac{1}{2}\sin(0)\cos(0) - \frac{1}{2}(0)\right) = \pi$$

Then, $\|f\| = \sqrt{\pi}$

  - Moreover, $\sin x$ and $\cos x$ are orthogonal

~~Inner product is zero~~ $\int_0^{2\pi} \sin x \cos x dx = \frac{1}{2}\int_0^{2\pi} \sin(2x) dx = \frac{1}{4}\left[\cos 2x\right]_0^{2\pi} = 0$

- Every function in the list is orthogonal to the others:
  $\cos(0x)$, $\sin x$, $\cos x$, $\sin 2x$, $\cos 2x$, $\sin 3x$, $\cos 3x$, ...

Fourier Series

- The Fourier Series of a function $y(x)$ is its expansion into sines and cosines

$$y(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

- We have an orthogonal basis! Our functions repeat every 2π.
- Remember, our list is infinite, we avoided $v = (1, 1, \dots)$ because its infinite length, so we avoid functions like $\frac{1}{2} + \cos x + \cos 2x + \cos 3x$ (Note: This is π times the Dirac Delta Function). All points inside $0 < x < 2\pi$ has a finite height, $\int \delta^2(x)dx$ is ∞, so it is not included in the Hilbert space
- Compute the length $f(x)$

$$(f, f) = \int_0^{2\pi} (a_0 + a_1\cos x + b_1 \sin x + a_2 \cos 2x + \dots)^2 dx$$

$$= \int_0^{2\pi} (a_0^2 + a_1^2\cos^2 x + b_1^2 \sin^2 x + a_2^2 \cos^2 2x + \dots) dx$$

$$\|f\|^2 = 2\pi a_0^2 + \pi(a_1^2 + b_1^2 + a_2^2 + \dots)$$

  - The step 1 to 2 uses orthogonality. All products like $\cos x \cos 2x$ integrate to zero. Line 2 contains what's left
  - If we divide each term by its length, we get an orthonormal basis for our function space

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots$$

- We combine the basis with coefficients $A_0, A_1, B_1, A_2, \ldots$ to yield a function $F(x)$. Then the 2π and π's leave the formula

Function length = vector length $\|F\|^2 = \langle F, F\rangle = A_0^2 + A_1^2 + B_1^2 + A_2^2$

- The function has finite length exactly when the vector of coefficients has finite length.
- The Fourier Series connects our function space and infinite dimensional Hilbert Space.
The function is in L^2, the coefficients in ℓ^2.
- The function space contains $f(x)$ exactly when the Hilbert space contains the vector $v = (a_0, a_1, b_1, \ldots)$ of Fourier coefficients. Both $f(x)$ and v have finite length
- Example
  - Suppose $f(x)$ is a square wave:

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x < \pi \\ -1 & \text{if } \pi \le x < 2\pi \end{cases} \quad \text{forever}$$

  - This is an odd function, and so all terms in its Fourier Series are sines

$$f(x) = \frac{4}{\pi}\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \ldots\right]$$

  - The length is $\sqrt{2\pi}$, because at every point $(f(x))^2$ is $(-1)^2$ or $(1)^2$!

$$\|f\|^2 = \int_0^{2\pi} (f(x))^2\,dx = \int_0^{2\pi} dx = 2\pi$$

  - At $x = 0$, the sines are zero and so is the Fourier Series. This is halfway up the jump from -1 to $+1$. Also, at $x = \frac{\pi}{2}$, the square wave equals 1 and the sines alternate sign

Formula for π: $1 = \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots\right) \Rightarrow \pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots\right)$

The Fourier Coefficients

- How do we find the a and b's that multiply the cosines and sines. For a given function $f(x)$, we are asking for its Fourier coefficients:

$$f(x) = a_0 + a_1\cos x + b_1 \sin x + a_2 \cos 2x$$

  - Multiply both sides by $\cos x$ and then integrate from 0 to 2π.
From orthogonality, all integrals on the right side are 0 except $\cos^2 x$.

$$\int_0^{2\pi} f(x)\cos x\,dx = \int_0^{2\pi} a_1 \cos^2 x\,dx = \pi a_1$$

Divide by π to get a_1. We can do this for all a and b.

