Chapter 10 : Finite-State Markov Chains

10.2 : The Steady-State Vector and Page Rank

Topics and Objectives

Topics

1. Review of Markov chains
2. Theorem describing the steady state of a Markov chain
3. Applying Markov chains to model website usage.
4. Calculating the PageRank of a web.

Learning Objectives

1. Determine whether a stochastic matrix is regular.
2. Apply matrix powers and theorems to characterize the long-term behaviour of a Markov chain.
3. Construct a transition matrix, a Markov Chain, and a Google Matrix for a given web, and compute the PageRank of the web.

Where is Chapter 10?

- The material for this part of the course is covered in Section 10.2
- Chapter 10 is not included in the **print** version of the book, but it is in the **on-line version**.
- If you read 10.2, and I recommend that you do, you will find that it requires an understanding of 10.1.
- You are not required to understand the material in 10.1.
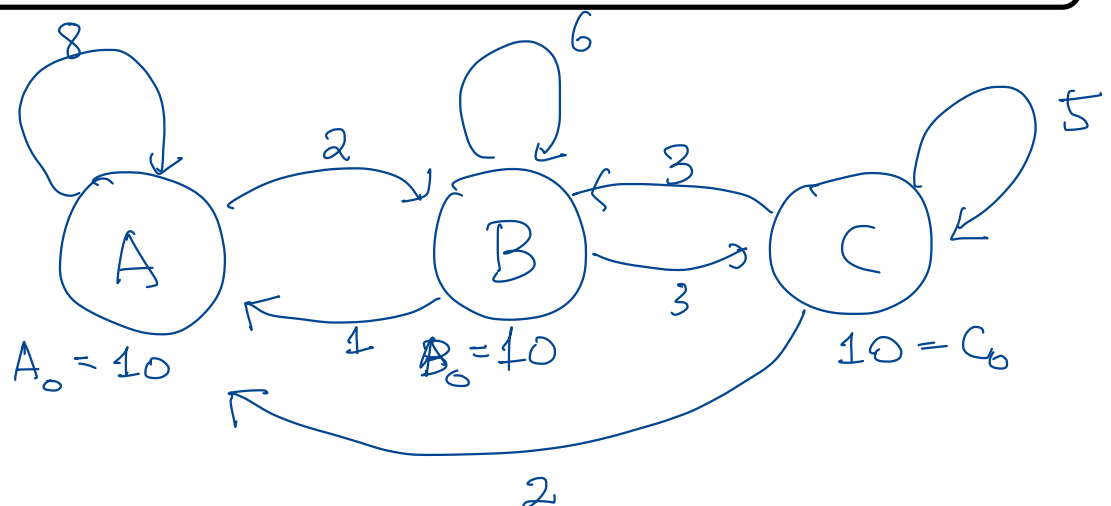
Steady State Vectors

Recall the car rental problem from our Section 4.9 lecture.

Problem

A car rental company has 3 rental locations, A, B, and C.

		rented from		
		A	B	C
	A	.8	.1	.2
returned to	B	.2	.6	.3
	C	.0	.3	.5

There are 10 cars at each location today, what happens to the distribution of cars after a long time?

$$A_1 = 0.8 \cdot A_0 + 0.1 \, B_0 + 0.2 \cdot C_0$$

$$B_1 = 0.2 \, A_0 + 0.6 \cdot B_0 + 0.3 \, C_0$$

$$C_1 = 0 \cdot A_0 + 0.3 \, B_0 + 0.5 \, C_0$$

$$\begin{bmatrix} A_1 \\ B_1 \\ C_1 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 & 0.2 \\ 0.2 & 0.6 & 0.3 \\ 0 & 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} A_0 \\ B_0 \\ C_0 \end{bmatrix}$$

$$P$$

Long Term Behaviour

Can use the transition matrix, P, to find the distribution of cars after 1 week:

$$\vec{x}_1 = P\vec{x}_0$$

The distribution of cars after 2 weeks is:

$$\vec{x}_2 = P\vec{x}_1 = PP\vec{x}_0 = P^2 \cdot \vec{x}_0$$

The distribution of cars after n weeks is:

$$\vec{x}_n = P \cdot \vec{x}_{n-1} = P \cdot P \cdot \vec{x}_{n-2} = \cdots = P^n \cdot \vec{x}_0$$

$$\lim_{n \to \infty} \vec{x}_n \qquad , \qquad \lim_{n \to \infty} P^n$$

Probability vector

$\downarrow$ P : transition matrix

$\begin{bmatrix} 0.8 & 0.1 & 0.2 \\ 0.2 & 0.6 & 0.3 \\ 0 & 0.3 & 0.5 \end{bmatrix}$

$\left\{ \text{Sum 1} \atop \geq 0 \right\}$

" Stochastic matrix

P is regular if
there exists an integer k
such that

all entries of P^k $\underset{\text{(after } k \text{ steps)}}{}$ is
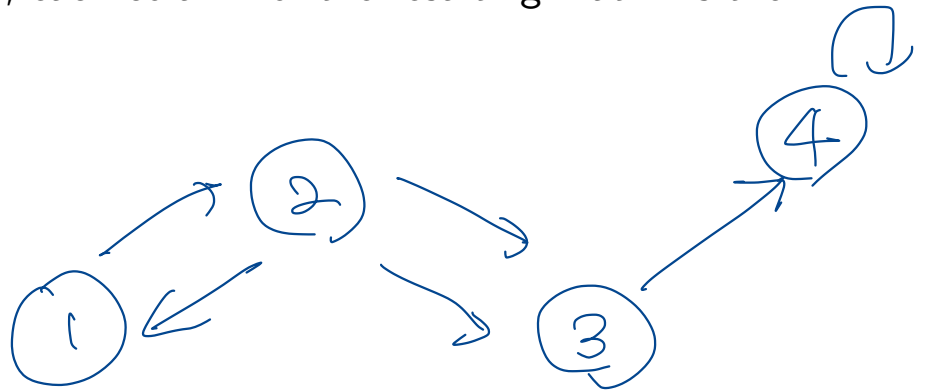
strictly positive.

positive prob.
from any·
to any place

Long Term Behaviour

To investigate the long-term behaviour of a system that has a regular
transition matrix P, we could:

1. compute the **steady-state vector**, $\vec{q}$, by solving $\vec{q} = P\vec{q}$.

2. compute $P^n \vec{x}_0$ for large n.

