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

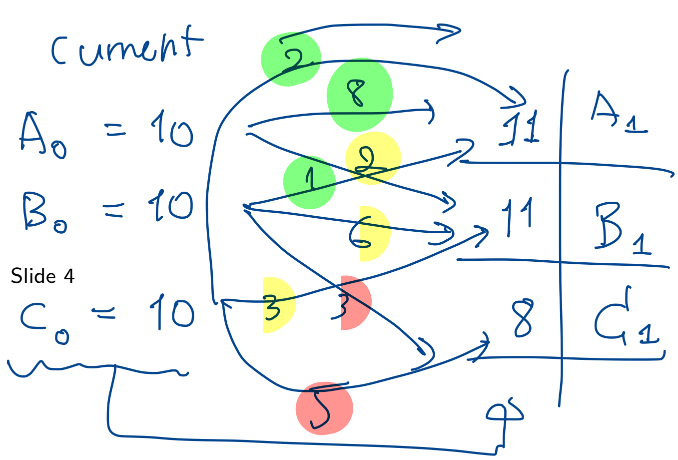
Recall the car rental problem from our Section 4.9 lecture.

Problem

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

		rented from		
		A	B	C
returned to	A	.8	.1	.2
	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?



$$A_1 = 0.8 \cdot A_0 + 0.1 B_0 + 0.2 C_0$$
$$B_1 = 0.2 A_0 + 0.6 B_0 + 0.3 C_0$$
$$C_1 = 0 \cdot A_0 + 0.3 B_0 + 0.5 C_0$$

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

Long Term Behaviour

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

$$\begin{bmatrix} A_1 \\ B_1 \\ C_1 \end{bmatrix} = \vec{x}_1 = P\vec{x}_0 = \begin{bmatrix} A_0 \\ B_0 \\ C_0 \end{bmatrix}$$

The distribution of cars after 2 weeks is:

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

The distribution of cars after n weeks is:

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

$$\cdots = P^n \cdot \vec{x}_0$$

probability vectors.

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

Sum. 1 1 1

stochastic matrix.

Long Term Behaviour

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

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

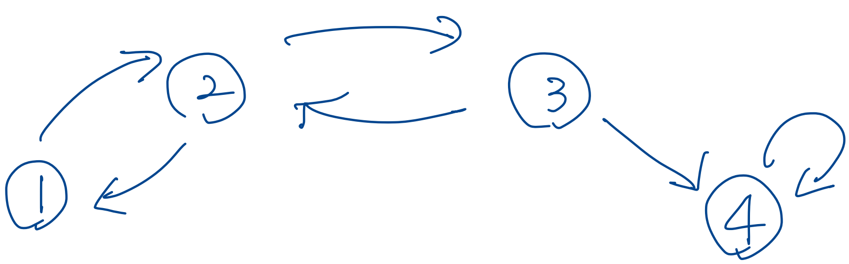
1. compute the **steady-state vector**, $\vec{q}$, by solving $\vec{q} = P\vec{q}$.
2. compute $P^n\vec{x}_0$ for large n.
3. compute P^n for large n, each column of the resulting matrix is the steady-state

<u>Recall</u> P stochastic matrix.

P is regular if there exists $k \in \{1, 2, \cdots\}$ such that P^k entries are strictly positive.

(after k steps)

(positive probability to reach every state)



$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} P^{n+1} = \lim_{n\to\infty} P^n \cdot [\vec{x}_1 \quad \vec{x}_2 \cdots \vec{x}_m]$$

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

Theorem 1

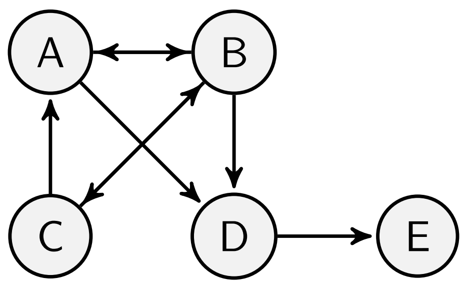
If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

1. There is a stochastic matrix Π such that
$$\lim_{n\to\infty} P^n = \Pi = [\vec{q} \quad \vec{q} \cdots \vec{q}]$$
2. Each column of Π is the same probability vector $\vec{q}$.
3. For any initial probability vector $\vec{x}_0$,
$$\lim_{n\to\infty} P^n \vec{x}_0 = \vec{q}$$
4. P has a unique eigenvector, $\vec{q}$, which has eigenvalue $\lambda = 1$. $(P\vec{q} = \vec{q} = 1 \cdot \vec{q})$
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.

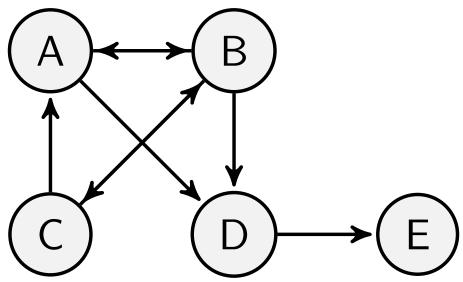
Use these assumptions to construct a Markov chain that represents how users navigate the above web.

$$P = \begin{array}{c|ccccc|c} \text{From} & A & B & C & D & E & \\ \hline & 0 & 1/3 & 1/2 & 0 & 0 & A \\ & 1/2 & 0 & 1/2 & 0 & 0 & B \\ & 0 & 1/3 & 0 & 0 & 0 & C \\ & 1/2 & 1/3 & 0 & 0 & 0 & D \\ & 0 & 0 & 0 & 1 & 1 & E \end{array}$$

Not regular.

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?

$$P^* = \begin{array}{c} \begin{array}{cccccc} \text{From} & A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1/3 & 1/2 & 0 & 1/5 \\ 1/2 & 0 & 1/2 & 0 & 1/5 \\ 0 & 1/3 & 0 & 0 & 1/5 \\ 1/2 & 1/3 & 0 & 0 & 1/5 \\ 0 & 0 & 0 & 1 & 1/5 \end{pmatrix} \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}$$

regular

$\Updownarrow$

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:

$$\overset{\text{w/ prob. } p}{P^*} \qquad\qquad \overset{\text{w/ prob } (1-p)}{} = K$$

$$G = p \cdot \begin{pmatrix} 0 & 1/3 & 1/2 & 0 & 1/5 \\ 1/2 & 0 & 1/2 & 0 & 1/5 \\ 0 & 1/3 & 0 & 0 & 1/5 \\ 1/2 & 1/3 & 0 & 0 & 1/5 \\ 0 & 0 & 0 & 1 & 1/5 \end{pmatrix} + (1-p) \cdot \begin{pmatrix} 1/5 & \cdots & 1/5 \\ 1/5 & \cdots & 1/5 \\ 1/5 & \ddots & \\ 1/5 & & \\ 1/5 & & 1/5 \end{pmatrix}$$

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:

$$G = 0.85 \cdot \begin{pmatrix} 0 & 1/3 & 1/2 & 0 & 1/5 \\ 1/2 & 0 & 1/2 & 0 & 1/5 \\ 0 & 1/3 & 0 & 0 & 1/5 \\ 1/2 & 1/3 & 0 & 0 & 1/5 \\ 0 & 0 & 0 & 1 & 1/5 \end{pmatrix} + 0.15 \cdot \begin{pmatrix} 1/5 & \cdots & 1/5 \\ 1/5 & \cdots & 1/5 \\ 1/5 & \ddots & \\ 1/5 & & \\ 1/5 & & 1/5 \end{pmatrix}$$

Computing Page Rank

WOLFRAM CLOUD Plan

(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=

$$G = \begin{pmatrix} 0.03 & 0.455 & 0.455 & 0.03 & 0.2 \\ 0.455 & 0.03 & 0.455 & 0.455 & 0.2 \\ 0.03 & 0.455 & 0.03 & 0.03 & 0.2 \\ 0.455 & 0.03 & 0.03 & 0.03 & 0.2 \\ 0.03 & 0.03 & 0.03 & 0.455 & 0.2 \end{pmatrix}$$