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx \quad \text{and similarly} \quad b_k = \int_0^{2\pi} f(x)\sin kx\,dx$$

  - The integral of $f(x)\sin kx$ was $4/k$ for odd k in the square wave.
  - The constant term is excluded
  - The constant term a_0 is given by:

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\cdot 1\,dx = \text{average value of } f(x)$$

Compare Linear Algebra in $\mathbb{R}^n$

- Notice how similar the infinite and finite dimensional cases are. We can still use an orthogonal basis even in the infinite case.
- If we had a finite vector space, we construct construct a vector b using the orthogonal basis:

  $b = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$

  - Multiply both sides by v_1^T; orthogonality leads to zeros

  coefficient c_1: $v_1^T b = c_1 v_1^T v_1 + 0 + 0 + \dots + 0$. Therefore $c_1 = v_1^T b / v_1^T v_1$

  $$c_n = \frac{\langle \text{nth basis vector}, b \rangle}{\|\text{nth basis vector}\|^2} = \frac{v_n^T b}{v_n^T v_n} = \frac{\int \cos nx \, f(x)\, dx}{\pi}$$, where $\langle \cdot, \cdot \rangle$ is the inner product

Chapter 8.6: Linear Algebra for Statistics and Probability

- Data tends to go in rectangular matrices so we'd expect a lot of A^TA
- This chapter will move to beyond least squares.

Weighted Least Squares

- In least squares, some measurements will be more reliable or important than others, and least squares must reflect that.
- To include weights in the m equations $Ax=b$, multiply each equation i by a weight w_i. We replace $Ax=b$ with $WAx = Wb$ and we use least squares just as before.

  $A^TA\hat{x} = A^Tb \Rightarrow (WA)^T(WA)\hat{x} = (WA)^T Wb$. Let $C = W^TW$. Then:

  weighted least squares: $C = W^TW$ is in the n equations for $\hat{x}$ $A^TCA\hat{x} = A^TCb$

- When $n=1$ and A = column of 1's, $\hat{x}$ changes from an average to a weighted average.

  Simplest case $\hat{x} = \frac{b_1 + \dots + b_m}{m}$ changes to $\hat{x}_W = \frac{w_1^2 b_1 + \dots + w_m^2 b_m}{w_1^2 + \dots + w_m^2}$

- How do we choose the weights w_i? That depends on the reliability of b_i. If that observation has variance σ_i^2, then the root mean square error in b_i is σ_i. When we divide the equations by $\sigma_i, \dots, \sigma_m$, all variance will equal 1. So the weight is $w_i = 1/\sigma_i$ and the diagonal of $C = W^TW$ contains the numbers $1/\sigma_i^2$

  The statistically correct matrix is $C = \text{diag}(1/\sigma_1^2, \dots, 1/\sigma_m^2)$

  - This is correct provided the errors e_i and e_j in different equations are statistically independent. If they are dependent, off-diagonal entries show up in the covariance matrix Σ_1. The good choice is still $C = \Sigma_1^{-1}$.

Mean and Variance

- Two crucial numbers for a random variable are its mean and variance m and σ^2. The expected value $E[e]$ is found from the probabilities $p_1, p_2, \dots$ of the possible errors $e_1, e_2, \dots$ (and the variance σ^2 is always measured from the mean)
- For a discrete random variable, the error e_j has probability p_j (the p_j add to 1)

  Mean $m = E[e] = \sum_1^{-1} e_j p_j$ variable $\sigma^2 = E[(e-m)^2] = \sum_1^{-1} (e_j - m^2) p_j$

- Example
  - Flip a coin. The result is 0 or 1 (for heads or tails). These events both have $p_0 = p_1 = \frac{1}{2}$.

    Mean $= (0)\frac{1}{2} + (1)\frac{1}{2} = \frac{1}{2}$ Variance $= (0-\frac{1}{2})^2(\frac{1}{2}) + (1-\frac{1}{2})^2(\frac{1}{2}) = \frac{1}{4}$
- Example:
  - Flip the coin N times and count heads. With 3 flips, we see $M = 0, 1, 2,$ or 3 heads. The chances are 1/8, 3/8, 3/8, 1/8
  - For all N, the number of ways to see M heads is N choose M. Divide that by the total 2^N possible outcomes to get the probability for each M.
  - The variance is $\sigma^2 = \frac{N}{4}$

The Covariance Matrix

- If we run m different experiments at once, they could be independent or there might be some correlation between them. Each measurement b is now a vector with m components. These components are the output b_i from the m experiments.
- If we measure distances from the means m_i, each error $e_i = b_i - m_i$ has mean zero. If two errors e_i and e_j are independent, their product $e_i e_j$ also has mean 0. However, if the measurements are at the same time by the same observer, e_i and e_j tend to have the same sign or size. The errors could be correlated. The products $e_i e_j$ are weighted by p_{ij} (their probability); covariance $\sigma_{ij} = \sum \sum p_{ij} e_i e_j$. The sum of $e_i^2 p_{ii}$ is the variance σ_i^2.