3. compute P^n for large n, each column of the resulting matrix is the
   steady-state

$$\lim_{n \to \infty} P^n = \lim_{n \to \infty} P^{n+1} = \lim_{n \to \infty} P^n \cdot [\vec{x_1} \quad \vec{x_2} \quad \cdots \quad \vec{x_m}]$$

$$= \lim_{n \to \infty} [P^n \cdot \vec{x_1} \quad P^n \cdot \vec{x_2} \quad \cdots \quad P^n \cdot \vec{x_m}]$$

$$= [\vec{q} \quad \vec{q} \quad \cdots \quad \vec{q}] = \Pi$$

Theorem 1

If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

1. There is a stochastic matrix Π such that

$$\lim_{n \to \infty} P^n = \Pi$$

2. Each column of Π is the same probability vector $\vec{q}$.

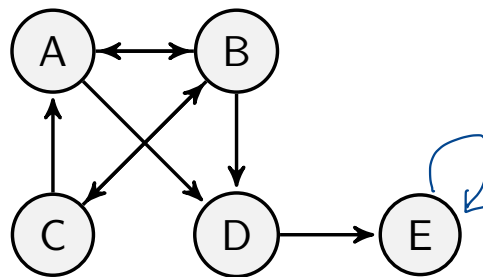3. For any initial probability vector $\vec{x}_0$,

   does not dep. on $\vec{x}_0$

$$\lim_{n \to \infty} P^n \vec{x}_0 = \vec{q}$$

   probability   *steady state*

4. P has a unique eigenvector, $\vec{q}$, which has eigenvalue $\lambda = 1$.

5. The eigenvalues of P satisfy $|\lambda| \leq 1$.

We will apply this theorem when solving PageRank problems.

Example 1

A set of web pages link to each other according to this diagram.

Page A has links to pages __B, D__ .

Page B has links to pages __A, C D__ .

We make two assumptions:

a) A user on a page in this web is equally likely to go to any of the pages that their page links to.

b) If a user is on a page that does not link to other pages, the user stays at that page.

Use these assumptions to construct a Markov chain that represents how users navigate the above web.

From A B C D E

$$
P = \begin{pmatrix}
0 & \frac{1}{3} & \frac{1}{2} & 0 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
0 & \frac{1}{3} & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1
\end{pmatrix}
\begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}
$$

Not regular.

⇕

steady state prob. vector is NOT unique.

Solution

Use the assumptions on the previous slide to construct a Markov chain that represents how users navigate the web.
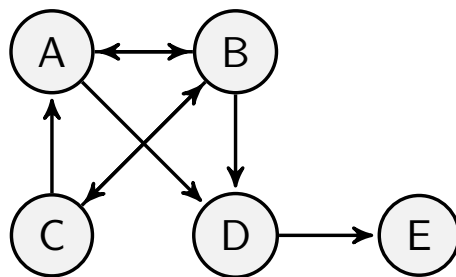
Transition Matrix, Importance, and PageRank

- The square matrix we constructed in the previous example is a **transition matrix**. It describes how users transition between pages in the web.

- The steady-state vector, $\vec{q}$, for the Markov-chain, can characterize the long-term behavior of users in a given web.

- If $\vec{q}$ is unique, the **importance** of a page in a web is given by its corresponding entry in $\vec{q}$.

- The **PageRank** is the ranking assigned to each page based on its importance. The highest ranked page has PageRank 1, the second PageRank 2, and so on.

- Two pages with same importance receive the same PageRank (some other method would be needed to resolve ties)

Is the transition matrix in Example 1 a regular matrix?

$$P^* = \begin{pmatrix} 0 & 1/3 & 1/2 & 0 & 1/5 \\ 1/2 & 0 & 1/2 & 0 & 1/5 \\ 0 & 1/3 & 0 & 0 & 1/5 \\ 1/2 & 1/3 & 0 & 0 & 1/5 \\ 0 & 0 & 0 & 1 & 1/5 \end{pmatrix}$$

regular.
⇓
unique $\vec{q}$

Adjustment 1

Adjustment 1

If a user reaches a page that does not link to other pages, the user will choose any page in the web, with equal probability, and move to that page.

Let's denote this modified transition matrix as P_*. Our transition matrix in Example 1 becomes:

Adjustment 2

> ### Adjustment 2
>
> A user at any page will navigate to any page among those that their page links to with equal probability p, and to any page in the web with equal probability $1 - p$. The transition matrix becomes
>
> $$G = pP_* + (1-p)K$$
>
> All the elements of the $n \times n$ matrix K are equal to $1/n$.

p is referred to as the **damping factor**, Google is said to use $p = 0.85$.

With adjustments 1 and 2, our the Google matrix is:

Computing Page Rank

- Because G is stochastic, for any initial probability vector $\vec{x}_0$,

$$\lim_{n \to \infty} G^n \vec{x}_0 = \vec{q}$$

- We can obtain steady-state evaluating $G^n \vec{x}_0$ for large n, by solving $G\vec{q} = \vec{q}$, or by evaluating $\vec{x}_n = G\vec{x}_{n-1}$ for large n.
- Elements of the steady-state vector give the importance of each page in the web, which can be used to determine PageRank.
- Largest element in steady-state vector corresponds to page with PageRank 1, second largest with PageRank 2, and so on.

On an exam,

- problems that require a calculator will not be on your exam
- you may construct your G matrix using factions instead of decimal expansions

Computing Page Rank

(unnamed)

```
In[1]:= P := {{0, 1/2, 1/2, 0, 1/5}, {1/2, 0, 1/2, 1/2, 1/5}, {0, 1/2, 0, 0, 1/5}, {1/2, 0, 0, 0, 1/5}, {0, 0, 0, 1/2, 1/5}}

In[2]:= K := {{1/5, 1/5, 1/5, 1/5, 1/5}, {1/5, 1/5, 1/5, 1/5, 1/5}, {1/5, 1/5, 1/5, 1/5, 1/5}, {1/5, 1/5, 1/5, 1/5, 1/5}, {1/5, 1/5, 1/5, 1/5, 1/5}}

In[3]:= G[p_] := p * P + (1 - p) * K

In[14]:= MatrixForm[MatrixPower[ G[0.85], 1]]
```

$$Out[14]//MatrixForm= \begin{pmatrix} 0.03 & 0.455 & 0.455 & 0.03 & 0.2 \\ 0.455 & 0.03 & 0.455 & 0.455 & 0.2 \\ 0.03 & 0.455 & 0.03 & 0.03 & 0.2 \\ 0.455 & 0.03 & 0.03 & 0.03 & 0.2 \\ 0.03 & 0.03 & 0.03 & 0.455 & 0.2 \end{pmatrix}$$

```
In[15]:= MatrixForm[MatrixPower[ G[0.85], 2]]
```

$$Out[15]//MatrixForm= \begin{pmatrix} 0.241225 & 0.241225 & 0.241225 & 0.313475 & 0.234 \\ 0.253975 & 0.4346 & 0.253975 & 0.1456 & 0.319 \\ 0.228475 & 0.04785 & 0.228475 & 0.300725 & 0.149 \\ 0.04785 & 0.228475 & 0.228475 & 0.1201 & 0.149 \\ 0.228475 & 0.04785 & 0.04785 & 0.1201 & 0.149 \end{pmatrix}$$

```
In[16]:= MatrixForm[MatrixPower[ G[0.85], 4]]
```