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

Out[15]//MatrixForm=

$$G^2 = \begin{pmatrix} 0.241225 & 0.241225 & 0.241225 & 0.313475 & 0.234 \\ 0.253975 & 0.4346 & 0.253975 & 0.1456 & 0.319 \\ 0.228475 & 0.04785 & 0.228475 & 0.300725 & 0.149 \\ 0.04785 & 0.228475 & 0.228475 & 0.1201 & 0.149 \\ 0.228475 & 0.04785 & 0.04785 & 0.1201 & 0.149 \end{pmatrix}$$

```
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=

$$G^{10} = \begin{pmatrix} 0.251705 & 0.251712 & 0.251712 & 0.251715 & 0.251709 \\ 0.298496 & 0.298551 & 0.298359 & 0.298349 & 0.2985 \\ 0.176611 & 0.176552 & 0.176745 & 0.176755 & 0.176605 \\ 0.156761 & 0.156774 & 0.156774 & 0.15676 & 0.156768 \\ 0.116427 & 0.116411 & 0.116411 & 0.116421 & 0.116418 \end{pmatrix}$$

A : Page Rank 2

B : Page Rank 1.

q

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

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

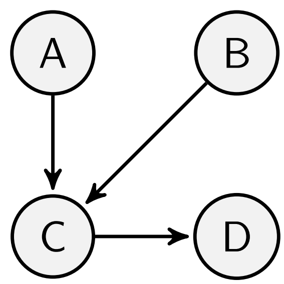
$$\begin{pmatrix} .8 & .1 & .2 \\ .2 & .6 & .3 \\ .0 & .3 & .5 \end{pmatrix}^{10}$$

- 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

A is a square matrix.

Definition

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

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

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

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

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

$$\ne \qquad\qquad \ne$$

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

$A \in \mathbb{R}^{m\times n}$ $A^T \in \mathbb{R}^{n\times m}$ $A^T \cdot A \in \mathbb{R}^{n\times n}$

$(A^T \cdot A)^T = A^T \cdot (A^T)^T = A^T \cdot A \Rightarrow$ Symmetric.

$A^T A$ is Symmetric

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

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

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

matrix multiplication

$\underset{\text{row}}{x^T} \cdot \underset{\text{column}}{x} = x \cdot x$

dot product.

$x \cdot y = y \cdot x$

$\parallel$

$\underbrace{x^T \cdot y}_{\in \mathbb{R}^{1\times 1}} = (x^T \cdot y)^T$

$= y^T \cdot (x^T)^T = y^T \cdot x$

Note A is a matrix (λ_1, v_1), (λ_2, v_2) eigenvalues, eigenvectors

If $\lambda_1 \neq \lambda_2$ then v_1, v_2 are linearly independent.

Symmetric Matrices and their Eigenspaces

Theorem

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.

$\lambda_1 \neq \lambda_2$

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

Proof:

matrix multi.

Recall: $\vec{x} \cdot (A\vec{y}) = \underbrace{x^T \cdot (Ay)}_{\in \mathbb{R}^{1\times 1}} = (x^T \cdot A \cdot y)^T = y^T \cdot (A^T) \cdot (x^T)^T$

$= y^T \cdot (A^T \cdot x) = y \cdot (A^T x) = A^T x \cdot y$

$\lambda_1 \langle v_1, v_2 \rangle = \underbrace{\langle \overset{\lambda_1 v_1}{\overset{\parallel}{A v_1}}, v_2 \rangle}_{\text{dot product}} = \langle v_1, \underset{\underset{\lambda_2 v_2}{\parallel}}{A v_2} \rangle = \lambda_2 \langle v_1, v_2 \rangle$

A^T

dot product

$\parallel$

$Av_1 \cdot v_2$

$\underbrace{(\lambda_1 - \lambda_2)}_{\neq 0} \cdot \langle v_1, v_2 \rangle = 0 \Rightarrow v_1 \cdot v_2 = 0$

v_1, v_2 orthogonal.

Example 1

$A = P \cdot D \cdot P^{-1}$ P : orthogonal

Diagonalize A using an orthogonal matrix. Eigenvalues of A are given.

Symmetric $\rightarrow$ $$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda = -1, 1$$

Hint: Gram-Schmidt

$E_1 = \mathrm{Nul}\,(A - 1 \cdot I) = \mathrm{Nul}\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} = \mathrm{Nul}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

$x - z = 0$

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

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

$E_{-1} = \mathrm{Nul}\,(A - (-1) \cdot I) = \mathrm{Nul}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

$= \mathrm{Nul}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$

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

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

$= P \cdot D \cdot P^T$.

orthogonal diagonalization

$P^T P = I$

$\Rightarrow P^T = P^{-1}$

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.
>
>   geometric multiplicity,
> 3. The eigenspaces are mutually orthogonal. (have seen before)
> 4. A can be diagonalized: $A = PDP^T$, where D is diagonal and P is orthogonal.

Proof (if time permits):

Suppose $\lambda \in \mathbb{C}$ is an eigenvalue w/ eigenvector $\vec{v} \in \mathbb{C}^n$

$\bar{\lambda} = \lambda$

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

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

$= \bar{v}^T \cdot A v \quad (Av = \lambda v)$