  Covariance $\sigma_{ij} = \sigma_{ji} = E[e_i e_j]$ = expected value of $e_i e_j$

  This is the ijth and jith entry of the covariance matrix Σ. The iith entry is σ_i^2

Principle Component Analysis

- Start by measuring m properties of n samples (e.g. grades in m courses for n students). From each row, subtract its average so the sample means are zeros. We look for a combination of courses and/or a combination of students for which the data provides the most information.
- Information is "distance" from randomness and is measured by variance. A large variance in course grades means greater information than a small variance.
- The key matrix idea is the SVD $A = U\Sigma V^T$. The singular values in the diagonal matrix Σ are in decreasing order and σ_1 is the most important. Weighting the m courses by the components of u_1 gives a "master course" or "eigencourse" with the most significant grades
- Example
  - Suppose the grades A, B, C, F are worth 4, 2, 0, and -6 points. If each course and each student has one of each grade, then all means are 0

$$\begin{bmatrix} -6 & 2 & 0 & 4 \\ 0 & 4 & -6 & 2 \\ 4 & 0 & 2 & -6 \\ 2 & -6 & 4 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & & \\ & 2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$\sigma = 12, 8, 4$ written as 3, 2, 1 to keep all entries as integers

- Weighting the rows (the courses) by $u_1 = \frac{1}{2}(-1,-1,1,1)$ will give the eigencourse.
  Weighting the columns (the students) by $v_1 = \frac{1}{2}(1,-1,1,-1)$ gives the eigenstudent

Chapter 8.7: Computer Graphics

- Computer graphics deal with 4 transformations to move objects around in 3D space. These objects are projected onto 2D space to make images. These transformation matrices are 4 by 4 because they also must translate objects, which is not possible with 3D linear transformations.
- These transformations are:
  - Translation: Shift the origin to another point $P_0 = (x_0, y_0, z_0)$
  - Rescaling: By c in all directions or by different factors c_1, c_2, c_3
  - Rotation: around an axis through the origin or through P_0
  - Projection: onto a plane through the origin or P_0
- We change the origin's coordinates to $(0,0,0,1)$. The homogenous coordinates of (x,y,z) are $(x,y,z,1)$
- Translation
  - Shift the whole three-dimensional space along the vector v_0. The origin moves to (x_0, y_0, z_0). This vector v_0 is added to every point v in $\mathbb{R}^3$. Using homogenous coordinates, the 4 by 4 matrix T shifts the whole space by v_0:

Translation Matrix $T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_0 & y_0 & z_0 & 1 \end{bmatrix}$

  - Computer graphics work with row vector times matrix instead of matrix times column vector
  - With the translation matrix, we have $[0\ 0\ 0\ 1]T = [x_0\ y_0\ z_0\ 1]$
  - To translate a row vector v, do $[v\ 1]T = [v_{new}\ 1]$
- Scaling
  - To scale, we multiply the homogenous coordinate by S

Scaling Matrix $S = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

  - If instead, we have cI (4 by 4), multiplying a point gives (cx, cy, cz, c), which represents the same point as $(x,y,z,1)$. The special property of homogenous coordinates is that multiplying cI does not move the point
  - We can scale by different factors in each direction with

Scaling Matrix $S = \begin{bmatrix} c_x & 0 & 0 & 0 \\ 0 & c_y & 0 & 0 \\ 0 & 0 & c_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

- The point (x,y,z) in $\mathbb{R}^3$ has homogeneous coordinates $(x,y,z,1)$ in $\mathbb{P}^3$. This "projective space" is not the same as $\mathbb{R}^4$. It is still three-dimensional. To achieve such a thing, we define (cx, cy, cz, c) as the same point as $(x,y,z,1)$. Those points of projective space are really lines through the origin in $\mathbb{R}^4$.
- Computer graphics use affine transformations (linear + shift). An affine transformation T is executed on $\mathbb{P}^3$ by a 4 by 4 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix} = \begin{bmatrix} T(1,0,0) & 0 \\ T(0,1,0) & 0 \\ T(0,0,1) & 0 \\ T(0,0,0) & 1 \end{bmatrix}$$

shows translation

- Rotation
  - A rotation in $\mathbb{R}^2$ and $\mathbb{R}^3$ is achieved by an orthogonal matrix Q. The determinant is 1.

