

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

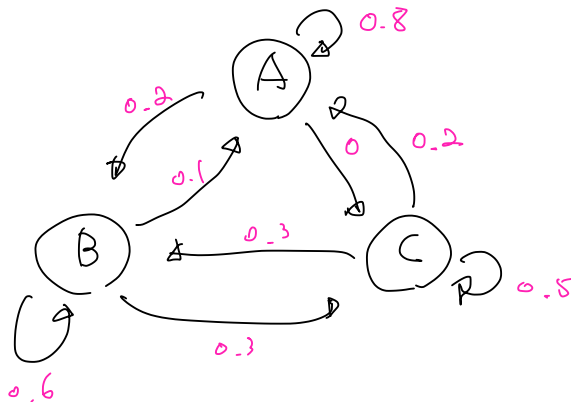
Recall the car rental problem from our Section 4.9 lecture.

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?



Transition matrix

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

Section 10.2 Slide 4

$$\vec{x}_0 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$$

$$\vec{x}_1 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = P \cdot \vec{x}_0$$

↑
after 1 week

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:

$$\vec{x}_2 = \underbrace{P}_{\leftarrow} \underbrace{\vec{x}_1}_{\leftarrow} = \underbrace{PP}_{\leftarrow} \vec{x}_0 = P^2 \vec{x}_0$$

The distribution of cars after n weeks is:

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

$$Q : \quad \lim_{n \rightarrow \infty} \vec{x}_n = ? \quad \lim_{n \rightarrow \infty} P^n = ?$$

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 $\vec{q} = P\vec{q}$.
Probability vector
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

fact P is stochastic $\Rightarrow P^k$ is stochastic. Every column is a probability vector

Definition We say P is regular if there exists $k \geq 1$ an integer s.t. P^k has no zero entries

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

$$\lim_{n \rightarrow \infty} P^n = \Pi = [\vec{q} \quad \vec{q} \quad \dots \quad \vec{q}]$$

2. Each column of Π is the same probability vector \vec{q} .
3. For any initial probability vector \vec{x}_0 ,

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

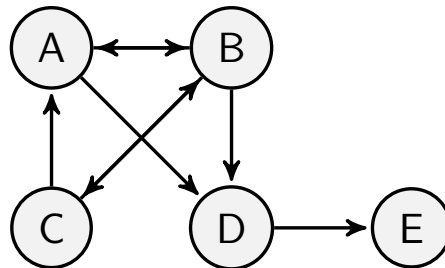
doesn't depend on the choice of \vec{x}_0

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.

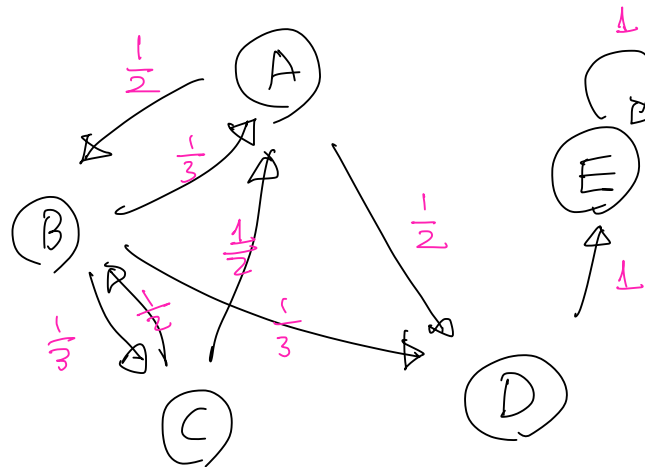
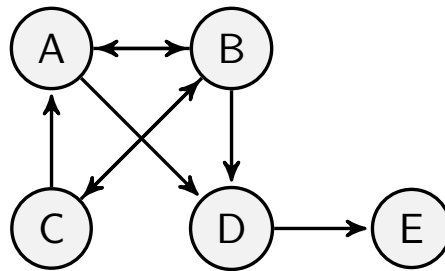
We make two assumptions:

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

Solution

Use the assumptions on the previous slide to construct a Markov chain that represents how users navigate the web.



Section 10.2 Slide 9

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

↖ Not Regular.

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)

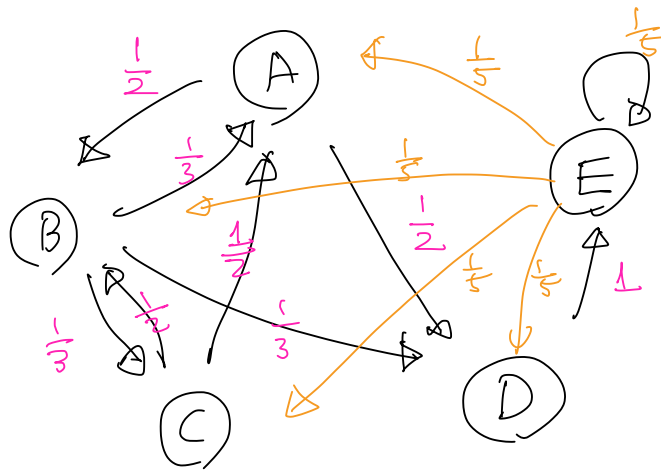
Is the transition matrix in Example 1 a regular matrix?

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:



Section 10.2 Slide 11

$$P_* = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{5} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 1 & \frac{1}{5} \end{bmatrix}$$

← Regular.

We know there exists a unique \vec{q} .

Q: How to find \vec{q} ?

$$P\vec{q} = \vec{q} \Rightarrow (P-I)\vec{q} = \vec{0}$$

Approximation of \vec{q}

$$P^n = [\vec{u}_1 \dots \vec{u}_m]$$

SS \vec{q}

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

$$G = pP_* + (1-p)K$$

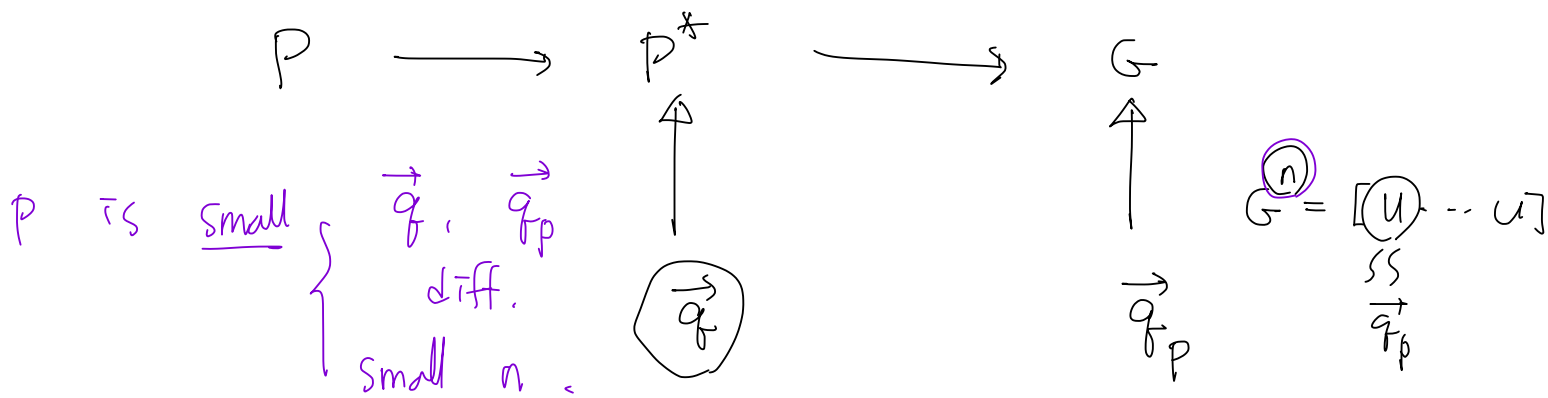
All the elements of the $n \times n$ matrix K are equal to $1/n$.