$= \lambda \cdot (\bar{v}^T \cdot v)$, $\bar{v}^T \cdot v \neq 0$

$\bar{\lambda} \cdot (\bar{v}^T \cdot v)$, $\bar{v}^T \cdot v \in \mathbb{R}$, $\neq 0$

$\Rightarrow \bar{\lambda} = \lambda$

$\bar{v}^T = [\bar{c}_1 \quad \bar{c}_2]$

$\bar{v}^T \cdot v = \bar{c}_1 \cdot c_1 + \bar{c}_2 \cdot c_2$

$= |c_1|^2 + |c_2|^2$

≥ 0 (real)

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?

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

Quadratic Forms

For a given matrix A,

Definition

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

column vector

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

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

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

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

$$= [x_1 \ x_2 \ \cdots \ x_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1^2 + x_2^2 + \cdots + x_n^2 = \vec{x} \cdot \vec{x}$$

dot product

$$= \|\vec{x}\|^2$$

Example 1

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

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

① $Q_A(\vec{x}) = \underbrace{[x_1\ x_2]\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 & 3x_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2$

② $Q_B(\vec{x}) = \underbrace{[x_1\ \ x_2]\begin{bmatrix} 4 & 1 \\ 1 & -3 \end{bmatrix}}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [4x_1 + x_2 \quad x_1 - 3x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

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

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

In general,

$\underbrace{[x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \\ a_{31} & & \ddots & \\ \vdots & & & a_{nn} \end{bmatrix}}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$= [a_{11} \cdot x_1 + a_{21} \cdot x_2 + \cdots,\ a_{12}x_1 + a_{22}x_2 + \cdots,\ \cdots]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$= a_{11} \cdot x_1 \cdot x_1 + a_{21} \cdot x_2 \cdot x_1 + a_{31}x_1 \cdot x_3 + \cdots + a_{12}x_1 \cdot x_2 + \cdots$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}\, x_i \cdot x_j = Q(\vec{x}).$$

$Q_A(x_1, x_2) = 4x_1^2 + 3x_2^2 = y$

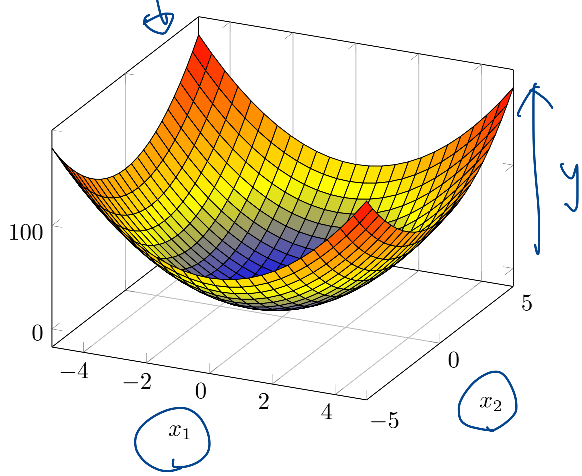
Graph

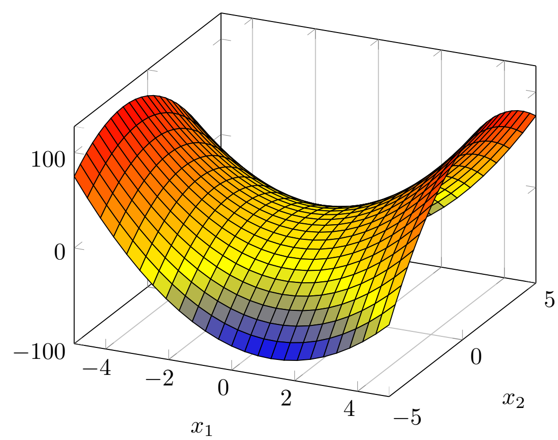
$Q_B(x_1, x_2) = 4x_1^2 + 2x_1x_2 - 3x_2^2 = y$

Graph

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.

Example 2

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

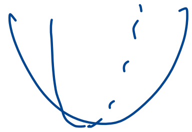
$$Q(x) = 5x_1^2 - x_2^2 + 3x_3^2 + 6x_1x_3 - 12x_2x_3$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}}_{\text{Symmetric}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \underset{=5}{a_{11}} x_1 \cdot x_1 + 2\underset{=0}{a_{12}} \cdot x_1 \cdot x_2 + 2\underset{=6}{a_{13}} \cdot x_1 \cdot x_3$$
$$+ \cancel{a_{21}\, x_2 \cdot x_1} + \underset{=-1}{a_{22}} \cdot x_2 \cdot x_2 + 2\underset{=-12}{a_{23}}\, x_2 \cdot x_3$$
$$+ \cancel{a_{31}\, x_3 \cdot x_1} + \cancel{a_{32}\, x_3 \cdot x_2} + \underset{=3}{a_{33}}\, x_3 \cdot x_3$$

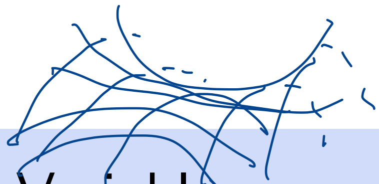
$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 3 \\ 0 & -1 & -6 \\ 3 & -6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$y = Q(x_1, x_2) = x_1^2 + 2x_2^2 \geq 0$ $\quad = 0$ only if $x_1, x_2 = 0$



$y = Q(x_1, x_2) = 5x_1^2 - 2x_2^2$

no max, no min.



Change of Variable

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

Invertible

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

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

Ex

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

Invertible.

$$x_1 = ay_1 + by_2$$
$$x_2 = cy_1 + dy_2$$

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

$$\cancel{x_1 = 2y_1 + y_2}$$
$$\cancel{x_2 = 4y_1 + 2y_2}$$

Example 3

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.

Symmetric ⇒ orthogonally diagonalizable

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} = PDP^T$$

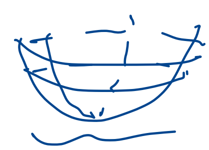
$A = PDP^{-1}$

$P^TP = I$

$P^{-1} = P^T$

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

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



$\boxed{2y_1^2 + 7y_2^2}$

$$Q_A(\vec{x}) = [x_1 \quad x_2] \underbrace{\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}}_{=A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 4x_1x_2 + 6x_2^2$$

$$= \vec{x}^T \cdot P \cdot D \cdot \underbrace{P^T \cdot \vec{x}}_{=\vec{y}}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P^T \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{2x_1 - x_2}{\sqrt{5}} \\ \frac{x_1 + 2x_2}{\sqrt{5}} \end{bmatrix}$$