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ becomes } R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  - The above matrix rotates around the origin. How do we rotate around an arbitrary point like $(4,5)$. We translate $(4,5)$ to $(0,0)$, rotate by θ, then translate $(0,0)$ back to $(4,5)$

$$vT_-RT_+ = [x\ y\ 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -5 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 5 & 1 \end{bmatrix}$$

  - The center of rotation $(4,5,1)$ moves to $(0,0,1)$. Rotation doesn't change it. Then it moves back to $(4,5,1)$
  - In $\mathbb{R}^3$, we get

$$R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

leaves z-axis alone
leaves origin alone

  - The rotation matrix around a unit vector $a = (a_1, a_2, a_3)$

$$Q = (\cos\theta)I + (1-\cos\theta) \begin{bmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{bmatrix} - \sin\theta \begin{bmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} & Q & & 0 \\ & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Projection
  - A plane through the origin is a vector space. Other planes are affine spaces, aka "flats". We want to project three-dimensional vectors onto planes. Start with a plane through the origin, whose normal vector is n. The vectors in the plane satisfy $n^T v = 0$. The usual projection onto the plane is $I - nn^T$
  - In homogeneous coordinates, this becomes 4 by 4

$$\text{projection onto the plane } n^T v = 0 \quad P = \left[\begin{array}{c|c} I - nn^T & \begin{matrix}0\\0\\0\end{matrix} \\ \hline 0\ 0\ 0 & 1 \end{array}\right]$$

  - Now, project onto a flat $n^T(v - v_0) = 0$, with v_0 as a point on the plane. First we translate v_0 to the origin by T_-, project it along the n direction, and translate back along the row vector v_0:

$$\text{Projection onto a flat} \quad T_- P T_+ = \begin{bmatrix} I & 0 \\ -v_0 & 1 \end{bmatrix} \begin{bmatrix} I - nn^T & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ v_0 & 1 \end{bmatrix}$$

- Reflection
  - A reflection simply moves a point in the direction of a projection, but twice as far
  - Instead of $I - nn^T$, we have $I - 2nn^T$

$$\text{Reflection across a flat} \quad T_- R T_+ = \begin{bmatrix} I & 0 \\ -v_0 & 1 \end{bmatrix} \begin{bmatrix} I - 2nn^T & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ v_0 & 1 \end{bmatrix}$$

- The projection matrix P gives a parallel projection. All points move parallel to n until they reach the plane. Another choice in computer graphics is a perspective projection, in which objects get smaller with distance.

Chapter 10: Complex Vectors and Matrices

Chapter 10.1: Complex Numbers

- $i = \sqrt{-1}$
- $\mathrm{Re}(a+bi) = a$, $\mathrm{Im}(a+bi) = b$
- $\overline{z_1}\,\overline{z_2} = \overline{z_1 z_2}$
- $\overline{z_1} + \overline{z_2} = \overline{z_1 + z_2}$
- If $Ax = \lambda x$ and A is real, then $A\bar{x} = \bar{\lambda}\bar{x}$
- The number $z = a + bi$ is also $z = r\cos\theta + ir\sin\theta = re^{i\theta}$
- Powers and multiplication is easy with complex numbers
- $z^n = r^n(\cos n\theta + i\sin n\theta)$
- Set $w = e^{2\pi i/n}$. The nth powers of $1, w, w^2, \ldots, w^{n-1}$ all equal 1

Chapter 10.2: Hermitian and Unitary Matrices

- We introduce a new matrix: a conjugate transpose. Transpose the matrix then take the conjugate of each entry.
- One reason to go $\bar{Z}$ is that the length^2 of a real vector is $x_1^2 + x_2^2 + \ldots$, but this is not the case for a complex vector. $|z| \neq z_1^2 + z_2^2 + \ldots + z_n^2$. Then the vector $\begin{bmatrix} 1 \\ i \end{bmatrix}$ would have length 0 — not good. Instead of $(a+bi)^2$, we want $a^2 + b^2$, the modulo squared. This is $(a+bi)(a-bi)$
- For each component, we want z_j times $\bar{z}_j$, which is $|z_j|^2 = a_j^2 + b_j^2$. That comes from $\bar{z}^T z$.

