Section 5.3 : Diagonalization

Chapter 5 : Eigenvalues and Eigenvectors

Math 1554 Linear Algebra

Motivation: it can be useful to take large powers of matrices, for example A^k, for large k.

But: multiplying two $n \times n$ matrices requires roughly n^3 computations. Is there a more efficient way to compute A^k?

Topics and Objectives

Topics

1. Diagonal, similar, and diagonalizable matrices
2. Diagonalizing matrices

Learning Objectives
For the topics covered in this section, students are expected to be able to do the following.

1. Determine whether a matrix can be diagonalized, and if possible diagonalize a square matrix.
2. Apply diagonalization to compute matrix powers.

Diagonal Matrices

A matrix is **diagonal** if the only non-zero elements, if any, are on the main diagonal.

The following are all diagonal matrices.

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 \end{bmatrix}, \quad I_n, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We'll only be working with diagonal square matrices in this course.

Powers of Diagonal Matrices

If A is diagonal, then A^k is easy to compute. For example,

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 0.5 \end{pmatrix}$$

$$A^2 =$$

$$A^k =$$

But what if A is not diagonal?

Diagonalization

Suppose $A \in \mathbb{R}^{n\times n}$. We say that A is **diagonalizable** if it is similar to a diagonal matrix, D. That is, we can write

$$A = PDP^{-1}$$

Diagonalization

Theorem

If A is diagonalizable $\Leftrightarrow$ A has n linearly independent eigenvectors.

Note: the symbol $\Leftrightarrow$ means " if and only if ".

Also note that $A = PDP^{-1}$ if and only if

$$A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_n]^{-1}$$

where $\vec{v}_1, \ldots, \vec{v}_n$ are linearly independent eigenvectors, and $\lambda_1, \ldots, \lambda_n$ are the corresponding eigenvalues (**in order**).

Example 1

Diagonalize if possible.

$$\begin{pmatrix} 2 & 6 \\ 0 & -1 \end{pmatrix}$$

Example 2

Diagonalize if possible.

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

Distinct Eigenvalues

Theorem

If A is $n \times n$ and has n distinct eigenvalues, then A is diagonalizable.

Why does this theorem hold?

Is it necessary for an $n \times n$ matrix to have n distinct eigenvalues for it to be diagonalizable?

Non-Distinct Eigenvalues

Theorem. Suppose

- A is $n \times n$
- A has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, $k \leq n$
- a_i = algebraic multiplicity of λ_i
- d_i = dimension of λ_i eigenspace ("geometric multiplicity")

Then

1. $d_i \leq a_i$ for all i
2. A is diagonalizable $\Leftrightarrow \Sigma d_i = n \Leftrightarrow d_i = a_i$ for all i
3. A is diagonalizable $\Leftrightarrow$ the eigenvectors, for all eigenvalues, together form a basis for $\mathbb{R}^n$.

Example 3

The eigenvalues of A are $\lambda = 3, 1$. If possible, construct P and D such that $AP = PD$.

$$A = \begin{pmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{pmatrix}$$

Additional Example (if time permits)

Note that

$$\vec{x}_k = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \vec{x}_{k-1}, \quad \vec{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad k = 1, 2, 3, \ldots$$

generates a well-known sequence of numbers.

Use a diagonalization to find a matrix equation that gives the n^{th} number in this sequence.

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.

Other sources that you may find helpful are listed below.

1. PageRank Algorithm (Math Explorer's Club, Cornell Univ.) `http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html`
2. Austin, D. *How Google Finds Your Needle in the Web's Haystack*. Available at: `http://www.ams.org/samplings/feature-column/fcarc-pagerank`
3. Bryan, K., Leise, T. *The $25,000,000,000 Eigenvector: The Linear Algebra behind Google*. SIAM Review, 48(3). Available at: `http://userpages.umbc.edu/~kogan/teaching/m430/GooglePageRank.pdf`

Steady State Vectors

Recall the car rental problem from our Section 4.9 lecture.

Problem

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

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

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

Long Term Behaviour

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

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

The distribution of cars after 2 weeks is:

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

The distribution of cars after n weeks is:

Long Term Behaviour

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

1. compute $P^n\vec{x}_0$ for large n.
2. compute the **steady-state vector**, $\vec{q}$, by solving $\vec{q} = P\vec{q}$.

To solve PageRank problems, we will rely on the first approach.

Theorem 1

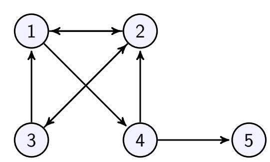
If P is a regular $m \times m$ transition matrix with $m \geq 2$, then the following statements are all true.

1. There is a stochastic matrix Π such that
$$\lim_{n\to\infty} P^n = \Pi$$
2. Each column of Π is the same probability vector $\vec{q}$.
3. For any initial probability vector $\vec{x}_0$,
$$\lim_{n\to\infty} P^n\vec{x}_0 = \vec{q}$$
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

Suppose we have 4 web pages that link to each other according to this diagram.



Page 1 has links to pages ________________ .