$$= y^T \cdot D \cdot y$$

$$= [y_1 \quad y_2] \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$x^T \cdot P = (P^T \cdot x)^T = y^T$$

$$= 2y_1^2 + 7y_2^2$$

$$= 2 \cdot \left(\frac{2x_1 - x_2}{\sqrt{5}} \right)^2 + 7 \cdot \left(\frac{x_1 + 2x_2}{\sqrt{5}} \right)^2 \geq 0$$

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 ___real symmetric___ matrix then there exists an orthogonal change of variable $\vec{x} = P\vec{y}$ that transforms $\vec{x}^T A\vec{x}$ to $\vec{y}^T D\vec{y}$ with no cross-product terms.

Proof (if time permits):

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{pmatrix}$$

no crossing terms.

$$Q_A(\vec{x}) = x^T \cdot A \cdot x = x^T P D P^T \cdot x = y^T \cdot D \cdot y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

(A : symmetric)

$\Downarrow$

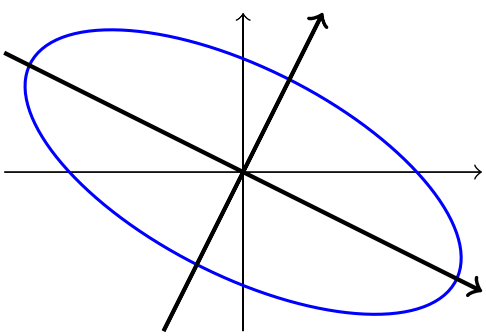
$A = PDP^T$

P : orthogonal

$y = P^T \cdot x$ (change of variables)

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.



$A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix} = P D P^T$

$D = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$ $\quad P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$

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

$= \frac{1}{\sqrt{5}} \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$Q(\vec{x}) = x^T \cdot A \cdot x = y^T \cdot D \cdot y = 4y_1^2 + 9y_2^2$

① Eigenvalues : $\phi(\lambda) = \det(A - \lambda I) \overset{2\times 2 \text{ case}}{=} \lambda^2 - tr(A)\cdot\lambda + \det(A)$

characteristic polynomial

$= \lambda^2 - 13\lambda + 36$

$= (\lambda - 4)(\lambda - 9) = 0$

$\lambda = 4, 9$

② $E_4 = Nul(A - 4I) = Nul \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ $\quad \boxed{x + 2y = 0} \quad \boxed{x = -2y}$

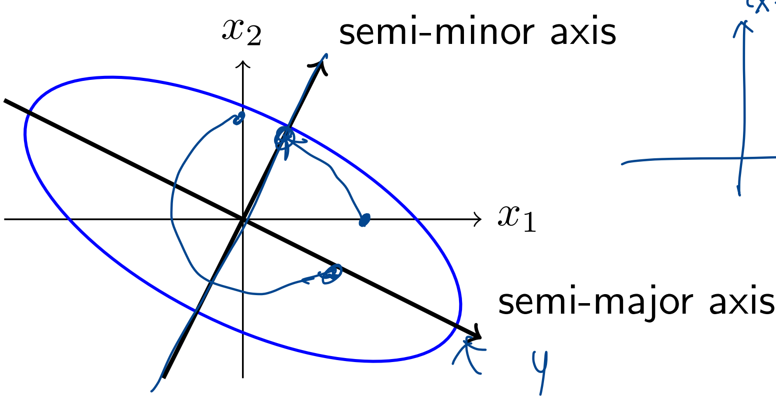
$= Span\left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$, $\quad v_1 = \underbrace{\frac{1}{\sqrt{5}} \cdot \begin{bmatrix} 2 \\ -1 \end{bmatrix}}$

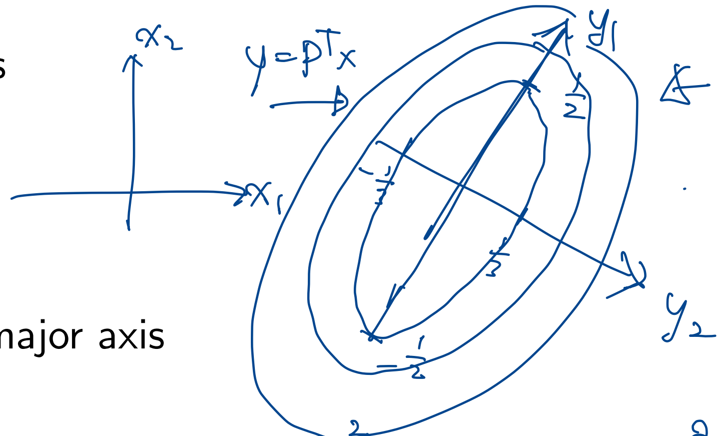
③ $E_9 = Nul(A - 9I) = Nul \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix}$ $\quad 2x - y = 0 \quad 2x = y.$

$= Span\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$, $\quad v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

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.





$$Q(\vec{x}) = x^T \cdot A \cdot x = y^T \cdot D \cdot y = 4y_1^2 + 9y_2^2 = 4\left(\frac{2x_1 - x_2}{\sqrt{5}}\right)^2 + 9\left(\frac{x_1 + 2x_2}{\sqrt{5}}\right)^2$$

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix} = P D P^T$$

$$\boxed{1 = 4 y_1^2 + 9 y_2^2}$$

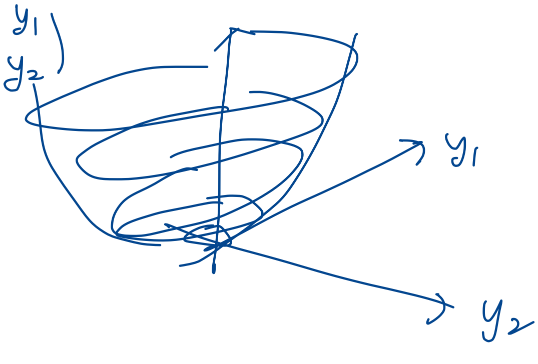
Fix z. $z = 1$

$$D = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} \qquad P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

$$y = P^T x = \frac{1}{\sqrt{5}} \cdot \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$x_1 = 1,\ x_2 = 0 \Rightarrow y_1 = \frac{2}{\sqrt{5}},\ y_2 = \frac{1}{\sqrt{5}}$

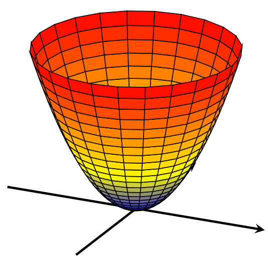
$x_1 = 0,\ x_2 = 1 \Rightarrow y_1 = -\frac{1}{\sqrt{5}},\ y_2 = \frac{2}{\sqrt{5}}$



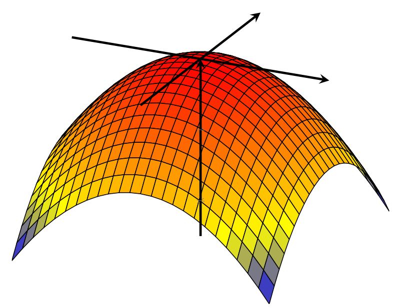
Classifying Quadratic Forms

$Q(0) = 0 = 0^T \cdot A \cdot 0$

$Q = x_1^2 + x_2^2$



$Q = -x_1^2 - x_2^2$



Definition

A quadratic form Q is

1. **positive definite** if $Q(x) > 0$ for all $\vec{x} \neq \vec{0}$.
2. **negative definite** if $Q(x) < 0$ for all $\vec{x} \neq \vec{0}$.
3. **positive semidefinite** if $Q(x) \geq 0$ for all $\vec{x}$.
4. **negative semidefinite** if $Q(x) \leq 0$ for all $\vec{x}$.
5. **indefinite** if $Q(x) > 0$ or $Q(x) = 0$ or $Q(x) < 0$

$Q = 0$ only if $x = 0$

$Q(x) = x^T A x = y^T \cdot D \cdot y$

$= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \quad \boxed{> 0}$

$\lambda_1 > 0$, $y_1 \geq 0$, $= 0$; $\lambda_2 > 0$, $y_2 \geq 0$; $\lambda_n > 0$, $y_n \geq 0$

$\mathbb{R}^n$ vector

$f(\vec{x})$ has local minimum if $\begin{cases} f'(\ ) = 0 \\ f'' \end{cases}$ P.D

$n \times n$ matrix $\longrightarrow f''$

Quadratic Forms and Eigenvalues

Theorem

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

1. **positive definite** iff λ_i > 0 for all $i = 1, \dots, n$
2. **negative definite** iff λ_i < 0 for all $i = 1, \dots, n$
3. **indefinite** iff λ_i > 0 for some i $\lambda_j < 0$ for some j.

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.

Q: Find eigenvalues by solving optimization problem for quadratic forms.

Example 1

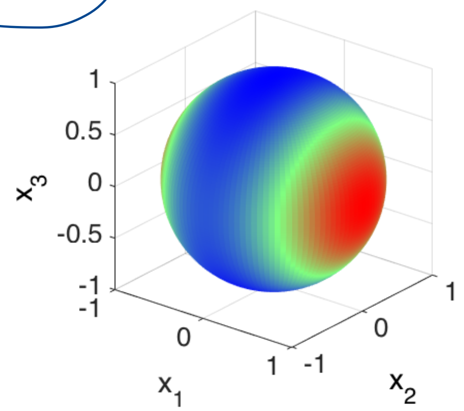
$\{ x \in \mathbb{R}^3 : \|x\| = 1 \}$

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.