Length squared $[\bar{z}_1 \ldots \bar{z}_n]\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = |z_1|^2 + \ldots + |z_n|^2$. This is $\bar{z}^T z = \|z\|^2$

- Now, $(1, i)$ has length^2 $1^2 + (i)(-i) = 2$. Then it has a length $\sqrt{2}$ and the zero vector is the only vector with length 0
- $\bar{Z}^T = Z^H$. $\square^H$ denotes the conjugate transpose of vectors and matrices. H for Hermitian ("Hermeeshan")
- Example

$$A = \begin{bmatrix} 1 & i \\ 0 & 1+i \end{bmatrix} \Rightarrow A^H = \begin{bmatrix} 1 & 0 \\ -i & 1-i \end{bmatrix}$$

Complex Inner Products

- For real vectors, $\|v\|^2 = v \cdot v = v^T v$. We want this property to hold. That is, $z \cdot z = \|z\|^2$. To make that happen, instead of $v^T v$, we do $z^H z$, and we define the inner product as such.
- For real or complex vectors u and v

$$\langle u, v \rangle = u \cdot v = u^H v = [\bar{u}_1 \ldots \bar{u}_n]\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \bar{u}_1 v_1 + \ldots + \bar{u}_n v_n$$

- With complex vectors, $u^H v$ is not necessarily $v^H u$. In fact, $v^H u = \bar{v}_1 u_1 + \ldots + \bar{v}_n u_n$ is the complex conjugate of $u^H v$
- Example
  - Given $u = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $v = \begin{bmatrix} i \\ 1 \end{bmatrix}$, $u^H v = [1 \ -i]\begin{bmatrix} i \\ 1 \end{bmatrix} = 0$. The vectors u and v are orthogonal
  - The inner product of Au with v equals the inner product of u with $A^H v$: A^H = "adjoint" of A $(Au)^H v = u^H (A^H v)$
- The rule for the transpose of products remain: $(AB)^H = B^H A^H$

Hermitian Matrices

- Hermitian matrices are complex square matrices equal to their conjugate transpose. $A = A^H$; $a_{ij} = \overline{a_{ji}}$. Every real symmetric matrix is Hermitian.
- Example:

$$A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$$ is Hermitian.

- If $A = A^H$ and z is any vector, the number $z^H A z$ is real
  - Proof: $z^H A z$ is 1 by 1 (i.e. a number)
    $(z^H A z)^H = z^H A^H (z^H)^H = z^H A z$
  - $z^H A z$ equals its own conjugate transpose and is 1 by 1, therefore it must be real
  - This number $z^H A z$ often represents energy
  - with the previous matrix

$$[\bar{z}_1 \ \bar{z}_2]\begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underbrace{\overbrace{2\bar{z}_1 z_1}^{\text{real}} + \overbrace{5\bar{z}_2 z_2}^{\text{real}}}_{\text{diagonal}} + \underbrace{\overbrace{(3-3i)\bar{z}_1 z_2 + (3+3i) z_1 \bar{z}_2}^{\text{conjugates} \therefore \text{real}}}_{\text{off diagonal}}$$

- Every eigenvalue of a Hermitian matrix is real
  - Proof: suppose $Az = \lambda z$. Multiply both sides by z^H to get $z^H A z = \lambda z^H z$.
    $z^H A z$ is real and so is $z^H z$. Therefore, λ is also real
- The eigenvectors of a Hermitian Matrix can be chosen orthonormal (when they correspond to different eigenvalues). If $Az = \lambda z$ and $Ay = \beta y$ then $y^H z = 0$
  - Proof: Multiply $Az = \lambda z$ on the left by y^H and multiply $y^H A^H = \beta y^H$ on the right by z

    $y^H A z = \lambda y^H z$ and $y^H A^H z = \beta y^H z$

  - The left sides are equal because $A = A^H$. Therefore the right sides are equal. Since $\lambda \neq \beta$, the factor $y^H z$ must be 0 and the eigenvectors y and z are orthogonal
- The eigenvector matrix can be chosen orthonormal, with complex entries. We call such a matrix Unitary

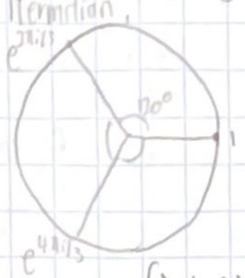
Unitary Matrices

- A Unitary matrix U is a (complex) square matrix that has orthonormal columns. U is the complex equivalent of Q

  Unitary Matrix $\quad U = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}$

- The matrix test for real orthonormal columns is $Q^T Q = I$. When Q^T multiplies Q, the zero inner products appear off diagonal. In the complex case, for this to happen, we use the conjugate transpose $U^H U = I$
- Every matrix U with orthonormal columns has $U^H U = I$
  If U is square, it is a unitary matrix. Then $U^H = U^{-1}$
- Suppose U (with orthonormal columns) multiplies any z. The vector length stays the same, because $z^H U^H U z = z^H z$. If z is an eigenvector of U we learn something more: The eigenvalues of unitary (and orthogonal) matrices all have absolute value $|\lambda| = 1$.
  If U is unitary then $\|Uz\| = \|z\|$. Therefore $Uz = \lambda z$ leads to $|\lambda| = 1$

- Example:
  - The 3 by 3 Fourier Matrix is in fig 10.2.1. Is it Hermitian and or unitary? F is certainly symmetric, but does not equal its conjugate transpose. It is not Hermitian.