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:

$$G = \begin{matrix} p \\ 0.85 \end{matrix} \cdot \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{5} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 1 & \frac{1}{5} \end{bmatrix} + \begin{matrix} (1-p) \\ 0.15 \end{matrix} \cdot \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \dots & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \dots & \frac{1}{5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{5} & \frac{1}{5} & \dots & \frac{1}{5} \end{bmatrix}$$

P_*
 K



Computing Page Rank

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

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

- 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 fractions instead of decimal expansions

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=
```

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

```
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=
```

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

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

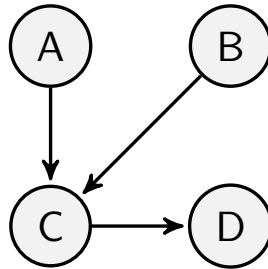
$$\begin{pmatrix} .8 & .1 & .2 \\ .2 & .6 & .3 \\ .0 & .3 & .5 \end{pmatrix}^{10}$$

- 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

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

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

$$A = [*] = A^T$$

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

$$\parallel$$

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

$$C = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = C^T$$

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

$$\neq$$

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

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

$$\neq$$

$$3 \times 2$$

$$F = \begin{bmatrix} 4 & 2 & 0 & 1 \\ 2 & 0 & 7 & 4 \\ 0 & 7 & 6 & 0 \\ 1 & 4 & 0 & 3 \end{bmatrix}$$

Symm? Yes.

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

$$2 \times 3$$

$A^T A$ is Symmetric

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

$$\textcircled{1} \quad A^T A = \begin{bmatrix} \text{---} & a_1^T & \text{---} \\ \text{---} & a_2^T & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & a_n^T & \text{---} \end{bmatrix} \begin{bmatrix} | & | & \dots & | \\ a_1 & a_2 & \dots & a_n \\ | & | & \dots & | \end{bmatrix}$$

$$\vec{a}_1 \cdot \vec{a}_1 = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \dots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \dots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \dots & a_n^T a_n \end{bmatrix}$$

Entries are the **dot products** of columns of A

Symm. $x \cdot y = y \cdot x$

$$\textcircled{2} \quad (A^T A)^T = A^T \cdot (A^T)^T = A^T \cdot A$$

$$\textcircled{3} \quad \underbrace{A^T}_{n \times m} \cdot \underbrace{A}_{m \times n} \in \mathbb{R}^{n \times n} \text{ symm. only for } A \in \mathbb{R}^{n \times n} ?$$

Section 7.1

Slide 4

No. for $A \in \mathbb{R}^{m \times n}$

$$\frac{1}{2} (A^T + (A^T)^T)$$

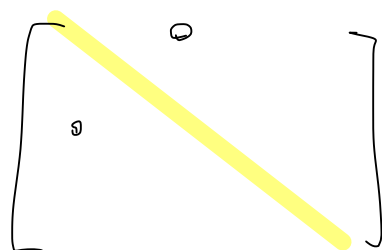
Ex

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

$$B = \frac{A + A^T}{2}$$

$$B^T = \frac{1}{2} (A^T + A)^T$$

$$= B$$



$$\Rightarrow A = \underbrace{\left(\frac{A + A^T}{2} \right)}_{\text{Symm.}} + \underbrace{\left(\frac{A - A^T}{2} \right)}_{C^T = -C}$$

$$C^T = \left(\frac{A - A^T}{2} \right)^T = \frac{A^T - (A^T)^T}{2} = \frac{A^T - A}{2}$$

$$= - \frac{A - A^T}{2} = -C$$

Recall

+ Symm.
 $A \in \mathbb{R}^{n \times n}$