$$3(x_1^2+x_2^2+x_3^2) \le Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2 \le 9x_1^2 + 9x_2^2 + 9x_3^2 = 9\cdot\underbrace{(x_1^2+x_2^2+x_3^2)}_{=1} = 9$$

$x_1 = x_2 = 0$, $x_3 = 1$

$x_1 = 1$, $x_2 = 0$, $x_3 = 0$ $\in$ Sphere

largest eigenvalue = largest value

$= 3$ smallest number.

= smallest eigenvalue.

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

$\|\vec{u}_i\| = 1$

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

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:

$$Q(\vec{x}) = x^T \cdot A \cdot x = x^T \cdot P \cdot D \cdot \underbrace{P^T \cdot x}_{=:y}$$

$P = [\vec{u}_1, \vec{u}_2 \cdots \vec{u}_n]$: orthogonal matrix

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

$$Q(\vec{x}) = [y_1 \cdots y_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = P^T \cdot \vec{x} = \begin{bmatrix} - u_1^T - \\ - u_2^T - \\ \vdots \\ - u_n^T - \end{bmatrix} \cdot \vec{x} = \begin{bmatrix} u_1^T \cdot x \\ u_2^T \cdot x \\ \vdots \\ u_n^T \cdot x \end{bmatrix} = \begin{bmatrix} x \cdot u_1 = y_1 \\ x \cdot u_2 = y_2 \\ \vdots \\ x \cdot u_n = y_n \end{bmatrix}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

$$Q(\vec{x}) = \lambda_1 \cdot (\vec{x} \cdot \vec{u}_1)^2 + \lambda_2 (\vec{x} \cdot \vec{u}_2)^2 + \cdots + \lambda_n (\vec{x} \cdot \vec{u}_n)^2$$

$$\|\vec{y}\|^2 = \vec{y} \cdot \vec{y} = \|\underset{\substack{\uparrow \\ \text{orthogonal matrix}}}{P^T} \cdot \vec{x}\|^2 = \|x\|^2 = 1.$$

Parseval's identity

$$= (\vec{x} \cdot u_1)^2 + (\vec{x} \cdot u_2)^2 + \cdots + (\vec{x} \cdot u_n)^2 = 1$$

$$Q(\vec{x}) \leq \lambda_1 (\vec{x} \cdot \vec{u}_1)^2 + \lambda_1 (\vec{x} \cdot \vec{u}_2)^2 + \cdots + \lambda_1 (\vec{x} \cdot \vec{u}_n)^2$$

$$= \lambda_1 (\underbrace{(\vec{x} \cdot u_1)^2 + \cdots + (\vec{x} \cdot u_n)^2}_{=1})$$

$$\leq \lambda_1$$

Equality holds when $\begin{cases} (\vec{x} \cdot \vec{u}_1)^2 = 1 \\ (\vec{x} \cdot \vec{u}_2)^2 = \cdots = (\vec{x} \cdot \vec{u}_n)^2 = 0 \end{cases}$

when $x = \pm \vec{u}_1$

$$Q(\vec{x}) \geq \lambda_n ((\vec{x} \cdot u_1)^2 + \cdots + (\vec{x} \cdot u_n)^2) = \lambda_n$$

Equality holds if $\begin{cases} (x \cdot u_n)^2 = 1 \\ (x \cdot u_1)^2 = \cdots = (x \cdot u_{n-1})^2 = 0 \end{cases}$

if $x = \pm \vec{u}_n$.

$\max \{ Q(\vec{x}) : \|\vec{x}\| = 1 \} = \lambda_1$: largest eigenvalue

$\min \{ Q(\vec{x}) : \|\vec{x}\| = 1 \} = \lambda_n$: smallest "

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.

$$Q(\vec{x}) = 1x_1^2 + \boxed{2}x_2x_3 \quad = \quad x^T \cdot A \cdot x .$$

$$+ \; 0 \cdot x_2^2 \; + \; 0 \cdot x_3^2$$

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

(2) $\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{bmatrix} = (1-\lambda) \cdot \det \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}$

$= (1-\lambda) \cdot (\lambda^2 - 1) = (1-\lambda)(\lambda+1)(\lambda-1)$

$= -(\lambda+1)(\lambda-1)^2 = 0$

$\lambda = 1, -1$

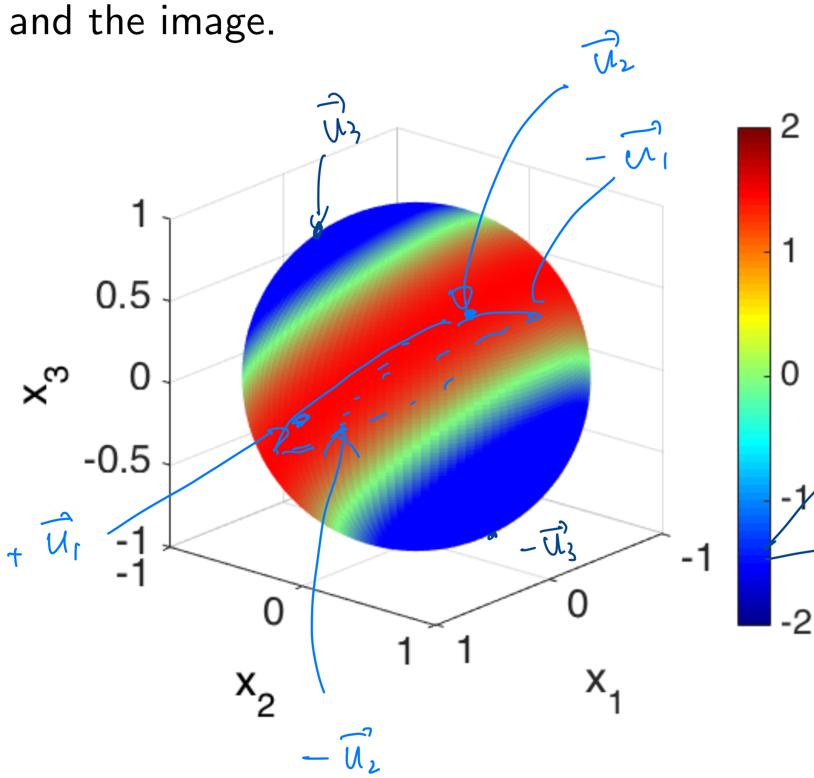
(3) $E_1 = \text{Nul}(A - I) = \text{Nul}\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} = \text{Span}\left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{u_1}, \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{u_2} \right\}$

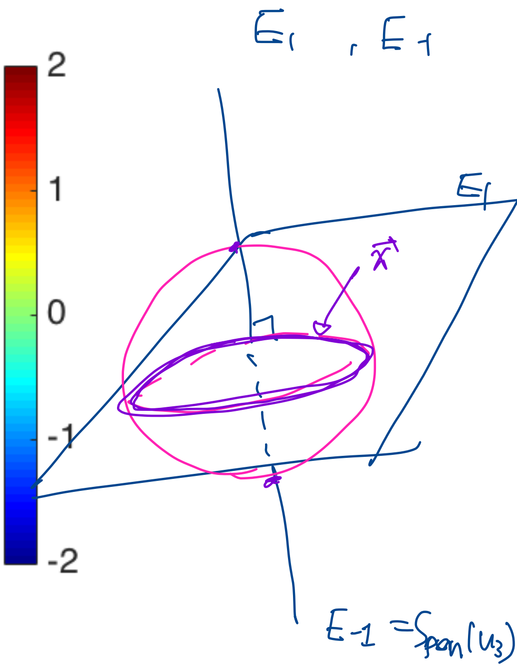
$E_{-1} = \text{Nul}(A + I) = \text{Nul}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \text{Span}\left\{ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}}_{u_3} \right\}$

max $= 1$ when $\vec{x} = \pm\vec{u}_1, \pm\vec{u}_2, a\vec{u}_1 + b\vec{u}_2$ & $a^2 + b^2 = 1$ / min $= -1$ when $\vec{x} = \pm\vec{u}_3$.

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.





$$Q(\vec{x}) = \lambda_1 (x\cdot u_1)^2 + \lambda_2 (x\cdot u_2)^2 + \lambda_3 (x\cdot u_3)^2$$

$$= \left(\underbrace{(x\cdot u_1)^2 + (x\cdot u_2)^2}\right) - (\underline{x\cdot u_3})^2$$

If $x = \pm u_3 \Rightarrow (x\cdot u_1)^2 = (x\cdot u_2)^2 = 0 \Rightarrow$ Min.

If $\vec{x} \in E_1$ & $\underline{\|x\| = 1} \Rightarrow x\cdot u_3 = 0$

$$\Rightarrow \|x\|^2 = (x\cdot u_1)^2 + (x\cdot u_2)^2 = 1$$

$\lambda_1 = \max\{ Q(\vec{x}) : \|x\| = 1\} \underset{?}{=} \max\left\{ \frac{Q(\vec{x})}{\|\vec{x}\|^2} : \vec{x} \neq 0 \right\}$

$Q(k\vec{x}) = k\, Q(\vec{x})$?

$\lambda_n = \min\{ Q(\vec{x}) : \|x\| = 1\} = \min\left\{ \frac{Q(\vec{x})}{\|\vec{x}\|^2} : \vec{x} \neq 0 \right\}$

$\vdots$

$Q(\vec{x}) = \lambda_1 (\underbrace{x \cdot u_1}_{=0})^2 + \lambda_2 (x \cdot u_2)^2 + \cdots + \lambda_n (x \cdot u_n)^2$

$\lambda_2, \lambda_3, \cdots$?

$\Rightarrow \lambda_2 = \max\{ Q(\vec{x}) : \|x\| = 1, \ x \cdot u_1 = 0\}$