Page 2 has links to pages ________________ .

If a user on a page in this web is **equally likely** to go to any of the pages that their page links to, construct a Markov chain that represents how users navigate this 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.

Is the transition matrix in Example 1 a regular matrix?

Adjustment 1

Adjustment 1

If a user reaches a page that doesn't link to other pages, then the user will choose any page in the web, with equal probability, and move to that page.

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

Adjustment 2

Adjustment 2

A user at any page will navigate any page among those that their page links to with equal probability p, and to any page in the web with equal probability $1-p$. The transition matrix becomes

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

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

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

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

Computing Page Rank

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

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

- In practice we can compute the page rank for each page in the web by evaluating:

$$G^n \vec{x}_0$$

for large n. The elements of the resulting vector give the page ranks of each page in the web.

On a MATH 1554 exam,

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

Example 2 (if time permits)

Construct the Google Matrix for the web below (your instructor would provide the web).

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 Syntax for Matrix Powers

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
```

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

Chapter 5 : Eigenvalues and Eigenvectors

5.5 : Complex Eigenvalues

Topics and Objectives

Topics

1. Complex numbers: addition, multiplication, complex conjugate
2. Diagonalizing matrices with complex eigenvalues
3. Eigenvalue theorems

Learning Objectives

1. Diagonalize 2×2 matrices that have complex eigenvalues.
2. Use eigenvalues to determine identify the rotation and dilation of a linear transform.
3. Apply theorems to characterize matrices with complex eigenvalues.

Motivating Question
What are the eigenvalues of a rotation matrix?

Imaginary Numbers

Recall: When calculating roots of polynomials, we can encounter square roots of negative numbers. For example:

$$x^2 + 1 = 0$$

The roots of this equation are:

We usually write $\sqrt{-1}$ as i (for "imaginary").

Addition and Multiplication

The imaginary (or complex) numbers are denoted by $\mathbb{C}$, where

$$\mathbb{C} = \{a + bi \mid a, b \text{ in } \mathbb{R}\}$$

We can identify $\mathbb{C}$ with $\mathbb{R}^2$: $\quad a + bi \leftrightarrow (a, b)$

We can add and multiply complex numbers as follows:

$(2 - 3i) + (-1 + i) =$

$(2 - 3i)(-1 + i) =$

Complex Conjugate, Absolute Value, Polar Form

We can **conjugate** complex numbers: $\overline{a + bi} =$ ___________

The **absolute value** of a complex number: $|a + bi| =$ ___________

We can write complex numbers in **polar form**: $a + ib = r(\cos\phi + i \sin\phi)$

Complex Conjugate Properties

If x and y are complex numbers, $\vec{v} \in \mathbb{C}^n$, it can be shown that:

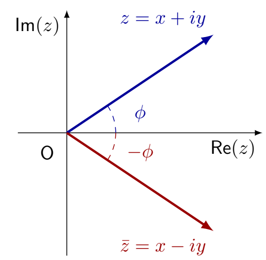
- $\overline{(x+y)} = \overline{x} + \overline{y}$
- $\overline{A\vec{v}} = A\overline{\vec{v}}$
- $\text{Im}(x\overline{x}) = 0$.

Example True or false: if x and y are complex numbers, then

$$\overline{(xy)} = \overline{x}\ \overline{y}$$

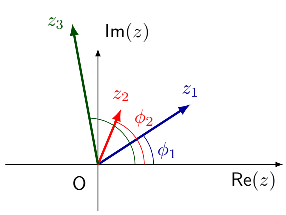
Polar Form and the Complex Conjugate

Conjugation reflects points across the real axis.



Euler's Formula

Suppose z_1 has angle ϕ_1, and z_2 has angle ϕ_2.



The product $z_1 z_2$ has angle $\phi_1 + \phi_2$ and modulus $|z|\,|w|$. Easy to remember using Euler's formula.

$$\boxed{z = |z|\, e^{i\phi}}$$

The product $z_1 z_2$ is:

$$z_3 = z_1 z_2 = (|z_1|\, e^{i\phi_1})(|z_2| e^{i\phi_2}) = |z_1|\, |z_2|\, e^{i(\phi_1+\phi_2)}$$

Complex Numbers and Polynomials

Theorem: Fundamental Theorem of Algebra

Every polynomial of degree n has exactly n complex roots, counting multiplicity.

Theorem

1. If $\lambda \in \mathbb{C}$ is a root of a real polynomial $p(x)$, then the conjugate $\overline{\lambda}$ is also a root of $p(x)$.
2. If λ is an eigenvalue of real matrix A with eigenvector $\vec{v}$, then $\overline{\lambda}$ is an eigenvalue of A with eigenvector $\overline{\vec{v}}$.

Example