v_1, v_2 eigenvectors with λ_1, λ_2

If $\lambda_1 \neq \lambda_2$, then $\{v_1, v_2\}$ is lin. indep.

+ \checkmark ?

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.

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

Proof:

$$\begin{cases} Av_1 = \lambda_1 v_1 \\ Av_2 = \lambda_2 v_2 \end{cases}$$

$$A^T = A$$

$$\lambda_1 \neq \lambda_2$$

WANT:

$$v_1 \cdot v_2 = 0$$

$$\lambda_1 (v_1 \cdot v_2) = (\lambda_1 v_1) \cdot v_2 = (Av_1) \cdot v_2$$

$$= (Av_1)^T \cdot v_2 = v_1^T \cdot A^T \cdot v_2$$

$$= v_1^T \cdot (A \cdot v_2) = v_1 \cdot (\lambda_2 v_2) = \lambda_2 (v_1 \cdot v_2)$$

$$(\lambda_1 - \lambda_2) \cdot (v_1 \cdot v_2) = 0 \Rightarrow v_1 \cdot v_2 = 0$$

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

$$P^T P = I$$

Recall P is orthogonal if P has orthonormal columns and P is square
 if $P^T = P^{-1}$

Example 1

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

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

Hint: Gram-Schmidt

$$A = \underset{\substack{\uparrow \\ \text{orthogonal}}}{P} \cdot \overset{\substack{\downarrow \\ \text{diagonal}}}{D} \cdot P^{-1} = P \cdot D \cdot P^T$$

$$\lambda = 1: \quad E_1 = \text{Null}(A - I) = \text{Null} \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \rightarrow \begin{matrix} -x_1 + x_3 = 0 \\ x_1 + x_3 = 0 \\ x_2 = 0 \end{matrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = x_2 v_1 + x_3 v_2$$

$$\lambda = -1: \quad E_{-1} = \text{Null}(A + I) = \text{Null} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{matrix} x_1 + x_3 = 0 \\ x_2 = 0 \end{matrix}$$

Section 7.1 Slide 6

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = x_3 v_3$$

$\{v_1, v_2, v_3\}$ eigenvectors, a basis for $\mathbb{R}^3 \Rightarrow$ Diagonalizable
 $\{u_1, u_2, u_3\}$ ONB.

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad u_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$A = P \cdot D \cdot P^T = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

u_1 u_2 u_3

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 dimension of each eigenspace is full, that is its dimension is **equal to** its algebraic multiplicity.
3. The eigenspaces are mutually orthogonal.
4. A can be diagonalized: $A = PDP^T$, where D is diagonal and P is orthogonal.

Proof (if time permits):

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\textcircled{1} \quad f(x_1, x_2, \dots, x_n) = \vec{x} \cdot (A\vec{x})$$

Maximize f subject to $\|\vec{x}\| = 1$.

Lagrange multi.

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

Let A be $n \times n$ symmetric.

Definition

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

$$Q_A(\vec{x}) = \underbrace{\vec{x}}_{(1 \times n)} \underbrace{A}_{n \times n} \underbrace{\vec{x}}_{n \times 1} = [x_1 \ x_2 \ \cdots \ x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

input

$= \vec{x} \cdot (A\vec{x})$

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

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

Example 1

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

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

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

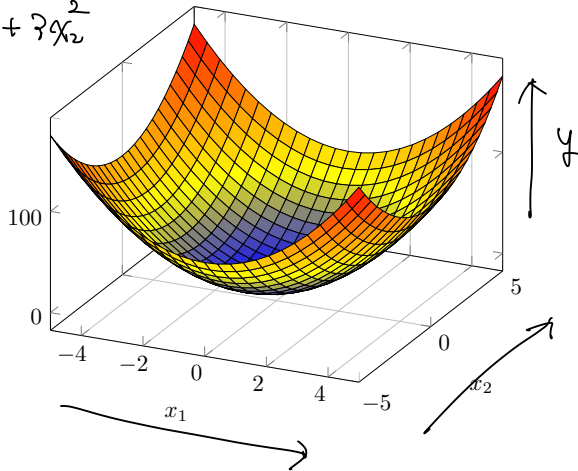
no $x_1 \cdot x_2$ term

$$\begin{aligned} Q_B(\vec{x}) &= \vec{x}^T \cdot B \cdot \vec{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} 4x_1 + x_2 & x_1 - 3x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= (4x_1 + x_2) \cdot x_1 + (x_1 - 3x_2) \cdot x_2 = 4x_1^2 + x_1x_2 + x_1x_2 - 3x_2^2 \\ &= 4x_1^2 + 2x_1x_2 - 3x_2^2. \end{aligned}$$