An Orthogonality Constraint

$\lambda_3 = \max\{ Q(\vec{x}) : \|x\| = 1, \ x \cdot u_1 = x \cdot u_2 = 0\}$

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

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

$Q(\vec{x}) = \boxed{\lambda_2} (x \cdot \vec{u}_2)^2 + \cdots$

Subject to the constraints $||\vec{x}|| = 1$ and $\boxed{\vec{x} \cdot \vec{u}_1 = 0}$,

- The maximum value of $Q(\vec{x}) = \lambda_2$, attained at $\vec{x} = \pm\vec{u}_2$.
- The minimum value of $Q(\vec{x}) = \lambda_n$, attained at $\vec{x} = \pm\vec{u}_n$.

Note that λ_2 is the second largest eigenvalue of A.

$Q(\vec{x}) = \lambda_1 (\underbrace{\vec{x} \cdot \vec{u}_1}_{=0})^2 + \lambda_2 (\vec{x} \cdot \vec{u}_2)^2 + \cdots + \lambda_n (\vec{x} \cdot \vec{u})^2$

$\max\{ Q(\vec{x}) : \|\vec{x}\| = 1\} = \lambda_1 = \max\left\{ \frac{Q(\vec{x})}{\|x\|^2} : \vec{x} \neq 0 \right\}$

$\min\{ Q(\vec{x}) : \|\vec{x}\| = 1\} = \lambda_n = \min\left\{ \frac{Q(\vec{x})}{\|x\|^2} : \vec{x} \neq 0 \right\}$

Q: $\max\{ Q(\vec{x}) : \|\vec{x}\| = 4\} = 4^2 \cdot \lambda_1$.

unit vector

Suppose $\|x\| = 4$ $\quad \left\| \frac{x}{\|x\|} \right\| = 1$ $\quad Q(\vec{x}) = Q\left(\|\vec{x}\| \cdot \frac{x}{\|x\|}\right) = 4^2 Q\left(\frac{x}{\|x\|}\right)$

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.

$$Q(\vec{x}) = x_1^2 + 2x_2x_3, \qquad \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$\lambda_1 = 1 \qquad u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$\boxed{\lambda_2 = 1} \qquad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

$\lambda_3 = -1 \qquad u_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

ANS = 1

when

$x = \pm u_2$.

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.

$$Q(\vec{x}) = x_1^2 + 2x_2x_3$$

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.

Spectral Thm If A is a real symmetric matrix

then $A = P \cdot D \cdot P^T$ where

where $P = [\underbrace{u_1 \quad u_2 \quad \cdots \quad u_n}_{\text{orthonormal eigenvectors}}]$ and

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_n \end{bmatrix}$$

$$A = \underbrace{[u_1 \quad u_2 \quad \cdots \quad u_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - u_1^T - \\ - u_2^T - \\ \vdots \\ - u_n^T - \end{bmatrix}$$

$$= \lambda_1 \cdot u_1 \cdot u_1^T + \lambda_2 u_2 \cdot u_2^T + \cdots + \lambda_n u_n \cdot u_n^T$$

$\underbrace{u_1}_{n\times 1} \cdot \underbrace{u_1^T}_{1\times n} \in \mathbb{R}^{n\times n}$ $\qquad u_i \cdot u_i^T$: a symmetric matrix.

Goal : $A \in \mathbb{R}^{m\times n}$, $\qquad A = \sigma_1 \cdot u_1 \cdot v_1^T + \sigma_2 u_2 \cdot v_2^T + \cdots$

$A^TA \in \mathbb{R}^{n\times n}$, symmetric $\qquad$ SVD.

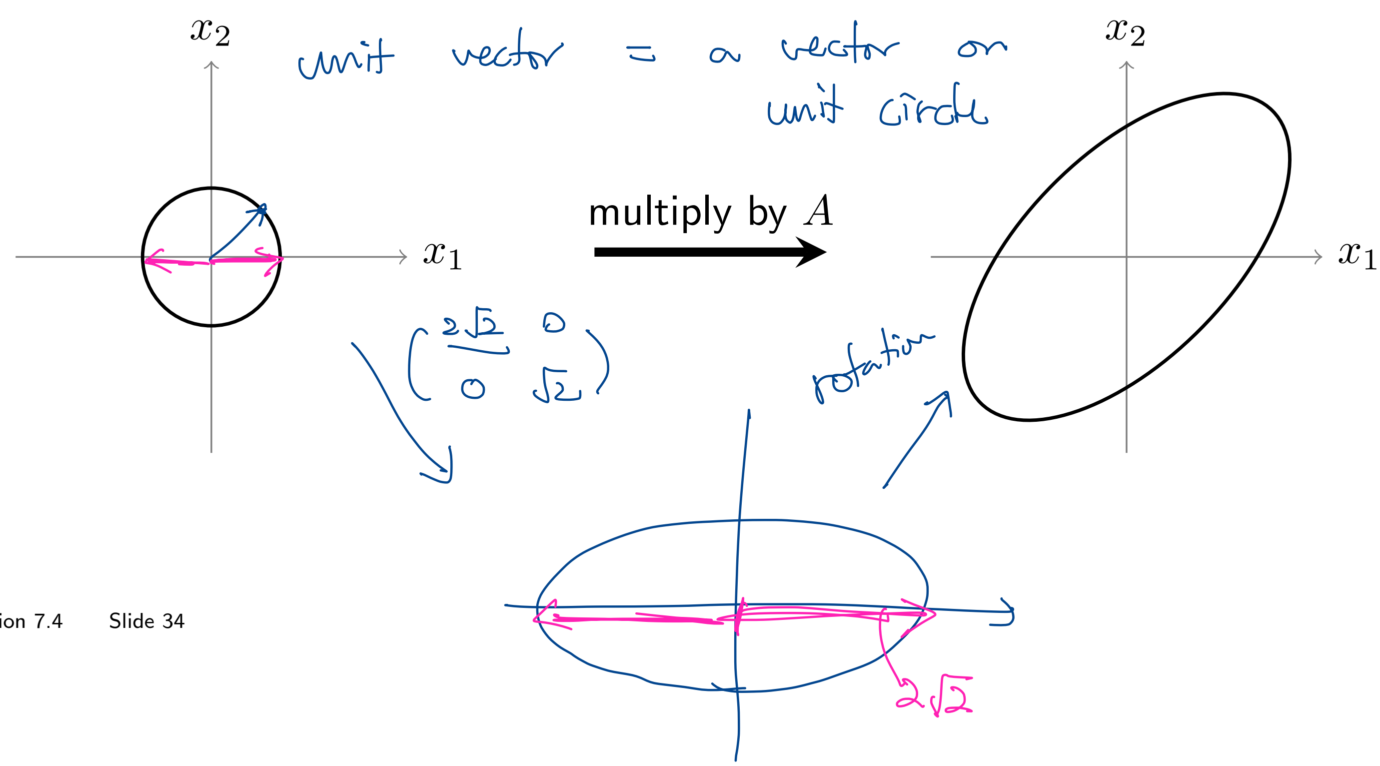
Example 1

The linear transform whose standard matrix is

$$A = \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}}_{\text{orthogonal, rotation}} \underbrace{\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}}_{\text{diagonal}} = \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}$$

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.

unit vector = a vector on unit circle



ANS : If $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ then $||A\vec{x}|| = 2\sqrt{2}$

Example 1 - Solution

$\underline{B = A^T \cdot A}$: $(n \times n)$ symmetric matrix $(A \in \mathbb{R}^{m \times n})$

① $Q_B(\vec{x}) = x^T \cdot B \cdot x = x^T (A^T \cdot A) x$: $\underline{\text{positive semidefinite}}$

(recall : Q is positive semidefinite if $Q(\vec{x}) \geq 0 \quad \forall \vec{x}$)

$Q_B(\vec{x}) = \underbrace{x^T \cdot A^T} \cdot A x = (Ax)^T \cdot (Ax) = \| Ax \|^2 \geq 0$

② $A^T A$ eigenvalues are nonnegative

Singular Values

3. The matrix A^TA 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

$$\|A\vec{v}_j\|^2 = (Av_j)^T \cdot (Av_j) = v_j^T \cdot (A^T \cdot A \cdot v_j) = v_j^T \cdot (\lambda_j v_j)$$
$$= \lambda_j \; v_j^T \cdot v_j = \lambda_j \, \|v_j\|^2 = \lambda_j$$

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

$$v_j \cdot v_k = 0 = v_j^T \cdot v_k$$

$$(Av_j) \cdot (Av_k) = (A\vec{v}_j)^T A\vec{v}_k = v_j^T \cdot (\underbrace{A^T \cdot A \cdot v_k}) = v_j^T \cdot (\lambda_k v_k)$$
$$= \lambda_k \cdot \underline{(v_j^T \cdot v_k)} = 0$$

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.

$\max \{ Q(\vec{x}) : \|\vec{x}\| = 1 \}$ = largest eigenvalue $= \lambda_1$

$\max \{ Q(\vec{x}) : \|\vec{x}\| = 1, \vec{x} \cdot \vec{u_1} = 0 \}$ = second largest eigenvalue

$\vec{u_1}$ eigenvector corresponding to λ_1