$$Out[16]//MatrixForm= \begin{pmatrix} 0.243031 & 0.257387 & 0.257387 & 0.249035 & 0.250914 \\ 0.30952 & 0.310825 & 0.2782 & 0.275068 & 0.305135 \\ 0.1679 & 0.16268 & 0.195305 & 0.201308 & 0.169779 \\ 0.16156 & 0.15634 & 0.15634 & 0.149293 & 0.158219 \\ 0.117989 & 0.112769 & 0.112769 & 0.125297 & 0.115953 \end{pmatrix}$$

```
In[17]:= MatrixForm[MatrixPower[ G[0.85], 8]]
```

$$Out[17]//MatrixForm= \begin{pmatrix} 0.251785 & 0.251656 & 0.251656 & 0.251754 & 0.251713 \\ 0.298582 & 0.29895 & 0.297885 & 0.297881 & 0.298641 \\ 0.176505 & 0.176171 & 0.177235 & 0.177204 & 0.176464 \\ 0.156692 & 0.156794 & 0.156794 & 0.156823 & 0.156753 \\ 0.116437 & 0.11643 & 0.11643 & 0.116338 & 0.11643 \end{pmatrix}$$

```
In[18]:= MatrixForm[MatrixPower[ G[0.85], 10]]
```

$$Out[18]//MatrixForm= \begin{pmatrix} 0.251705 & 0.251712 & 0.251712 & 0.251715 & 0.251709 \\ 0.298496 & 0.298551 & 0.298359 & 0.298349 & 0.2985 \\ 0.176611 & 0.176552 & 0.176745 & 0.176755 & 0.176605 \\ 0.156761 & 0.156774 & 0.156774 & 0.15676 & 0.156768 \\ 0.116427 & 0.116411 & 0.116411 & 0.116421 & 0.116418 \end{pmatrix}$$

$G^{10} =$

(handwritten annotations: PageRank 2, PageRank 1, 3rd, 4th, 5th; column labels $\vec{\alpha}$, $\vec{\alpha}$, $\vec{\alpha}$, $\vec{\alpha}$, $\vec{\alpha}$)

There is (of course) Much More to PageRank

The PageRank Algorithm currently used by Google is under constant development, and tailored to individual users.

- When PageRank was devised, in 1996, Yahoo! used humans to provide a "index for the Internet," which was 10 million pages.
- The PageRank algorithm was produced as a competing method. The patent was awarded to Stanford University, and exclusively licensed to the newly formed Google corporation.
- Brin and Page combined the PageRank algorithm with a webcrawler to provide regular updates to the transition matrix for the web.
- The explosive growth of the web soon overwhelmed human based approaches to searching the internet.

WolframAlpha and MATLAB/Octave Syntax

Suppose we want to compute

$$\begin{pmatrix} .8 & .1 & .2 \\ .2 & .6 & .3 \\ .0 & .3 & .5 \end{pmatrix}^{10}$$

- At wolframalpha.com, we can use the syntax:

  `MatrixPower[{{.8,.1,.2},{.2,.6,.3},{.0,.3,.5}},10]`
- In MATLAB, we can use the syntax

  `[.8 .1 .2 ;.2 .6 .3;.0 .3 .5]^10`
- Octave uses the same syntax as MATLAB, and there are several free, online, Octave compilers. For example: https://octave-online.net.

You will need to compute a few matrix powers in your homework, and in your future courses, depending on what courses you end up taking.

Example 2 (if time permits)

Construct the Google Matrix for the web below. Which page do you think will have the highest PageRank? How would your result depend on the damping factor p? Use software to explore these questions.
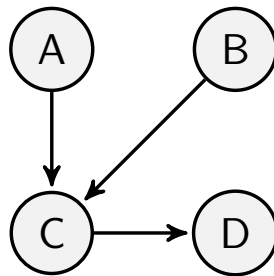
Section 7.1 : Diagonalization of Symmetric Matrices

Chapter 7: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Symmetric matrices
2. Orthogonal diagonalization

Learning Objectives

1. Construct an orthogonal diagonalization of a symmetric matrix, $A = PDP^T$.

Symmetric Matrices

Definition | square

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

Example. Which of the following matrices are symmetric? Symbols $*$ and $\star$ represent real numbers.

$$A = [*] \in \mathbb{R}^{1 \times 1} \qquad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = C^T = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{matrix} \text{\Vert} \\ B^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{matrix}$$

$$D = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \qquad E = \begin{bmatrix} 4 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad F = \begin{bmatrix} 4 & 2 & 0 & 1 \\ 2 & 0 & 7 & 4 \\ 0 & 7 & 6 & 0 \\ 1 & 4 & 0 & 3 \end{bmatrix} = F^T$$

$$\cancel{}$$

$$D^T = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad \cancel{} \qquad E^T = \begin{bmatrix} 4 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$$

$$(A^TA)^T = A^T \cdot (A^T)^T = A^T \cdot A \, .$$

$$A \in \mathbb{R}^{m \times n} \, , \quad A^T \in \mathbb{R}^{n \times m} \qquad A^TA \in \mathbb{R}^{n \times n}$$

$A^T A$ is Symmetric

A very common example: For **any** matrix A with columns $a_1, \ldots, a_n$,

$$A^T A = \begin{bmatrix} -- & a_1^T & -- \\ -- & a_2^T & -- \\ \vdots & \vdots & \vdots \\ -- & a_n^T & -- \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & \cdots & | \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}}_{\text{Entries are the dot products of columns of } A}$$

Notation

$$\vec{x} \cdot \vec{y} \underset{\substack{\uparrow \\ \text{dot product}}}{} = \langle \vec{x}, \vec{y} \rangle = x^T \cdot y \underset{\text{matrix multiplication}}{} \in \mathbb{R}^{1 \times 1}$$

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$$

Recall A is a square matrix. $(\lambda_1, \vec{v}_1)$ $(\lambda_2, \vec{v}_2)$

eigenvalues, eigenvectors.   If $\lambda_1 \neq \lambda_2$ then $\vec{v}_1, \vec{v}_2$ are
                                                                    linearly independent

Symmetric Matrices and their Eigenspaces

Theorem

$\lambda_1 \neq \lambda_2$

A is a **symmetric** matrix, with eigenvectors $\vec{v}_1$ and $\vec{v}_2$ corresponding to two distinct eigenvalues. Then $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.

More generally, eigenspaces associated to distinct eigenvalues are orthogonal subspaces.