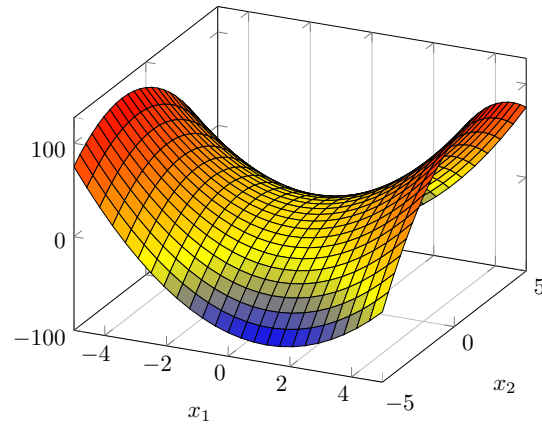
Example 1 - Surface Plots

The surfaces for Example 1 are shown below.

$$y = 4x_1^2 + 3x_2^2$$



$$y = 4x_1^2 + 2x_1x_2 - 3x_2^2$$



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 $\vec{x} \in \mathbb{R}^3$.

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

$$= \vec{x}^T \cdot A \cdot \vec{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \left[x_1 \cdot a_{11} + x_2 \cdot a_{21} + x_3 \cdot a_{31}, \quad x_1 \cdot a_{12} + x_2 \cdot a_{22} + x_3 \cdot a_{32}, \quad x_1 \cdot a_{13} + x_2 \cdot a_{23} + x_3 \cdot a_{33} \right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= x_1 \cdot a_{11} \cdot x_1 + x_2 \cdot a_{21} \cdot x_1 + x_3 \cdot a_{31} \cdot x_1 + x_1 \cdot a_{12} \cdot x_2 + x_2 \cdot a_{22} \cdot x_2 + x_3 \cdot a_{32} \cdot x_2$$

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

$$= \underbrace{a_{11}}_5 x_1^2 + \underbrace{a_{22}}_{-1} x_2^2 + \underbrace{a_{33}}_3 x_3^2 + 2 \underbrace{a_{12}}_0 x_1 x_2 + 2 \underbrace{a_{23}}_{-6} x_2 \cdot x_3 + 2 \underbrace{a_{31}}_3 x_3 \cdot x_1$$

$$A = \begin{bmatrix} 5 & 0 & 3 \\ 0 & -1 & -6 \\ 3 & -6 & 3 \end{bmatrix}$$

$n \times n$ invertible P

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:~~

$$\text{Ex} \quad P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad P^{-1} = P^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\begin{aligned} \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= P^{-1} \cdot \vec{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} (x_1 + x_2) \\ \frac{1}{\sqrt{2}} (x_1 - x_2) \end{bmatrix}. \end{aligned}$$

$$\begin{aligned} \text{Ex} \quad Q(\vec{x}) &= x_1^2 - 2x_1 x_2 + 2x_2^2 \geq 0 \quad \text{for all } x_1, x_2, \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

Section 7.2 Slide 14

$$\begin{aligned} &= (x_1^2 - 2x_1 x_2 + x_2^2) + x_2^2 \\ &= \underbrace{(x_1 - x_2)^2}_{y_1} + \underbrace{x_2^2}_{y_2} \geq 0 \end{aligned}$$

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

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.

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

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

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

$$Q_A(\vec{x}) = \vec{x}^T \cdot A \cdot \vec{x} = \underbrace{(\vec{x}^T \cdot P)}_{(P^T \cdot x)^T} \cdot D \cdot \underbrace{(P^T \cdot x)}_y = P^T \cdot x$$

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

$$Q_A(\vec{x}) = 3x_1^2 + 4x_1x_2 + 6x_2^2$$

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

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

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

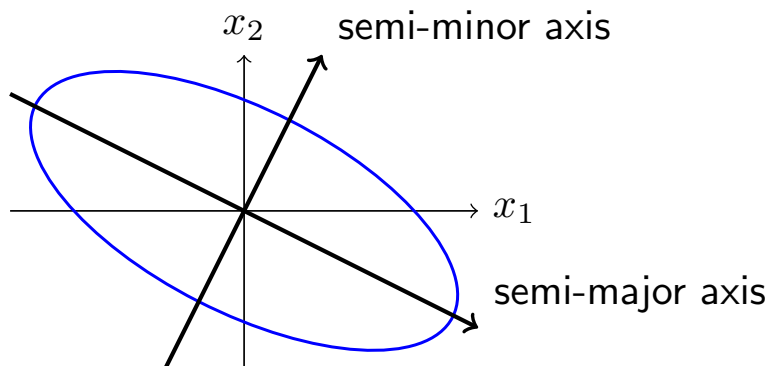
Theorem

If A is a symm 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.



$$= 4 \cdot y_1^2 + 9 \cdot y_2^2$$

$$Q(\vec{x}) = 5x_1^2 + 4x_1x_2 + 8x_2^2$$

$$\textcircled{1} \quad \lambda : \quad \det(A - \lambda I) = \lambda^2 - 13\lambda + 36 = 0$$

$$\lambda = 4, 9$$

Section 7.2 Slide 18

$$\textcircled{2} \quad E_4 = \text{Null} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \text{Span} \left\{ \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$$

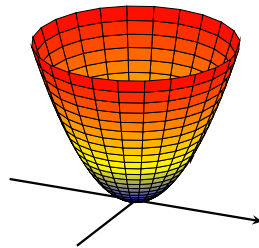
$$E_9 = E_4^\perp = \text{Span} \left\{ \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

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