Fourier Matrix $F = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \end{bmatrix}$

fig 10.2.1: The cube roots of 1 go into the Fourier Matrix $F = F_3$

  - F is unitary. The squared length of every column is $\frac{1}{3}(1+1+1) = 1$. The first column is orthogonal to the rest: $1 + e^{2\pi i/3} + e^{4\pi i/3} = 0$
  - From the figure on the left, it is really obvious that summing columns 2 and 3 ends up with 0.
  - Column 2 and 3 are orthogonal, so F is unitary

$$F_2^H F_3 = \frac{1}{3}(1\cdot 1 + e^{-2\pi i/3}e^{4\pi i/3} + e^{-4\pi i/3}e^{2\pi i/3}) = \frac{1}{3}(1 + e^{2\pi i/3} + e^{-2\pi i/3}) = 0$$

  - When we multiply by F, we are computing the Discrete Fourier Transform, and multiplying by F^{-1} computes the inverse transform. Since F is unitary and square, $F^{-1} = F^H$

Real vs Complex
- Length: $\|x\|^2 = x_1^2 + \ldots + x_n^2 \longleftrightarrow$ Length: $\|z\|^2 = |z_1|^2 + \ldots |z_n|^2$
- Transpose: $(A^T)_{ij} = A_{ji} \longleftrightarrow$ conjugate transpose: $(A^H)_{ij} = \overline{A_{ji}}$
- Product rule: $(AB)^T = B^TA^T \longleftrightarrow$ product rule: $(AB)^H = B^HA^H$
- Dot product: $x^Ty = x_1y_1 + \ldots + x_ny_n \longleftrightarrow$ inner product: $u^Hv = \bar{u}_1v_1 + \ldots + \bar{u}_nv_n$
- Orthogonality: $x^Ty = 0 \longleftrightarrow u^Hv = 0$
- $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$ (real Λ) $\longleftrightarrow A = U\Lambda U^{-1} = U\Lambda U^H$ (real Λ)
- $(Qx)^T(Qy) = x^Ty$ and $\|Qx\| = \|x\| \longleftrightarrow (Ux)^H(Uy) = x^Hy$ and $\|Uz\| = \|z\|$

Chapter 10.3: The Fast Fourier Transform
- We want to multiply quickly by F and F^{-1}. This is achieved by the Fast Fourier Transform. An ordinary product Fc uses n^2 multiplications (F has n^2 entries). The FFT only needs $\frac{1}{2}n\log_2 n$
- Fourier's idea is to represent f as a sum of harmonics c_ke^{ikx}. The function is seen in frequency space through the coefficients c_k, instead of physical space with values $f(x)$. FFT provides fast passage between c and f

Roots of Unity and the Fourier Matrix

- Equations of degree n have n roots (counting repetition). This is the Fundamental Theorem of Algebra, and to make it true, we need to allow complex roots
- This section focuses on $z^n = 1$. The solutions are the "nth roots of unity". They are n evenly spaced points around the unit circle in the complex plane
- For $z^8 = 1$, we have points spaced at $\frac{1}{8}(360°) = 45°$ degrees. The solutions are numbers $w = e^{i\theta} = e^{i2\pi/8}$, and the powers $w^2, w^3, \ldots w^8$. This is shown in fig 10.3.1

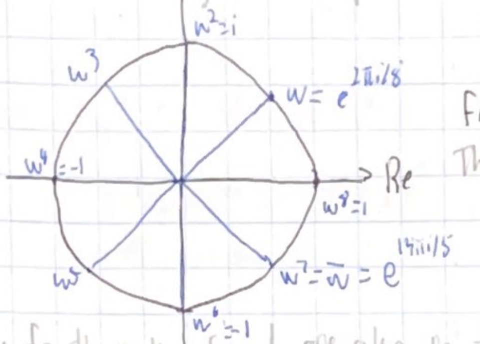


Fig 10.3.1
The eight solutions to $z^8 = 1$

- The fourth roots of 1 are also in this figure. They are $i, -1, -i$, and 1. The angle is now 90°. The idea behind FFT is to go from an 8 by 8 matrix to 4 by 4 to 2 by 2. By exploiting the connections between F_8 and F_4 and F_8 and F_{16}, we can make multiplication by F_{1024} very fast.
- The fourier matrix for $n=4$ is