Proof:

matrix multiplication

$$\vec{x} \cdot (A\vec{y}) = \langle x, Ay \rangle = x^T \cdot A \cdot y = (x^T \cdot A \cdot y)^T$$

$\uparrow$
$\mathbb{R}^{1 \times 1}$

$$= y^T \cdot A^T \cdot (x^T)^T = (y)^T \cdot (A^T \cdot x) = \langle y, A^T x \rangle$$

$$= \langle A^T x, y \rangle = (A^T x) \cdot y$$

$$\lambda_1 \langle v_1, v_2 \rangle = \langle A v_1, v_2 \rangle = \langle v_1, A v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$$

$\qquad\qquad\quad \underset{\lambda_1 v_1}{\underline{\parallel}} \qquad\qquad\qquad \underset{\lambda_2 v_2}{\underline{\parallel}}$

$$(\lambda_1 - \lambda_2) \cdot \langle v_1, v_2 \rangle = 0 \qquad \Rightarrow \qquad v_1 \cdot v_2 = 0$$

$\qquad\quad \underset{\neq 0}{\underline{}}$

$$\Rightarrow \quad \text{orthogonal.}$$

$$\underline{\underline{P}} = [\text{ eigen vectors }] \qquad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

$$A = P \cdot D \underbrace{(P^{-1})}_{= P^T}$$

Example 1

$P:$ orthogonal $\begin{cases} \text{columns} & \text{orthonormal} \\ P^T \cdot P = I \end{cases}$

square

Diagonalize A using an orthogonal matrix. Eigenvalues of A are given. $\quad P^{-1} = P^T$

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda = -1, 1$$

Hint: Gram-Schmidt

① $\lambda = 1$: $\quad E_1 = \text{Nul}(A - 1 \cdot I) = \text{Nul} \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}$

$= \text{Nul} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \quad x - z = 0$

$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ x \end{bmatrix}$

$= \text{Span} \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$

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

② $\lambda = -1$: $\quad E_{-1} = \text{Nul}(A - (-1) \cdot I)$

$= \text{Nul} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \text{Nul} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \quad \begin{array}{c} x + z = 0 \\ \hline y = 0 \end{array}$

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

$= \text{Span} \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$

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

Spectral Theorem

Recall: If P is an orthogonal $n \times n$ matrix, then $P^{-1} = P^T$, which implies $A = PDP^T$ is diagonalizable and symmetric.

> **Theorem: Spectral Theorem**
>
> An $n \times n$ symmetric matrix A has the following properties.
>
> 1. All eigenvalues of A are real.
>
> 2. The dimenison of each eigenspace is full, that it's dimension is **equal to** it's algebraic multiplicity. $\Leftarrow$ geometric multiplicity
>
> 3. The eigenspaces are mutually orthogonal. $\checkmark$
>
> 4. A can be diagonalized: $A = PDP^T$, where D is diagonal and P is orthogonal.

Proof (if time permits):

①   Suppose $\quad \lambda \in \mathbb{C}$ eigenvalue $\quad$ w/ $\quad v \in \mathbb{C}^n$. $\quad$ ev

$$\left(\overline{\lambda \cdot (\overline{v}^T \cdot v)} \right)^T = \left(\overline{v}^T \cdot Av \right)^T = \left(v^T \cdot A \cdot \overline{v} \right)^T$$

$$= \overline{\lambda} \cdot (\overline{v}^T \cdot v) \qquad\qquad = \overline{v}^T \cdot A \, v$$

$$\qquad\qquad\qquad\qquad\qquad = \lambda \cdot (\overline{v}^T \cdot v)$$

$$\overline{\lambda} = \lambda \, ,$$

$$v = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

$$\overline{v}^T = \begin{bmatrix} \overline{c_1} & \overline{c_2} & \cdots & \overline{c_n} \end{bmatrix}$$

$$\overline{v}^T \cdot v = \overline{c_1} \cdot c_1 + \cdots + \overline{c_n} \cdot c_n$$

$$= |c_1|^2 + \cdots + |c_n|^2$$

$$\geq 0 \qquad (\text{real})$$

$$= 0 \Rightarrow v = 0$$

Section 7.1   Slide 7

②   <u>Induction</u> $\boxed{W}$ = Span $\{\vec{v}\}$ eigenvalue

$$z \in W^\perp \quad \Rightarrow \quad Az \in W^\perp$$

$\Downarrow$

$$\langle z, w \rangle = 0 \quad \forall \ \underline{w \in W}$$

$$\langle z, c \cdot \vec{v} \rangle = 0$$

$$\langle \boxed{Az}, \vec{v} \rangle = \langle z, A\vec{v} \rangle = \lambda \langle z, \vec{v} \rangle = 0$$

$$\Rightarrow \qquad Az \in W^\perp$$

$$\mathbb{R}^n = \quad W \oplus W^\perp$$

Section 7.2 : Quadratic Forms

Chapter 7: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Quadratic forms
2. Change of variables
3. Principle axes theorem
4. Classifying quadratic forms

Learning Objectives

1. Characterize and classify quadratic forms using eigenvalues and eigenvectors.
2. Express quadratic forms in the form $Q(\vec{x}) = \vec{x}^T A \vec{x}$.
3. Apply the principle axes theorem to express quadratic forms with no cross-product terms.

Motivating Question Does this inequality hold for all x, y?

$$z = x^2 - 6xy + 9y^2 \geq 0$$

Quadratic Forms

Definition — For a given symmetric matrix A,

A **quadratic form** is a function $Q : \mathbb{R}^n \to \mathbb{R}$, given by

$$Q(\vec{x}) = \vec{x}^T A \vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

Matrix A is $n \times n$ and symmetric.

In the above, $\vec{x}$ is a vector of variables.

Ex $\quad\quad A = I \quad\quad Q(\vec{x}) = \quad x^T \cdot I \cdot x \quad = \quad x^T \cdot x$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2$$

$$= \|\vec{x}\|^2 = \vec{x} \underset{\uparrow}{\cdot} \vec{x}$$

$$\text{dot product}.$$

Example 1

Compute the quadratic form $\vec{x}^T A \vec{x}$ for the matrices below.