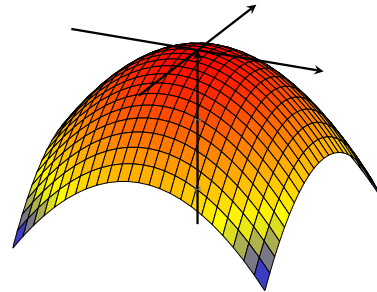
Classifying Quadratic Forms

$$\underline{x_1^2 - x_2^2}$$

$$Q = x_1^2 + x_2^2$$



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



Note: $Q(\vec{0}) = 0$

Definition

A quadratic form Q is

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

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 > 0
2. negative definite iff λ_i < 0
3. indefinite iff λ_i > 0
 < 0

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.

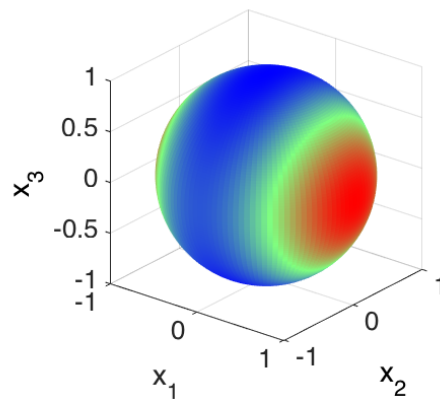
Example 1

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

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

and associated **normalized** eigenvectors

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

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

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.

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

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

$$\det(A - \lambda I) = (1 - \lambda) \det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} \\ = (1 - \lambda)(\lambda^2 - 1)$$

$$\lambda = 1, -1$$

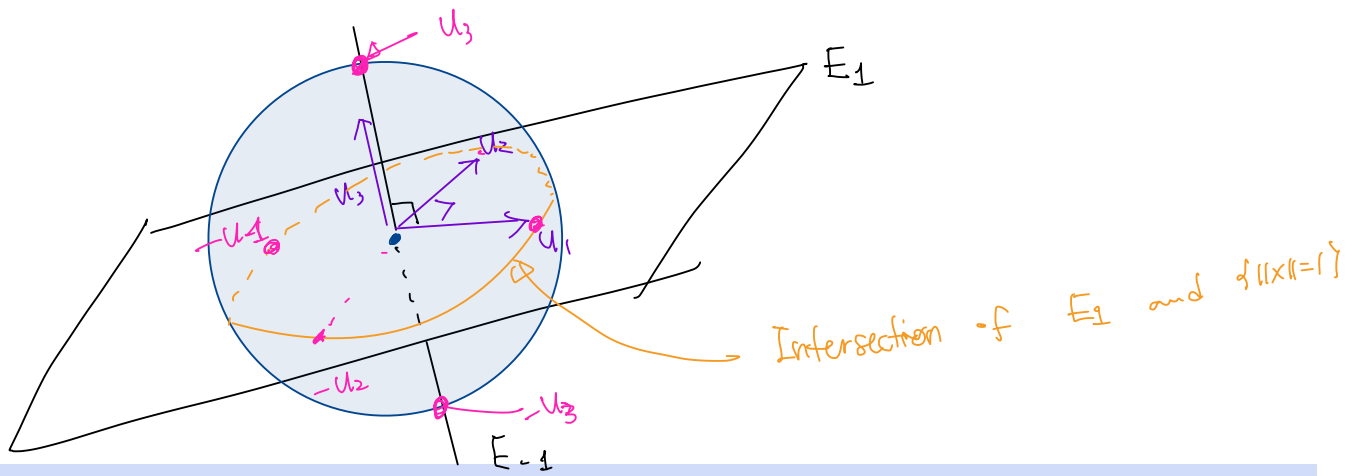
$$E_1 = \text{Null}(A - I) = \text{Null} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} = \text{Span} \left\{ \begin{matrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \end{matrix} \right\} \\ \begin{matrix} \text{u}_1 \\ \text{u}_2 \end{matrix}$$

$$E_{-1} = \text{Null}(A + I) = \text{Null} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \text{Span} \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\} \\ \text{u}_3$$

Section 7.3 Slide 27

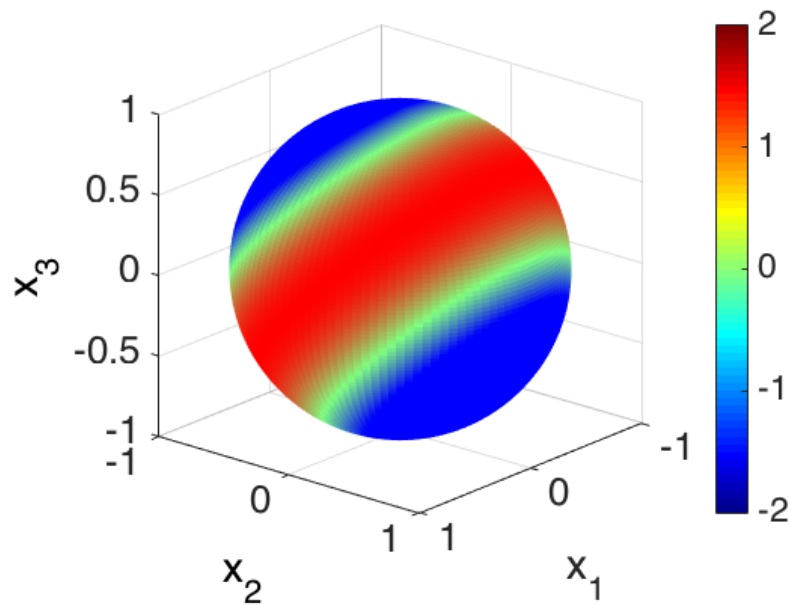
$$\text{Max} \{ Q(\vec{x}) : \|\vec{x}\| = 1 \} = 1 \\ \text{when } \vec{x} = \pm \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \pm \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{Min} \{ Q(\vec{x}) : \|\vec{x}\| = 1 \} = -1 \\ \text{when } \vec{x} = \pm \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$



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

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

and associated eigenvectors

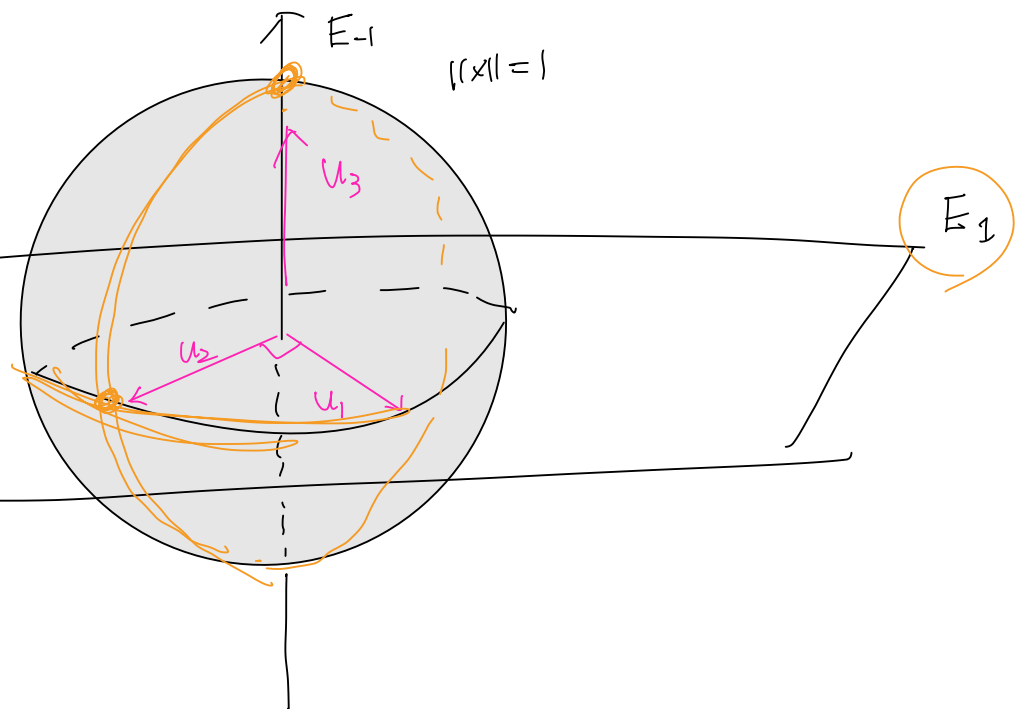
normalized.

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

Subject to the constraints $\|\vec{x}\| = 1$ and $\vec{x} \cdot \vec{u}_1 = 0$, *$\vec{x} \cdot \vec{u}_2 = 0$*

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

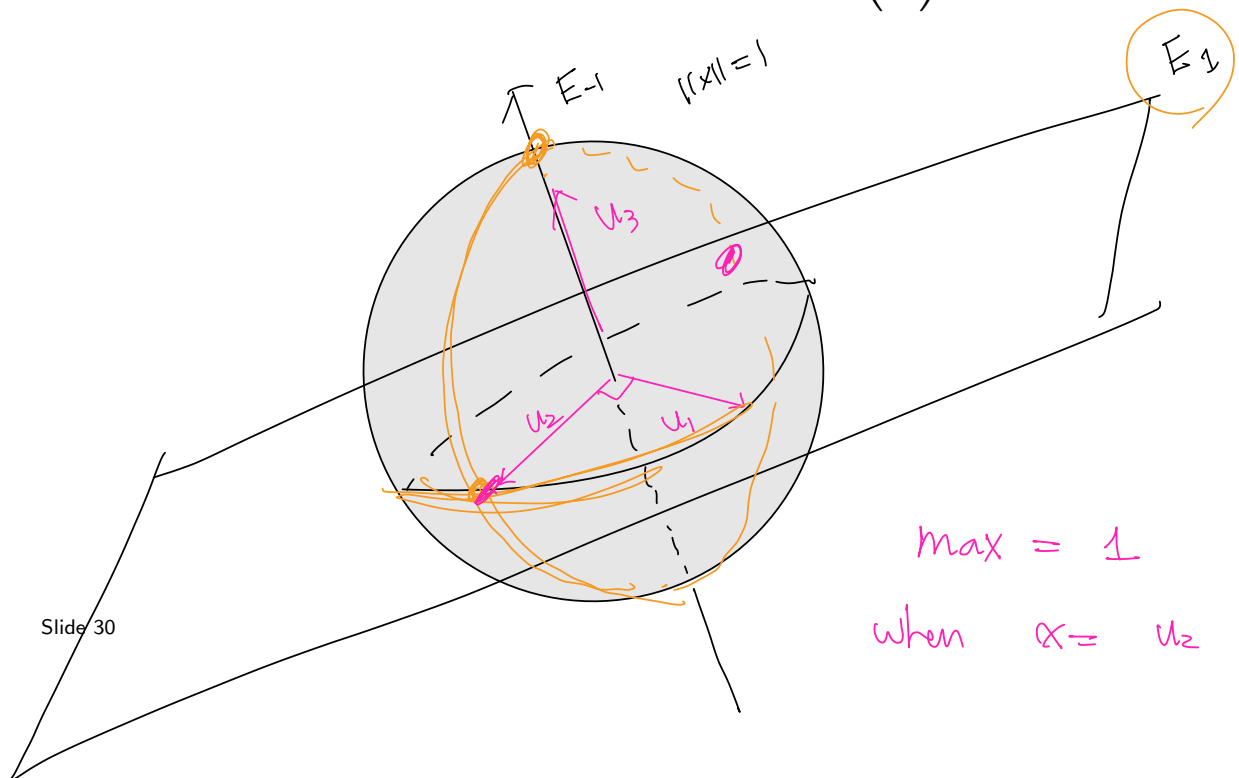


Example 3

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.

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

$\lambda_1 \geq \lambda_2 \geq \lambda_3$
 $1 \quad 1 \quad -1$
 $\quad \quad \quad \textcircled{1} \quad \quad -1$



- ① ³ $\sqrt{5}$, ② ¹¹ 5 , ③ ¹ 5^2 , ④ ¹ $\frac{1}{5}$, ⑤ ¹ ∞

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.

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

$$\begin{aligned} \vec{x} &= \|\vec{x}\| \cdot \left(\frac{\vec{x}}{\|\vec{x}\|} \right) \\ &= \|\vec{x}\| \cdot \vec{u} \\ &\quad \uparrow \\ &\quad \text{unit.} \end{aligned}$$

$$Q(\vec{x}) = \vec{x} \cdot (A \vec{x})$$

$$= (\|\vec{x}\| \cdot \vec{u}) \cdot (A (\|\vec{x}\| \cdot \vec{u}))$$

$$= \underbrace{\|\vec{x}\|^2} \cdot \underbrace{(\vec{u} \cdot (A \vec{u}))}$$

$$\begin{aligned} \max &= 1 \\ \min &= -1 \end{aligned}$$

$A : n \times n$ *Symm.*

$A = P \cdot D \cdot P^T = \underbrace{[u_1 \ u_2 \ \dots \ u_n]}_{\substack{\text{eigenvectors} \\ \text{orthonormal}}} \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix}$