$Q(\vec{x}) = \lambda_1 (x \cdot u_1)^2 + \lambda_2 (x \cdot u_2)^2 + \dots + \lambda_n (x \cdot u_n)^2$

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

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

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

$A^TA \in \mathbb{R}^{n \times n}$ symm. $\Rightarrow$ diagonalizable.

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

$\{v_1, v_2, \dots, v_n\}$ orthonormal eigenvectors

$\sigma_i = \sqrt{\lambda_i}$

$A = U \Sigma V^T$

U: $m \times m$, Σ: $m \times n$, V^T: $n \times n$

orthogonal matrices

$A \in \mathbb{R}^{m \times n}$

① $A^T A$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, $\{v_1, \cdots, v_n\}$

② $\Sigma_1 = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}$ ← need to know rank(A)

$\sigma_i = \sqrt{\lambda_i}$

③ $V = \underbrace{[\, v_1 \quad v_2 \quad \cdots \quad v_n \,]}$ make sure that they are orthonormal

④ $U = \underbrace{[\, u_1 \quad u_2 \quad \cdots \quad u_m \,]}$ orthonormal How?

<u>Note</u>

- $\|Av_j\|^2 = (Av_j) \cdot (Av_j) = (Av_j)^T \cdot (Av_j)$
  $= v_j^T \cdot \underbrace{A^T \cdot A v_j}_{= \lambda_j v_j} = \lambda_j \underbrace{v_j^T \cdot v_j}_{=1}$
  $= \lambda_j$

- For $j \neq k$, $Av_j \cdot Av_k = 0$

- $\dim(\mathrm{Col}(A)) = \mathrm{rank}(A) = r$

$\underbrace{\left\{ \underbrace{\frac{1}{\|Av_1\|} Av_1}_{u_1},\ \underbrace{\frac{1}{\|Av_2\|} Av_2}_{u_2},\ \cdots\ \underbrace{\frac{1}{\|Av_r\|} Av_r}_{u_r} \right\}}_{\text{orthogonal}} \subseteq \mathrm{Col}(A)$

$\sigma_1 = \|Av_1\|$, $\sigma_2 = \|Av_2\|$, $\sigma_r = \|Av_r\|$

orthogonal ⇒ lin. indep. ⇒ orthogonal basis

$\left\{ u_1 = \frac{1}{\sigma_1} Av_1, \cdots, u_r = \frac{1}{\sigma_r} Av_r, \underbrace{\qquad\qquad}_{\text{more}} \right\}$ ← Gram Schmidt

$\Rightarrow$ ONB for $\mathbb{R}^m$

$$A = \underline{U \Sigma} V^T$$

Why:

$$U \Sigma V^T = \underbrace{\frac{1}{\cancel{\sigma_1}} A v_1}_{= u_1} \cdot \cancel{\sigma_1} \cdot v_1^T + \frac{1}{\cancel{\sigma_2}} A v_2 \cdot \cancel{\sigma_2} \cdot v_2^T$$

$$+ \cdots + \frac{1}{\cancel{\sigma_r}} A v_r \, \cancel{\sigma_r} \, v_r^T$$

$$= A v_1 \cdot v_1^T + A v_2 \cdot v_2^T + \cdots + A v_r \cdot v_r^T .$$

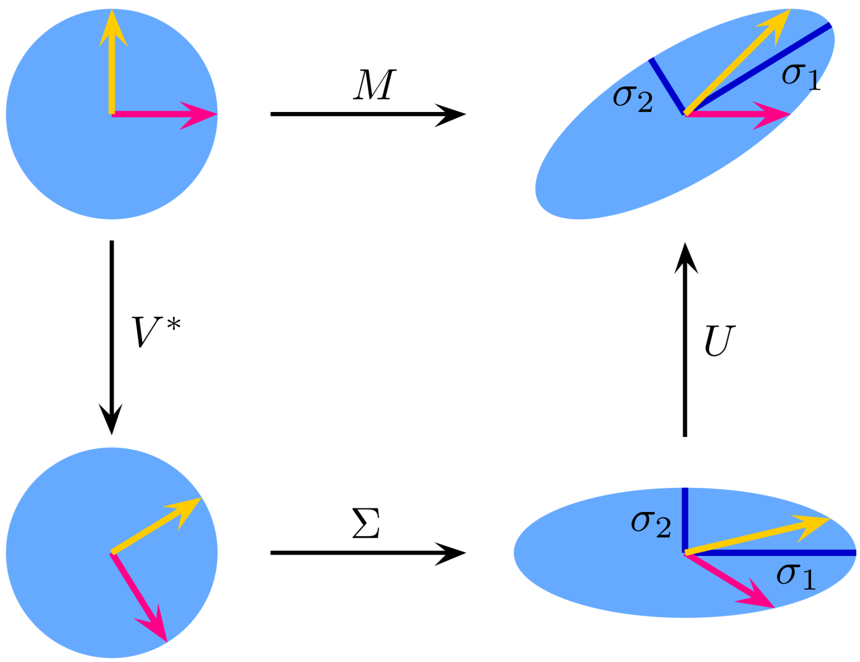
$$= A$$ Goal

$A = B \iff$ For any basis of $\mathbb{R}^n$, $\{v_1, \ldots, v_n\}$

$$A v_i = B v_i$$

check.

$$(U \Sigma V^T) \cdot v_j = (A v_1 \cdot v_1^T + \cdots + A v_r \cdot v_r^T) \cdot v_j \overset{\downarrow}{=} A v_j$$



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

$A = U \cdot \Sigma \cdot V^T$

Algorithm to find the SVD of A

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

$\sigma_i = \sqrt{\lambda_i}$

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

$\Downarrow$ spectral decomposition $\Rightarrow$ $\begin{cases} \lambda_i \geq 0 \\ v_i \text{ orthonormal} \end{cases}$

$\Rightarrow \quad \Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{pmatrix} \in \mathbb{R}^{m \times n}$

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

$\{v_1, v_2, \cdots v_n\}$ orthonormal eigenvectors for A^TA.

$\Rightarrow \quad V = [v_1 \ v_2 \ \cdots \ v_n] \in \mathbb{R}^{n \times n}$ orthogonal

3. Compute an orthonormal basis for 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.

$\{Av_1, Av_2, \cdots, Av_r\}$ orthogonal vectors in $\mathbb{R}^m$

$\|Av_j\|^2 = \lambda_j$

$\Rightarrow \quad \{\frac{1}{\sigma_1} Av_1, \frac{1}{\sigma_2} Av_2, \cdots, \frac{1}{\sigma_r} Av_r\}$: orthonormal vectors in $\mathbb{R}^m$

$\sigma_1 = \sqrt{\lambda_1}$

G-S

$\Rightarrow \quad \{\underbrace{\frac{1}{\sigma_1} Av_1}_{= u_1}, \cdots, \underbrace{\frac{1}{\sigma_r} Av_r}_{= u_r}, u_{r+1}, u_{r+2}, \cdots, u_m\}$ ONB for $\mathbb{R}^m$

$\Rightarrow \quad U = [u_1, u_2 \cdots u_m]$ orthogonal matrix $m \times n$

$\Rightarrow \quad A = U \Sigma V^T$.

Example 2: Write down the singular value decomposition for

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = U \cdot \Sigma' \cdot V^T$$

① $A^T \cdot A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$

$4 = \lambda_1 \to v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad 9 = \lambda_2 \to v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\sigma_1 = \sqrt{4} = 2 \qquad \sigma_2 = \sqrt{9} = 3. \qquad \Sigma' = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

② $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

③ $u_1 = \frac{1}{\sigma_1} A \cdot v_1 = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$u_2 = \frac{1}{\sigma_2} A v_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$