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 1 \\ 1 & -3 \end{bmatrix}$$

$$Q_A(\vec{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 & 3x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2 = y$$

$$Q_B(\vec{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 + x_2 & x_1 - 3x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= (4x_1 + x_2) \cdot x_1 + (x_1 - 3x_2) \cdot x_2 = 4x_1^2 + 2x_1 \cdot x_2 - 3x_2^2 = y$$

Example 1 - Surface Plots

The surfaces for Example 1 are shown below.
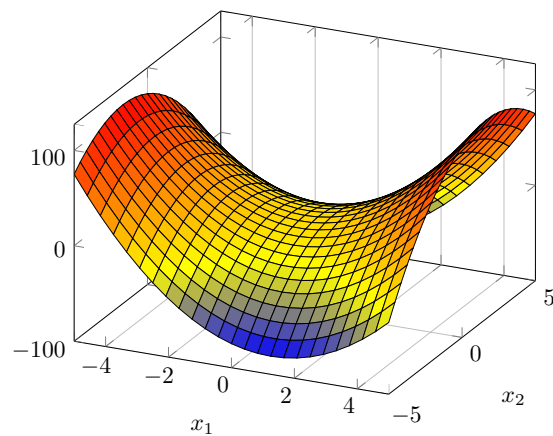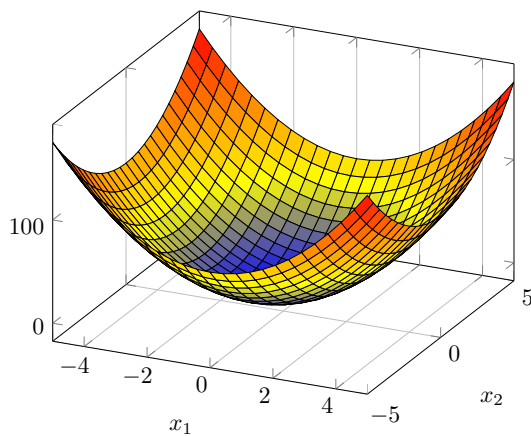
Students are not expected to be able to sketch quadratic surfaces, but it is helpful to see what they look like.

$$Q_I(\underset{\sim}{x}) = x \cdot x = B_A(x, x)$$

$$B(x, y) = y^T \underset{\approx}{A} x \quad \leftarrow \text{dot product.}$$

Example 2

Write Q in the form $\vec{x}^T A \vec{x}$ for $\vec{x} \in \mathbb{R}^3$.

$$Q(x) = 5x_1^2 - x_2^2 + 3x_3^2 + 6x_1x_3 - 12x_2x_3 \quad + \cancel{0} \quad + \cancel{5x_3}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad f \quad \cancel{x_1^2 x_3}$$

$$= \begin{bmatrix} a_{11} \cdot x_1 + a_{21} \cdot x_2 + a_{31} \cdot x_3 & , & a_{12} \cdot x_1 + a_{22} \cdot x_2 + a_{32} \cdot x_3 & , & \cdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \overset{5}{a_{11}} \cdot x_1 \cdot x_1 \quad + 2a_{21} \cdot x_2 \cdot x_1 \quad +2a_{31}^{=6} \cdot x_3 \cdot x_1$$

$$+ a_{12} \cdot x_1 \cdot x_2 \quad + a_{22}^{=-1} \cdot x_2 \cdot x_2 \quad +2a_{32}^{=-12} \cdot x_3 \cdot x_2$$

$$+ a_{13} \cdot x_1 \cdot x_3 \quad + a_{23} \cdot x_2 \cdot x_3 \quad + a_{33} \cdot x_3 \cdot x_3$$

$$\overset{3}{\underset{\underset{}{}}{=}} \sum_{j=1}^{3} \sum_{i=1}^{3 = n} a_{ij} \cdot x_i \cdot x_j$$

In general

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 3 \\ 0 & -1 & -6 \\ 3 & -6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Change of Variable

If $\vec{x}$ is a variable vector in $\mathbb{R}^n$, then a **change of variable** can be represented as

$$\vec{x} = P\vec{y}, \quad \text{or} \quad \vec{y} = P^{-1}\vec{x}$$

With this change of variable, the quadratic form $\vec{x}^T A \vec{x}$ becomes:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \longrightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$x_1 = a y_1 + b y_2$$
$$x_2 = c y_1 + d y_2$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$x_1 = 2 y_1 + 0 y_2$$
$$x_2 = 0 y_1 + 3 y_2$$

$$\vec{x} = P \cdot \vec{y}$$

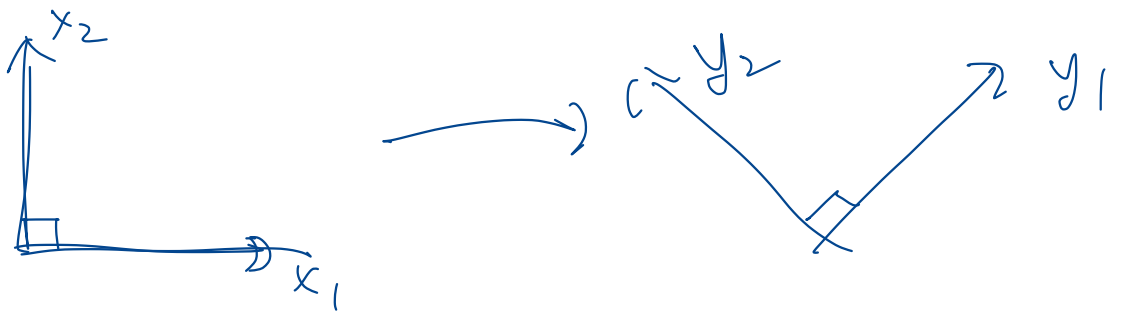Invertible

Example 3

$\uparrow$

symmetric

$\downarrow$

Make a change of variable $\vec{x} = P\vec{y}$ that transforms $Q = \vec{x}^T A \vec{x}$ so that it does not have cross terms. The orthogonal decomposition of A is given.

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} = PDP^T \qquad \underline{x^T \cdot P} = \underline{(P^T x)^T}$$

$3x_1^2 + 4x_1 x_2 + 6x_2^2$

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \leftarrow \text{rotation} \qquad = y^T$$

$\parallel$

$[x_1 \ x_2] \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix}$$

$\parallel$

$\prime\prime \ y$ (change of variable

$Q(\vec{x}) = \quad x^T \cdot A \cdot x \quad = \quad \underline{x^T \cdot P \cdot D \cdot \underline{P^T} \cdot x}$

$= \quad y^T \cdot D \cdot y \quad = \quad [y_1 \ y_2] \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$= \quad 2y_1^2 + 7y_2^2 \quad = \quad 2 \cdot \left(\dfrac{2x_1 - x_2}{\sqrt{5}} \right)^2 + 7 \left(\dfrac{x_1 + 2x_2}{\sqrt{5}} \right)^2$

$y = P^T \cdot x = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{bmatrix}$