\uparrow *orthogonal* \nwarrow *diagonal*

$= \lambda_1 \cdot \underbrace{u_1 \cdot u_1^T}_{\substack{n \times 1 \\ 1 \times n}} + \lambda_2 u_2 \cdot u_2^T + \dots + \lambda_n u_n \cdot u_n^T$

\swarrow *n x n matrix, symmetric*

$u_i^T \cdot u_i = u_i \cdot u_i = 1$

$\textcircled{1 \times n} \quad \textcircled{n \times 1}$

$\text{Col}(u_i \cdot u_i^T) = \text{Span}\{u_i\}$

Q: The rank of $u_i \cdot u_i^T = ?$ $\textcircled{=1}$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} [1 \ 2 \ 3] = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

In general $u \cdot v^T \leftarrow$ rank - 1 matrix

7.4 $A : m \times n$ general

Goal: $A = \sigma_1 \cdot u_1 \cdot v_1^T + \sigma_2 u_2 \cdot v_2^T + \dots + \sigma_r u_r \cdot v_r^T$

Singular Value Decomposition (SVD)

Idea: $A^T A : n \times n$ *Symm.* $\left\{ \begin{array}{l} \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0 \text{ why?} \\ v_1 \quad v_2 \quad \dots \quad v_n \text{ (ONB)} \end{array} \right.$

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

$= \|A x\|^2 \geq 0 \Rightarrow A^T A$ is positive semidefinite

$\Rightarrow \lambda_i \geq 0 \quad \forall i=1, \dots, n$

$\{A v_1, A v_2, \dots, A v_n\}$ Orthogonal.

$A v_i \cdot A v_j = (A v_i)^T \cdot A v_j = v_i^T \cdot (A^T A v_j) = v_i^T \cdot \lambda_j v_j$

$= \lambda_j \cdot v_i \cdot v_j = 0$

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

orthogonal rotation

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

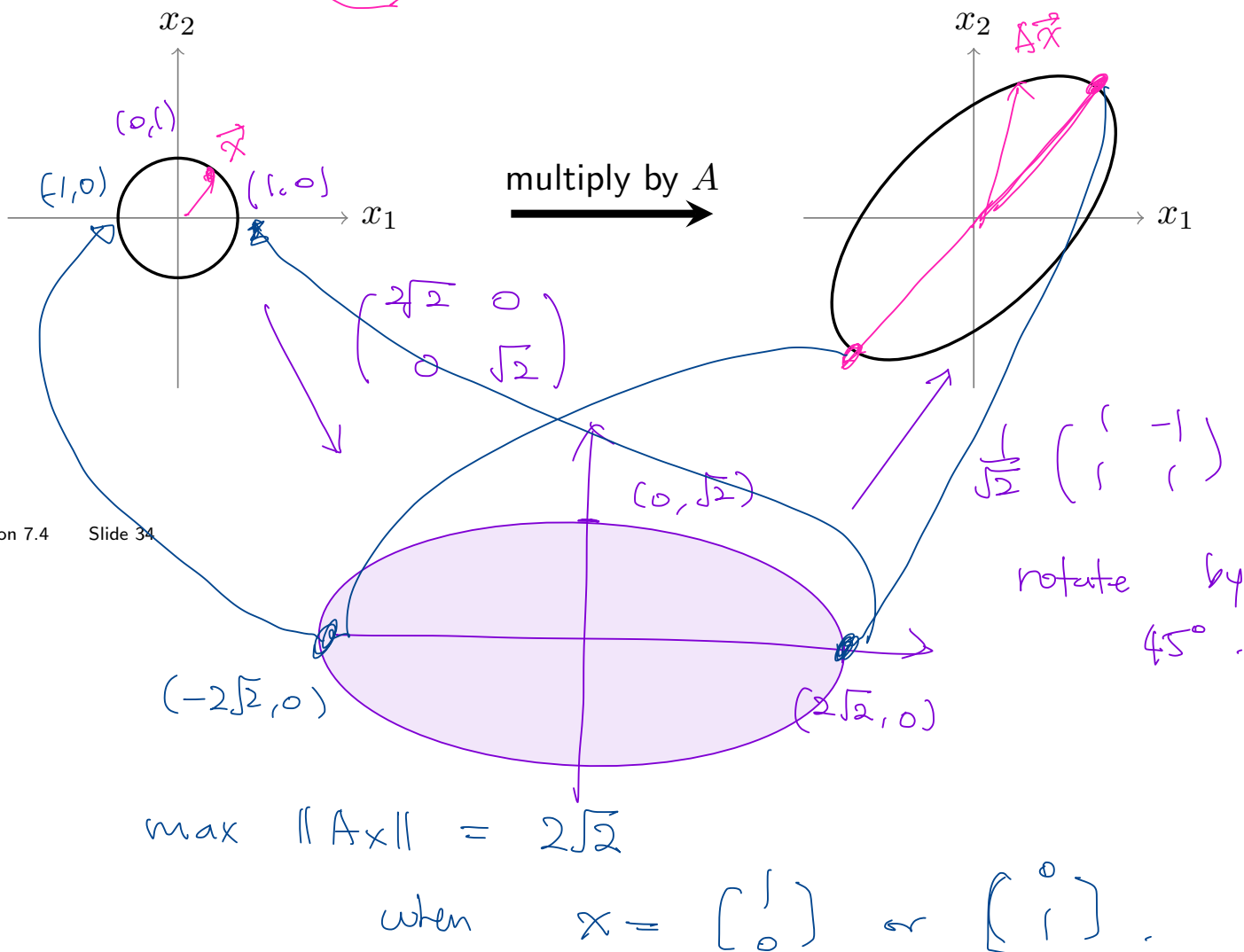
$$\theta = \frac{\pi}{4}$$

The linear transform whose standard matrix is

centered at $\vec{0}$
radius=1

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

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 $\{\vec{v}_1, \dots, \vec{v}_n\}$ be the associated orthonormal eigenvectors. Then

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

If the A has rank r , then $\{A\vec{v}_1, \dots, A\vec{v}_r\}$ 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 .

$$\dim(\text{Col}(A)) = \dim(\text{Row}(A))$$

The SVD = Singular Value Decomposition.

$$\text{rank}(A) = \dim(\text{Col}(A)) \leq \underline{n \text{ and } m}$$

Theorem: Singular Value Decomposition

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

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

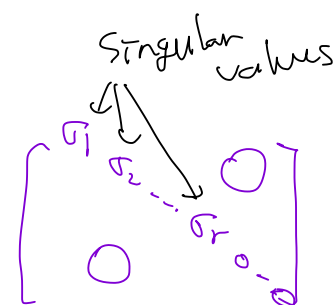
U is a $m \times m$ orthogonal matrix, and V is a $n \times n$ orthogonal matrix.

A : $m \times n$ matrix, $\text{rank}(A) = r$

$$\Rightarrow A = \begin{matrix} m \times n \\ U \end{matrix} \cdot \begin{matrix} m \times m & m \times n \\ \Sigma \end{matrix} \cdot \begin{matrix} n \times n \\ V^T \end{matrix}$$

↑ Orthogonal

diag.



$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$$

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

Step 1 : $A^T \cdot A$ $n \times n$, symm ($(A^T A)^T = A^T A$)
positive semidefinite

Γ proof. Need: $Q(x) = x \cdot (A^T A x) \geq 0$ for all x

$$\begin{aligned} x \cdot (A^T A x) &= (x^T \cdot A^T)(A x) = (A x)^T \cdot A x \\ &= A x \cdot A x = \|A x\|^2 \geq 0 \text{ for all } x \end{aligned}$$

Step 2 : $A^T A$

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

$$v_1, v_2, \dots, v_n$$

orthonormal eigenvector
because $A^T A$ is p.s.d
because $A^T A$ is symm .

$\{ A v_1, A v_2, \dots, A v_n \}$ orthogonal

Γ proof.

$$A v_i \cdot A v_j = (A v_i)^T \cdot A v_j$$