Fourier Matrix $n=4$
$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 \\ 1 & w^2 & w^4 & w^6 \\ 1 & w^3 & w^6 & w^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix}$$

- $\frac{1}{2}F$ is unitary, so $(\frac{1}{2}F^H)(\frac{1}{2}F) = I$
- The columns of F give $F^H F = 4I$. Its inverse is $\frac{1}{4}F^H$, which is $F^{-1} = \frac{1}{4}\bar{F}$
- The inverse changes from $w = i$ and $\bar{w} = -i$. That takes us from F to $\bar{F}$. The FFT gives a quick way to multiply by F and by F^{-1}.
- The unitary matrix is $U = F/\sqrt{n}$. We avoid the $\sqrt{n}$ and put it outside F^{-1}.
- The main point is to multiply F times the fourier coefficients c_0, c_1, c_2, c_3

4-point Fourier Series
$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} = Fc = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 \\ 1 & w^2 & w^4 & w^6 \\ 1 & w^3 & w^6 & w^9 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

- The input is four complex coefficients c_0, c_1, c_2, c_3. The output is four function values y_0, y_1, y_2, y_3. The first output $y_0 = c_0 + c_1 + c_2 + c_3$ is the value at the Fourier Series at $x = 0$. The second output is the value of the series $\sum c_k e^{ikx}$ at $x = 2\pi/4$. Likewise y_2 and y_3 are the outputs of $\sum c_k e^{ikx}$ at $x = 4\pi/4$ and $6\pi/6$

$$y_1 = c_0 + c_1 e^{i2\pi/4} + c_2 e^{i4\pi/4} + c_3 e^{i6\pi/4} = c_0 + c_1 w + c_2 w^2 + c_3 w^3$$

- These are finite Fourier Series. They contain $n=4$ terms and are evaluated at 4 equally space points in $[0, 2\pi)$

- The period then repeats for the next 2π.
- j and k will be zero-indexed for the rest of the section
- The n by n Fourier Matrix contains powers of $e^{2\pi i/n}$:

$$F_n c = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & w^{n-1} & w^{2(n-1)} & & w^{(n-1)^2} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix} = y \quad (F_n)_{ij} = w^{ij}, \text{ where } i, j$$

- F_n is symmetric but not Hermitian. Its columns are orthogonal and $F_n \overline{F_n} = nI$. Then $F_n^{-1} = \overline{F}_n / n$. The inverse contains powers of $\overline{w}_n = e^{-2\pi i/n}$.
- When we multiple c by F_n, we sum the series at n points. When we multiply y by F_n^{-1}, we find the coefficients c from function values y. The matrix F passes from "frequency space" to "physical space."
- Many authors prefer to use $\omega = e^{-2\pi i/N}$, the complex conjugate of w. Then F becomes $\overline{F}$.
- When a function $f(x)$ has period 2π, and we change x to $e^{i\theta}$, the function is defined around the unit circle (where $z = e^{i\theta}$). Then the DFT from y to c is matching n values of this $f(z)$ by a polynomial $p(z) = c_0 + c_1 z + \dots + c_{n-1} z^{n-1}$
- Interpolation: Find $c_0, \dots, c_{n-1}$ so that $p(z) = f(z)$ at n points $z = 1, \dots, w^{n-1}$
- The Fourier matrix is the Vandermonde Matrix for interpolation at those n points

One Step of the Fast Fourier Transform

- We want to multiply F and c as quickly as possible. Usually this takes n^2 multiplications. Since F follows a special pattern ($(F_n)_{ij} = w^{ij}$), we can factor it in a way that produces many zeroes. This is the FFT
- We connect F_n to two copies of $F_{n/2}$. For example, with F_4

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} \text{ and } \begin{bmatrix} F_2 & \\ & F_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & & \\ 1 & i^2 & & \\ & & 1 & 1 \\ & & 1 & i^2 \end{bmatrix}$$

- We can factorize F_4 as the following:

$$F_4 = \begin{bmatrix} 1 & & 1 & \\ & 1 & & i \\ 1 & & -1 & \\ & 1 & & -i \end{bmatrix} \begin{bmatrix} 1 & 1 & & \\ 1 & i^2 & & \\ & & 1 & 1 \\ & & 1 & i^2 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}$$

  - The last matrix is a permutation. It puts the even c's (c_0 and c_2) ahead of the odd c's (c_1 and c_3). The middle matrix performs half sized matrix transforms F_2 and F_2 on the evens and odds. The left matrix combines the two half-sized outputs in a way that correctly produces $y = F_4 c$.