$y_2 = 1$

y_2

y_1
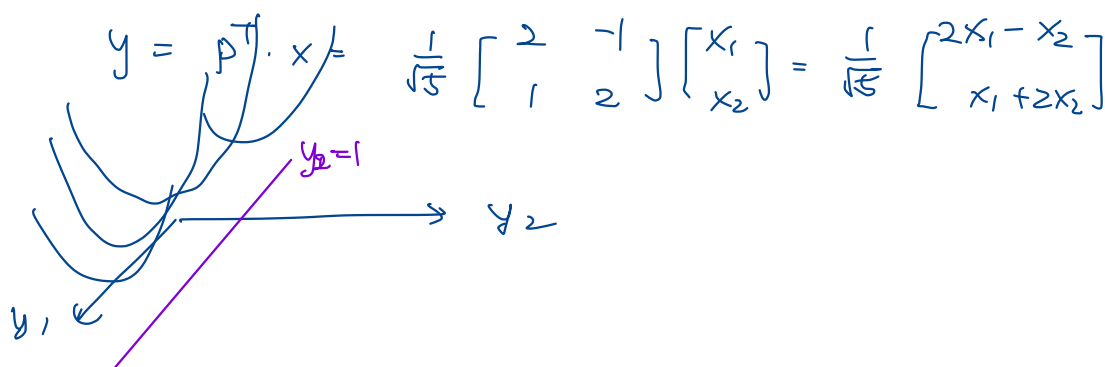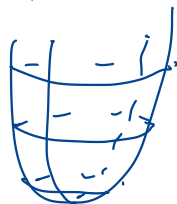
Geometry

Suppose $Q(\vec{x}) = \vec{x}^T A \vec{x}$, where $A \in \mathbb{R}^{n \times n}$ is symmetric. Then the set of $\vec{x}$ that satisfies

$$C = \vec{x}^T A \vec{x}$$

defines a curve or surface in $\mathbb{R}^n$.
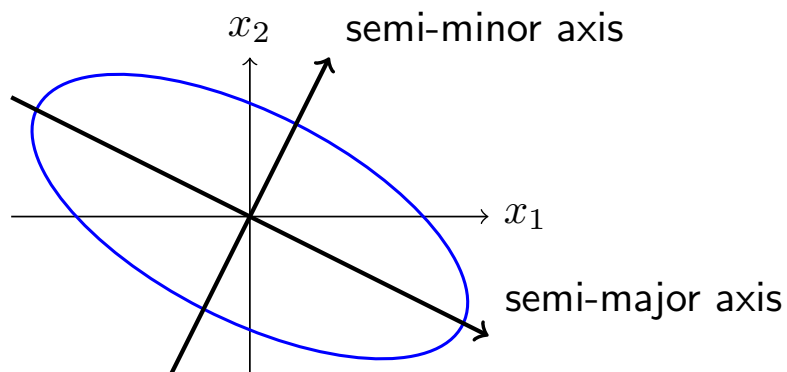
Principle Axes Theorem

> **Theorem**
>
> If A is a _____ matrix then there exists an orthogonal change of variable $\vec{x} = P\vec{y}$ that transforms $\vec{x}^T A \vec{x}$ to $\vec{x}^T D \vec{x}$ with no cross-product terms.
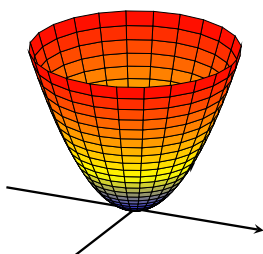
Proof (if time permits):

Example 5

Compute the quadratic form $Q = \vec{x}^T A \vec{x}$ for $A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$, and find a change of variable that removes the cross-product term. A sketch of Q is below.
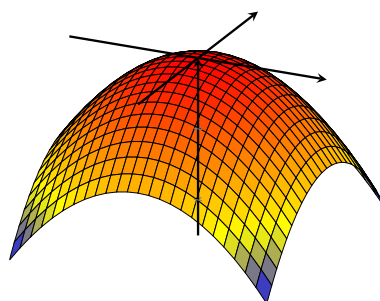
Classifying Quadratic Forms

$$Q = x_1^2 + x_2^2 \qquad Q = -x_1^2 - x_2^2$$

Definition

A quadratic form Q is
1. **positive definite** if _____ for all $\vec{x} \neq \vec{0}$.
2. **negative definite** if _____ for all $\vec{x} \neq \vec{0}$.
3. **positive semidefinite** if _____ for all $\vec{x}$.
4. **negative semidefinite** if _____ for all $\vec{x}$.
5. **indefinite** if _____

Quadratic Forms and Eigenvalues

Theorem

If A is a _____ matrix with eigenvalues λ_i, then $Q = \vec{x}^T A \vec{x}$ is

1. **positive definite** iff λ_i _____

2. **negative definite** iff λ_i _____

3. **indefinite** iff λ_i _____

Proof (if time permits):

Example 6

We can now return to our motivating question (from first slide): does this inequality hold for all x, y?

$$x^2 - 6xy + 9y^2 \geq 0$$

Section 7.3 : Constrained Optimization

Chapter 7: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Constrained optimization as an eigenvalue problem
2. Distance and orthogonality constraints

Learning Objectives

1. Apply eigenvalues and eigenvectors to solve optimization problems that are subject to distance and orthogonality constraints.
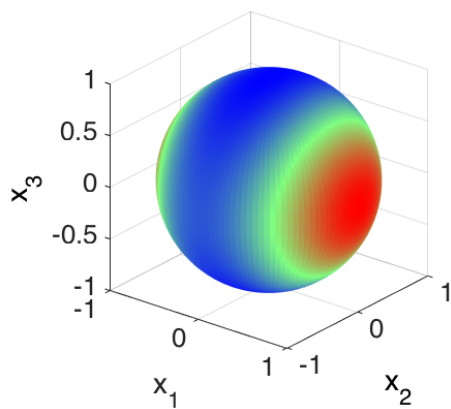
Example 1

The surface of a unit sphere in $\mathbb{R}^3$ is given by

$$1 = x_1^2 + x_2^2 + x_3^2 = ||\vec{x}||^2$$

Q is a quantity we want to optimize

$$Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

Find the largest and smallest values of Q on the surface of the sphere.

A Constrained Optimization Problem

Suppose we wish to find the maximum or minimum values of

$$Q(\vec{x}) = \vec{x}^T A \vec{x}$$

subject to

$$||\vec{x}|| = 1$$

That is, we want to find

$$m = \min\{Q(\vec{x}) : ||\vec{x}|| = 1\}$$
$$M = \max\{Q(\vec{x}) : ||\vec{x}|| = 1\}$$

This is an example of a **constrained optimization** problem. Note that we may also want to know where these extreme values are obtained.