$$\begin{aligned} i, j = 1, \dots, n \\ &= v_i^T \cdot (A^T \cdot A \cdot v_j) = \lambda_j v_j \\ &= \lambda_j v_i^T \cdot v_j \end{aligned}$$

$$= \lambda_j \underbrace{v_i \cdot v_j}_0 = 0 \quad \text{if } i \neq j$$

↑ because orthonormal

If $i = j$ $\|A v_i\|^2 = \lambda_i \cdot \|v_i\|^2 = \lambda_i$

$$\|A v_i\| = \sqrt{\lambda_i} = \sigma_i$$

$$\text{rank}(A) = r \leq n, m$$

$\left\{ \underbrace{\frac{1}{\sigma_1} A v_1}_{u_1}, \dots, \underbrace{\frac{1}{\sigma_r} A v_r}_{u_r} \right\}$: orthonormal basis for

$$\text{Col}(A) \subseteq \mathbb{R}^m$$

Extend

$\{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$: ONB for \mathbb{R}^m

Step 3

$$A v_1 = \sigma_1 u_1$$

$$A v_2 = \sigma_2 u_2$$

\vdots

$$A v_r = \sigma_r u_r$$

$$A v_{r+1} = 0$$

\vdots

$$A v_n = 0$$

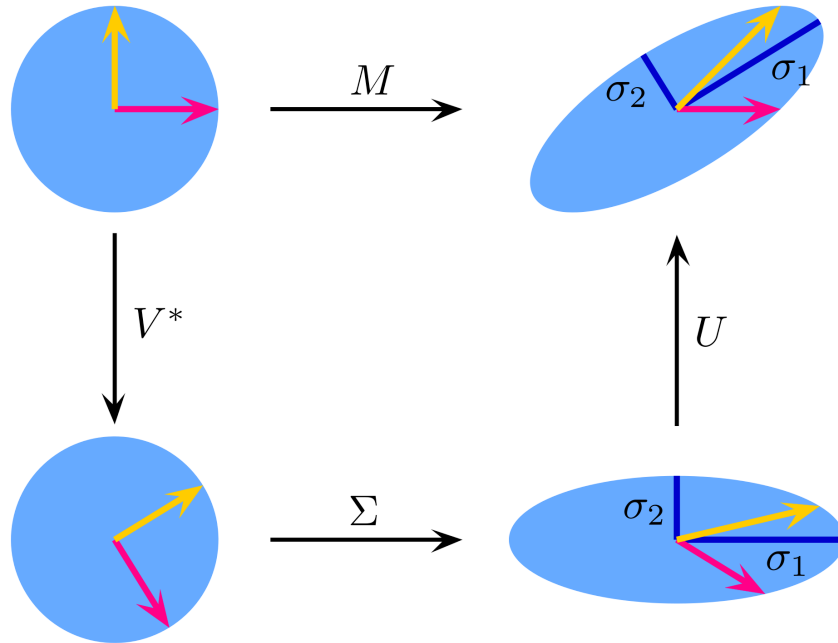
$$(\sigma_{r+1} = \dots = \sigma_n = 0)$$

$$A \underbrace{\begin{bmatrix} v_1 & v_2 & \dots & v_r & \dots & v_n \end{bmatrix}}_V = \underbrace{\begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_r & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}}_\Sigma$$

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

$$A V V^T = U \Sigma V^{-1}$$

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



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

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

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

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

Example 2: Write down the singular value decomposition for

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{matrix} u_1 & u_2 & u_3 & u_4 \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \begin{matrix} \\ \\ \\ \\ \end{matrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = V^T$$

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

$$\lambda_1 = 9 \geq \lambda_2 = 4$$

$$\sigma_1 = 3 \quad \sigma_2 = 2$$

↓

$$v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow V = \begin{bmatrix} v_1 & v_2 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}$$

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

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

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Singular Value Decomposition

$$A \in \mathbb{R}^{m \times n}, \quad A = \overset{m \times m}{U} \cdot \overset{m \times n}{\Sigma} \cdot \overset{n \times n}{V^T}$$

$\swarrow \quad \searrow$
 orthogonal matrices

diagonal.

① $A^T \cdot A$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0$$

$\neq 0 \quad \neq 0 \quad \quad \quad \neq 0 \quad \quad \quad \parallel 0 \quad \quad \quad \parallel 0$
 $v_1, v_2, \dots, v_r, v_{r+1}, \dots, v_n$

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

$\|Av_i\|^2 = \lambda_i$

$\Rightarrow Av_{r+1} = 0, Av_{r+2} = 0, \dots, Av_n = 0$

orthonormal eigenvalues

$$V = [v_1 \ v_2 \ \dots \ v_n] \quad \Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \dots & & \\ & & & \sigma_r & \\ & & & & 0 & \\ & & & & & \dots \\ & & & & & & 0 \end{bmatrix}$$

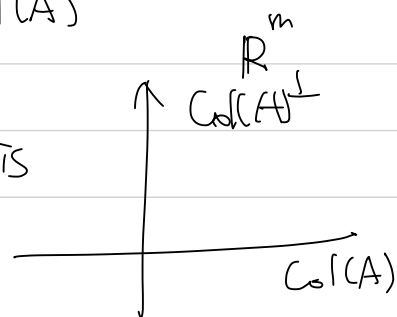
$\sigma_i = \sqrt{\lambda_i}, \sigma_2 = \sqrt{\lambda_2}, \dots, \sigma_r = \sqrt{\lambda_r}$

② $u_1 = \frac{1}{\sigma_1} Av_1, u_2 = \frac{1}{\sigma_2} Av_2, \dots, u_r = \frac{1}{\sigma_r} Av_r$

\downarrow : orthonormal basis for $\text{Col}(A)$

$\{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$: orthonormal basis

Find a basis for $\text{Col}(A)^\perp$ for \mathbb{R}^m .
 Apply G-S.



$$U = [u_1, \dots, u_m]$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \text{Char. Poly} = \lambda^2 - (a+d)\lambda + (ad-bc)$$