- This idea applies for larger n!

$$F_{1024} = \begin{bmatrix} I_{512} & D_{512} \\ I_{512} & -D_{512} \end{bmatrix} \begin{bmatrix} F_{512} & \\ & F_{512} \end{bmatrix} \begin{bmatrix} \text{even-odd} \\ \text{Permutation} \end{bmatrix}$$

  - I is the identity matrix and D is the identity matrix with entries $(1, w, \ldots, w^{511})$
- Here are the algebra formulas meaning the same thing:

(FFT) Set $m = \frac{1}{2}n$. The first m and last m components of $y = F_n c$ combine the half-size transforms $y' = F_m c'$ and $y'' = F_m c''$. The matrix equation shows the step from n to $m = n/2$ as $Iy' + Dy''$ and $Iy' - Dy''$:

$$y_j = y'_j + w_n^j y''_j, \quad j = 0, \ldots, m-1$$
$$y_{j+m} = y'_j - w_n^j y''_j, \quad j = 0, \ldots, m-1 \qquad (5)$$

Split c into c' and c'', transform them by F_m into y' and y'', and reconstruct y.

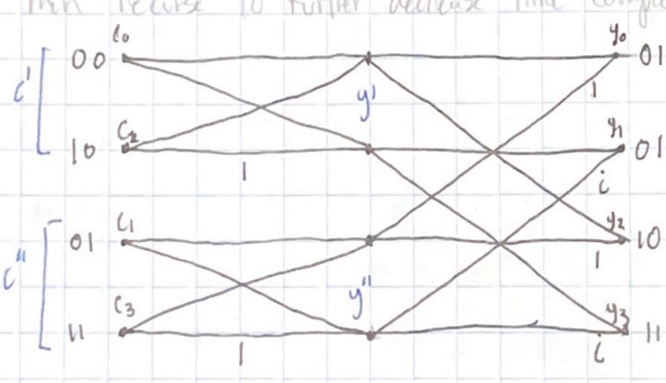
- These formulas come from separating even c_{2k} from odd c_{2k+1}:

$$y_j = \sum_0^{n-1} w^{jk} c_k = \sum_0^{m-1} w^{2jk} c_{2k} + \sum_0^{m-1} w^{j(2k+1)} c_{2k+1} \quad \text{with } m = \tfrac{1}{2}n$$

- The even c's go into $c' = (c_0, c_2, \ldots)$ and the odd c's go into $c'' = (c_1, c_3, \ldots)$. Then comes the transforms $F_m c'$ and $F_m c''$. The key is $w_n^2 = w_m$. This gives $w_n^{2jk} = w_m^{jk}$

Rewrite: $y_j = \sum w_m^{jk} c'_k + (w_n)^j \sum w_m^{jk} c''_k = y'_j + (w_n)^j y''_j$

- For $j \geq m$, the minus sign in (5) comes from factoring out $(w_n)^m = -1$
- The flow graph shows c' and c'' going through the half-size F_2. Those steps are called butterflies. Then the outputs y' and y'' are combined (multiplying y'' by $1, i$, and also by $-1, -i$) to produce $y = F_4 c$.
- This cuts the work almost by half — you see the zeros in the factorization. We then recurse to further decrease time complexity.



The Full FFT by Recursion

- We can further recurse:

$$\begin{bmatrix} F_{512} & \\ & F_{512} \end{bmatrix} = \begin{bmatrix} I & D & & \\ I & -D & & \\ & & I & D \\ & & I & -D \end{bmatrix} \begin{bmatrix} F & & & \\ & F & & \\ & & F & \\ & & & F \end{bmatrix} \begin{bmatrix} \text{pick } 0,4,8 \\ \text{pick } 2,6,10 \\ \text{pick } 1,5,9 \\ \text{pick } 3,7,11 \end{bmatrix}$$

- The total amount of computations for size $n=2^{\ell}$ is reduced from n^2 to $\frac{1}{2}n\ell$. There are ℓ levels, going down from $n=2^{\ell}$ to $n=1$. Each level has $n/2$ multiplications from the diagonal D's to reassemble the half-size outputs from the lower level. This gives our final count $\frac{1}{2}n\log_2 n$
- A rule for the order the c's enter the FFT after all the even-odd permutations. Write the numbers 0 to $n-1$ in base 2. Reverse the digits. The complete picture shows the bit-reversed order at the start, the $\ell=\log_2 n$ steps of the recursion, and the final output $y_0, \dots, y_{n-1}$ which is F_n times c
- The book ends with that fundamental idea, a matrix multiplying a vector. :)

Harry Lu, August 4th 2024, 4:25pm