Constrained Optimization and Eigenvalues

Theorem

If $Q = \vec{x}^T A \vec{x}$, A is a real $n \times n$ symmetric matrix, with eigenvalues

$$\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$$

and associated normalized eigenvectors

$$\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$$

Then, subject to the constraint $||\vec{x}|| = 1$,
- the **maximum** value of $Q(\vec{x}) = \lambda_1$, attained at $\vec{x} = \pm \vec{u}_1$.
- the **minimum** value of $Q(\vec{x}) = \lambda_n$, attained at $\vec{x} = \pm \vec{u}_n$.
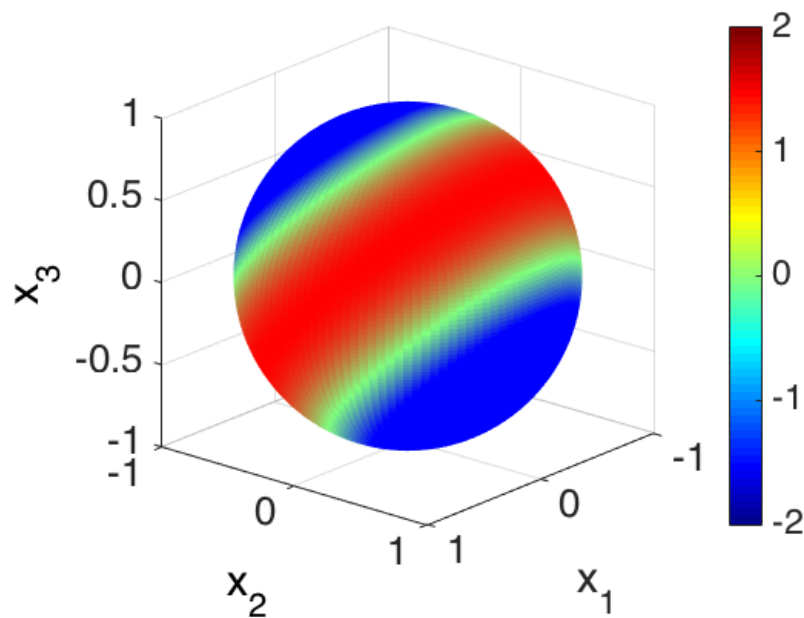
Proof:

Example 2

Calculate the maximum and minimum values of $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $\vec{x} \in \mathbb{R}^3$, subject to $||\vec{x}|| = 1$, and identify points where these values are obtained.

$$Q(\vec{x}) = x_1^2 + 2x_2 x_3$$

Example 2

The image below is the unit sphere whose surface is colored according to the quadratic from the previous example. Notice the agreement between our solution and the image.

An Orthogonality Constraint

Theorem

Suppose $Q = \vec{x}^T A \vec{x}$, A is a real $n \times n$ symmetric matrix, with eigenvalues

$$\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$$

and associated eigenvectors

$$\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$$

Subject to the constraints $||\vec{x}|| = 1$ and $\vec{x} \cdot \vec{u}_1 = 0$,
- The maximum value of $Q(\vec{x}) = \lambda_2$, attained at $\vec{x} = \vec{u}_*$.
- The minimum value of $Q(\vec{x}) = \lambda_n$, attained at $\vec{x} = \vec{u}_n$.

Note that λ_2 is the second largest eigenvalue of A.

Example 3

Calculate the maximum value of $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $\vec{x} \in \mathbb{R}^3$, subject to $||\vec{x}|| = 1$ and to $\vec{x} \cdot \vec{u}_1 = 0$, and identify a point where this maximum is obtained.

$$Q(\vec{x}) = x_1^2 + 2x_2 x_3, \qquad \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Example 4 (if time permits)

Calculate the maximum value of $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $\vec{x} \in \mathbb{R}^3$, subject to $||\vec{x}|| = 5$, and identify a point where this maximum is obtained.

$$Q(\vec{x}) = x_1^2 + 2x_2 x_3$$

Section 7.4 : The Singular Value Decomposition

Chapter 7: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. The Singular Value Decomposition (SVD) and some of its applications.

**Learning Objectives**

1. Compute the SVD for a rectangular matrix.
2. Apply the SVD to
   - ▶ estimate the rank and condition number of a matrix,
   - ▶ construct a basis for the four fundamental spaces of a matrix, and
   - ▶ construct a spectral decomposition of a matrix.

Example 1

The linear transform whose standard matrix is

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

maps the unit circle in $\mathbb{R}^2$ to an ellipse, as shown below. Identify the unit vector $\vec{x}$ in which $||A\vec{x}||$ is maximized and compute this length.
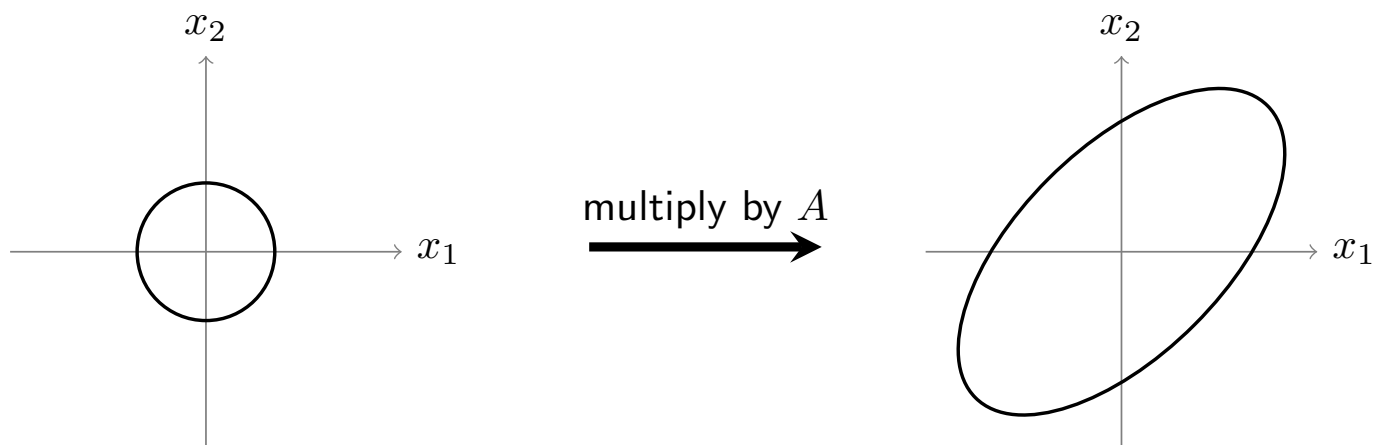
Example 1 - Solution