"
 $\det(A - \lambda I)$

Example 3: Construct the singular value decomposition of

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}_{3 \times 2} \quad A = U \cdot \Sigma \cdot V^T$$

$3 \times 3 \quad 3 \times 2 \quad 2 \times 2$

(It has rank 1.)

$$\textcircled{1} \quad A^T \cdot A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}_{2 \times 2}$$

$$\lambda^2 - (9+9)\lambda + (9^2 - (-9)^2) = \lambda^2 - 18\lambda = 0 \quad \lambda = 18, 0$$

$$\lambda_1 = 18 \quad \lambda_2 = 0 \quad \sigma_1 = \sqrt{18}, \quad \sigma_2 = 0.$$

$$\text{Null}(AA^T - 18I) = \text{Null}\left(\begin{bmatrix} -9 & -9 \\ -9 & -9 \end{bmatrix}\right) = \text{Span} \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \rightarrow v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

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

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

Section 7.4 Slide 41

$$\textcircled{2} \quad u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{18}} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{36}} \begin{bmatrix} 2 \\ -4 \\ 4 \end{bmatrix}$$

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

Need: $\{u_1, u_2, u_3\}$: ONB for \mathbb{R}^3

$$u_2, u_3 \in (\text{Span of } u_1)^\perp = \frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

$$= \left\{ \vec{x} : \vec{x} \cdot \vec{u}_1 = 0 \right\}$$

$$= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 - 2x_2 + 2x_3 = 0 \right\}$$

The set of vectors that satisfy an equation.

= Null space of ~~some matrix~~ ^{AT}.

$$= \left\{ \begin{bmatrix} 2x_2 - 2x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} : x_2, x_3 \in \mathbb{R} \right\}$$

\vec{y}_2 \vec{y}_3

$\{y_2, y_3\} \xrightarrow{G-S} \{\vec{z}_2, \vec{z}_3\}$ orthogonal

$$\vec{z}_2 = \vec{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{z}_3 = \vec{y}_3 - \left(\frac{\vec{y}_3 \cdot \vec{y}_2}{\vec{y}_2 \cdot \vec{y}_2} \right) \cdot \vec{y}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} - \frac{(-4)}{5} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$= \frac{1}{5} \begin{bmatrix} -10 + 8 \\ 4 \\ 5 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -2 \\ 4 \\ 5 \end{bmatrix}$$

$$u_2 = \frac{1}{\|\vec{z}_2\|} \vec{z}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$u_3 = \frac{1}{\|\vec{z}_3\|} \vec{z}_3 = \frac{1}{\sqrt{(-2)^2 + 4^2 + 5^2}} \begin{bmatrix} -2 \\ 4 \\ 5 \end{bmatrix} = \frac{1}{3\sqrt{5}} \begin{bmatrix} -2 \\ 4 \\ 5 \end{bmatrix}$$

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

$$\Rightarrow \text{rank}(A) = 1$$

$$\sigma_1 = \sqrt{18}$$

$$\sigma_2 = 0$$

$\text{Null}(A)^\perp$

$\text{Null}(A)$

$\{v_1, v_2\}$ ONB for \mathbb{R}^2

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

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

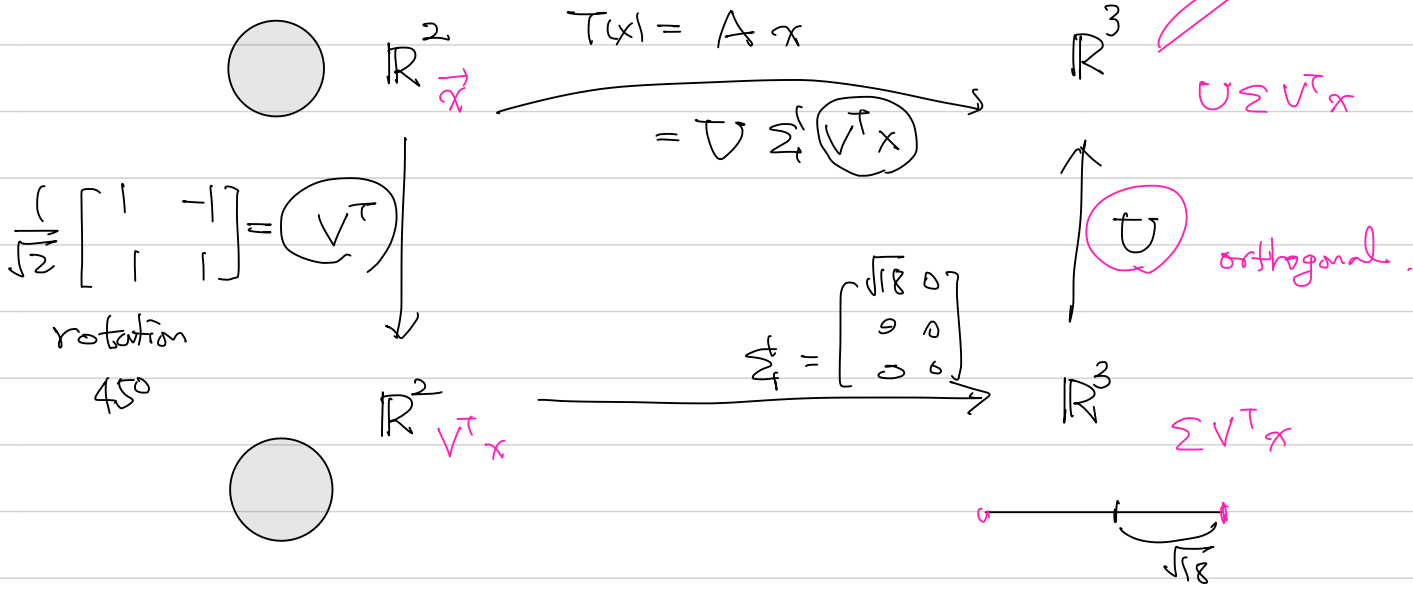
ONB for $\text{Col}(A)$

$$u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$u_3 = \frac{1}{3\sqrt{5}} \begin{bmatrix} -2 \\ 4 \\ 5 \end{bmatrix}$$

ONB for $\text{Col}(A)^\perp$
 \parallel
 $\text{Null}(A^T)$

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



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 $\text{Null}A$ 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

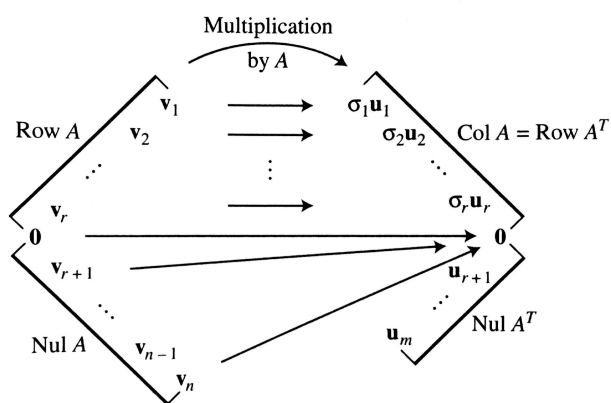


FIGURE 4 The four fundamental subspaces and the action of A .

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