$$\Big\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}}_{\text{orth-normal}}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Big\} \Rightarrow U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

$$\underset{m\times n}{A} = \underset{m\times m}{U} \underset{m\times n}{\Sigma} \underset{n\times n}{V}^T$$

Example 3: Construct the singular value decomposition of

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

(It has rank 1.)

① $A^TA = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$

$\phi(\lambda) = \lambda^2 - 18\lambda = \lambda(\lambda - 18) = 0$ $\quad \lambda_1 = 18$, $\lambda_2 = 0$

$\sigma_1 = \sqrt{18}$, $\sigma_2 = 0$.

$$\Sigma = \begin{bmatrix} \sqrt{18} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

② $\mathrm{Nul}(A^TA - 18I) = \mathrm{Nul}\begin{pmatrix} -9 & -9 \\ -9 & -9 \end{pmatrix} = \mathrm{Span}\left\{ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{= v_1} \right\}$

$\mathrm{Nul}(A^TA - 0\cdot I) = \mathrm{Nul}\begin{pmatrix} 9 & -9 \\ -9 & 9 \end{pmatrix} = \mathrm{Span}\left\{ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{= v_2} \right\}$

$$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

③ $\underset{3\times 3}{U} = [\; \underset{\substack{\| \\ \frac{1}{\sigma_1} A v_1 \\ \| \\ \frac{1}{\sqrt{18}} \cdot \sqrt{2} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}}}{u_1} \quad \underbrace{u_2 \quad u_3}\;]$

$\frac{1}{\sqrt{18}} = \frac{1}{\sqrt{9}}$

$$\begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2 \\ -4 \\ 4 \end{bmatrix} = \sqrt{2} \cdot \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

$$u_1 = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} \qquad u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \qquad u_3 = \frac{1}{\sqrt{18}}\begin{bmatrix} -4 \\ -1 \\ 1 \end{bmatrix}$$

orthogonal

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 1^{st} *two items in the list.*

ONB for Col(A)

ONB for $\mathrm{Col}(A)^{\perp} = \mathrm{Nul}(A^T)$

$$U = [\overbrace{u_1, \dots, u_r}, \underbrace{u_{r+1}, \dots, u_n}]$$

$$u_1 = \frac{1}{\sigma_1} A v_1 \quad \dots \quad u_r = \frac{1}{\sigma_r} A v_r$$

$$\underbrace{\frac{1}{\sigma_1} A v_1 \;\dots\; \frac{1}{\sigma_r} A v_r}_{\in \mathrm{Col}(A)}$$

By Gram-Schmidt

SVD $A = U \cdot \Sigma \cdot V^T$, $\text{rank}(A) = r \leq n$

A^TA : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, $\{v_1, \cdots, v_n\}$ Orthonormal eigenvectors

(Note : $\text{rank}(A) = \text{rank}(A^TA)$) $\Rightarrow$ $\lambda_1 \geq \cdots \geq \lambda_r > 0$ & $\lambda_{r+1} = \cdots = \lambda_n = 0$)

$\sigma_i = \sqrt{\lambda_i}$ $\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$ $V = [\underbrace{v_1, \cdots v_r}_{\text{ONB for } \text{Nul}(A)^\perp = \text{Col}(A^T) = \text{Row}(A)} \; \underbrace{v_{r+1} \cdots v_n}_{\text{ONB for Nul}(A)}]$

$A \cdot v_j = 0 \cdot v_j$ $j \geq r+1$

$v_j \in \text{Nul}(A)$

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.

$\left(V^* \overset{def}{=} \bar{V}^T \right)$

Example 4

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

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

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

$$V^T = \mathbf{V}^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix} \begin{matrix} \leftarrow \lambda_1 = 16 \\ \leftarrow \lambda_2 = 9 \\ \leftarrow \lambda_3 = 5 \\ \lambda = 0. \end{matrix}$$

A^TA.

$\sigma_i = \sqrt{\lambda_i}$

① Rank$(A) = 3$

② ONB for $\underbrace{\text{Nul}(A)}$

ONB for Nul(A).

$$= \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\sqrt{0.8} \\ 0 \\ 0 \\ 0 \\ \sqrt{0.2} \end{bmatrix} \right\}$$

③ ONB Col$(A)^{\perp}$ $= \left\{ \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \right\}$

ONB for Col(A) $\quad \in$ Col$(A)^{\perp}$

$$U = [\, u_1 \quad u_2 \quad u_3 \quad \underline{u_4} \,]$$

$u_1 = \frac{1}{4}Av_1$, $u_2 = \frac{1}{3}Av_2$, $u_3 = \frac{1}{\sqrt{5}}Av_3$

ONB.

Example 4 - Solution

The Four Fundamental Spaces

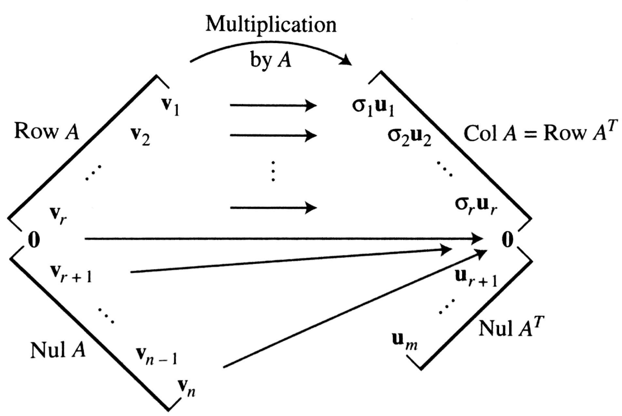


FIGURE 4 The four fundamental subspaces and the action of A.

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

$A = \sigma_1 \cdot \underset{\overset{\shortparallel}{\frac{1}{\sigma_1} A v_1}}{u_1} \cdot v_1^T + \sigma_2 \underset{\overset{\shortparallel}{\frac{1}{\sigma_2} A v_2}}{u_2} v_2^T + \cdots + \sigma_r \underset{\overset{\shortparallel}{\frac{1}{\sigma_r} A v_r}}{u_r} v_r^T$

$= A v_1 \cdot v_1^T + A v_2 \cdot v_2^T + \cdots + A v_r \cdot v_r^T.$

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 s^{th} 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.