Singular Values

The matrix $A^T A$ is always symmetric, with non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be the associated orthonormal eigenvectors. Then

$$\|A\vec{v}_j\|^2 =$$

If the A has rank r, then $\{A\vec{v}_1, \ldots, A\vec{v}_r\}$ is an orthogonal basis for $\text{Col}A$: For $1 \leq j < k \leq r$:

$$(A\vec{v}_j)^T A\vec{v}_k =$$

Definition: $\sigma_1 = \sqrt{\lambda_1} \geq \sigma_2 = \sqrt{\lambda_2} \cdots \geq \sigma_n = \sqrt{\lambda_n}$ are the singular values of A.
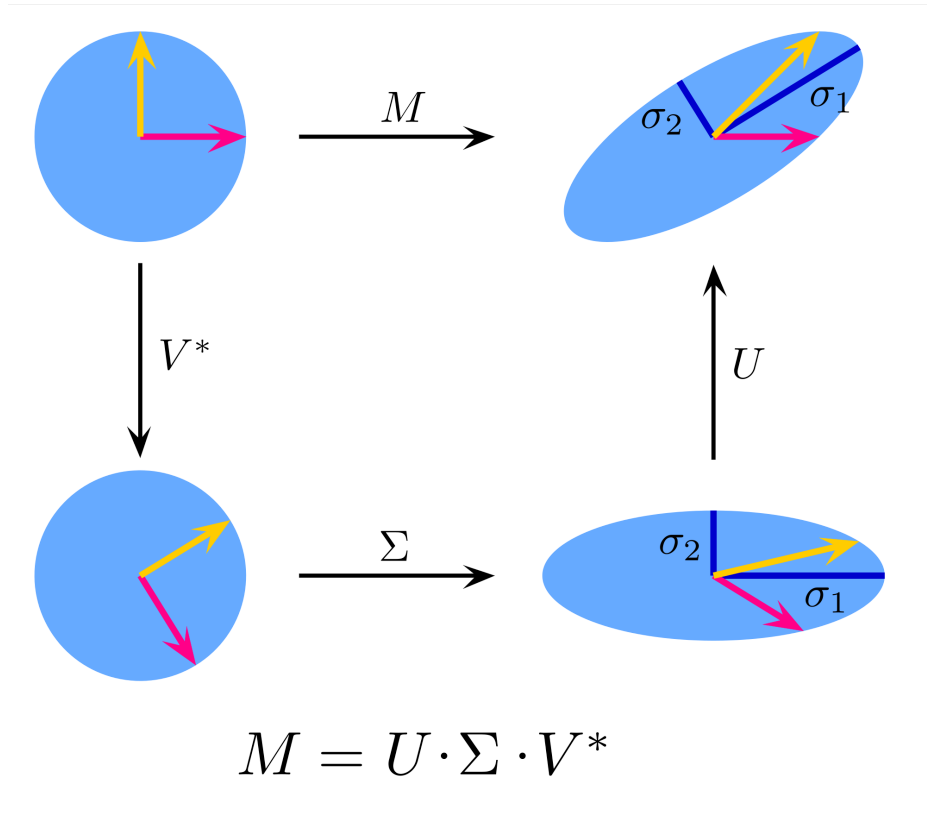
The SVD

Theorem: Singular Value Decomposition

A $m \times n$ matrix with rank r and non-zero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ has a decomposition $U\Sigma V^T$ where

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}_{m \times n} = \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 & \\ 0 & \sigma_2 & \ldots & \vdots & \mathbf{0} \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \ldots & \sigma_r & \\ & \mathbf{0} & & & \mathbf{0} \end{bmatrix}$$

U is a $m \times m$ orthogonal matrix, and V is a $n \times n$ orthogonal matrix.

$$M = U \cdot \Sigma \cdot V^*$$

Algorithm to find the SVD of A

Suppose A is $m \times n$ and has rank $r \leq n$.

1. Compute the squared singular values of $A^T A$, σ_i^2, and construct Σ.

2. Compute the unit singular vectors of $A^T A$, $\vec{v}_i$, use them to form V.

3. Compute an orthonormal basis for $\text{Col}A$ using

$$\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i, \quad i = 1, 2, \ldots r$$

Extend the set $\{\vec{u}_i\}$ to form an orthonomal basis for $\mathbb{R}^m$, use the basis for form U.

Example 2: Write down the singular value decomposition for

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

Example 3: Construct the singular value decomposition of

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

(It has rank 1.)

Applications of the SVD

The SVD has been applied to many modern applications in CS, engineering, and mathematics (our textbook mentions the first four).

- Estimating the rank and condition number of a matrix

- Constructing bases for the four fundamental spaces

- Computing the pseudoinverse of a matrix

- Linear least squares problems

- Non-linear least-squares
  https://en.wikipedia.org/wiki/Non-linear_least_squares

- Machine learning and data mining
  https://en.wikipedia.org/wiki/K-SVD

- Facial recognition
  https://en.wikipedia.org/wiki/Eigenface

- Principle component analysis
  https://en.wikipedia.org/wiki/Principal_component_analysis

- Image compression

Students are expected to be familiar with the 1st two items in the list.

The Condition Number of a Matrix

If A is an invertible $n \times n$ matrix, the ratio

$$\frac{\sigma_1}{\sigma_n}$$

is the **condition number** of A.

Note that:

- The condition number of a matrix describes the sensitivity of a solution to $A\vec{x} = \vec{b}$ is to errors in A.
- We could define the condition number for a rectangular matrix, but that would go beyond the scope of this course.

Example 4

For $A = U\Sigma V^*$, determine the rank of A, and orthonormal bases for Null$A$ and $(\text{Col}A)^\perp$.

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

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

Example 4 – Solution
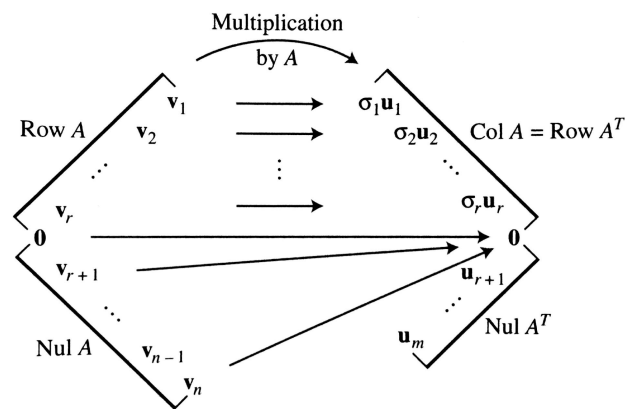
The Four Fundamental Spaces

FIGURE 4 The four fundamental subspaces and the action of A.

1. $A\vec{v}_s = \sigma_s \vec{u}_s$.
2. $\vec{v}_1, \ldots, \vec{v}_r$ is an orthonormal basis for $\text{Row} A$.
3. $\vec{u}_1, \ldots, \vec{u}_r$ is an orthonormal basis for $\text{Col} A$.
4. $\vec{v}_{r+1}, \ldots, \vec{v}_n$ is an orthonormal basis for $\text{Null} A$.
5. $\vec{u}_{r+1}, \ldots, \vec{u}_n$ is an orthonormal basis for $\text{Null} A^T$.

The Spectral Decomposition of a Matrix

The SVD can also be used to construct the spectral decomposition for any matrix with rank r

$$A = \sum_{s=1}^{r} \sigma_s \vec{u}_s \vec{v}_s^T,$$

where $\vec{u}_s, \vec{v}_s$ are the sth columns of U and V respectively.

For the case when $A = A^T$, we obtain the same spectral decomposition that we encountered in Section 7.2.