Four of the eigenvalues of a 7×7 matrix are $-2, 4+i, -4-i$, and i. What are the other eigenvalues?

Example

The matrix that rotates vectors by $\phi = \pi/4$ radians about the origin, and then scales (or dilates) vectors by $r = \sqrt{2}$, is

$$A = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

What are the eigenvalues of A? Express them in polar form.

Example

The matrix in the previous example is a special case of this matrix:

$$C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Calculate the eigenvalues of C and express them in polar form.

Diagonalization

Theorem

Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ (where $b \neq 0$) and associated eigenvector $\vec{v}$. Then we may construct the diagonalization

$$A = PCP^{-1}$$

where

$$P = (\mathrm{Re}\,\vec{v} \quad \mathrm{Im}\,\vec{v}) \quad \text{and} \quad C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Note the following.

- C is referred to as a **rotation dilation** matrix, because it is the composition of a rotation by ϕ and dilation by r.
- The proof for why the columns of P are always linearly independent is a bit long, it goes beyond the scope of this course.

Example

If possible, construct matrices P and C such that $AP = PC$.

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

Topics and Objectives

Topics

1. Dot product of vectors
2. Magnitude of vectors, and distances in $\mathbb{R}^n$
3. Orthogonal vectors and complements
4. Angles between vectors

Learning Objectives

1. Compute (a) dot product of two vectors, (b) length (or magnitude) of a vector, (c) distance between two points in $\mathbb{R}^n$, and (d) angles between vectors.
2. Apply theorems related to orthogonal complements, and their relationships to Row and Null space, to characterize vectors and linear systems.

Motivating Question
For a matrix A, which vectors are orthogonal to all the rows of A? To the columns of A?

The Dot Product

$$\vec{u} \cdot \vec{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Example 1: For what values of k is $\vec{u} \cdot \vec{v} = 0$?

$$\vec{u} = \begin{pmatrix} -1 \\ 3 \\ k \\ 2 \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \end{pmatrix}$$

Properties of the Dot Product

The dot product is a special form of matrix multiplication, so it inherits linear properties.

Theorem (Basic Identities of Dot Product)

Let $\vec{u}, \vec{v}, \vec{w}$ be three vectors in $\mathbb{R}^n$, and $c \in \mathbb{R}$.

1. (Symmetry) $\vec{u} \cdot \vec{w} =$ ____________
2. (Linear in each vector) $(\vec{v} + \vec{w}) \cdot \vec{u} =$ ____________
3. (Scalars) $(c\vec{u}) \cdot \vec{w} =$ ____________
4. (Positivity) $\vec{u} \cdot \vec{u} \geq 0$, and the dot product equals ____________

The Length of a Vector

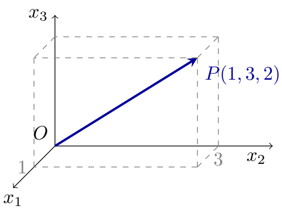
Definition

The **length** of a vector $\vec{u} \in \mathbb{R}^n$ is

$$\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

Example: the length of the vector $\overrightarrow{OP}$ is

$$\sqrt{1^2 + 3^2 + 2^2} = \sqrt{14}$$



Example

Let $\vec{u}, \vec{v}$ be two vectors in $\mathbb{R}^n$ with $\|\vec{u}\| = 5$, $\|\vec{v}\| = \sqrt{3}$, and $\vec{u} \cdot \vec{v} = -1$.
Compute the value of $\|\vec{u} + \vec{v}\|$.

Length of Vectors and Unit Vectors

Note: for any vector $\vec{v}$ and scalar c, the length of $c\vec{v}$ is

$$\|c\vec{v}\| =$$

Definition

If $\vec{v} \in \mathbb{R}^n$ has length one, we say that it is a **unit vector.**

Example: Let W be a subspace of $\mathbb{R}^4$ spanned by

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

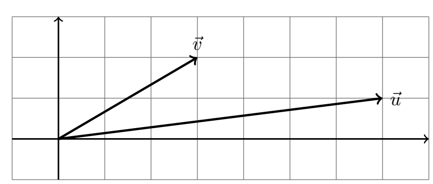
a) Construct a unit vector $\vec{u}$ in the same direction as $\vec{v}$.
b) Construct a basis for W using unit vectors.

Distance in $\mathbb{R}^n$

Definition

For $\vec{u}, \vec{v} \in \mathbb{R}^n$, the **distance** between $\vec{u}$ and $\vec{v}$ is given by the formula

Example: Compute the distance from $\vec{u} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.



Orthogonality

Definition (Orthogonal Vectors)

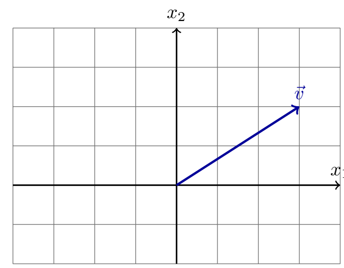
Two vectors $\vec{u}$ and $\vec{w}$ are **orthogonal** if $\vec{u} \cdot \vec{w} = 0$. This is equivalent to:

$$\|\vec{u} + \vec{w}\|^2 =$$

Note: The zero vector is orthogonal to every vector. But we usually only mean non-zero vectors.

Example

Sketch the subspace spanned by the set of all vectors $\vec{u}$ that are orthogonal to $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.



Orthogonal Compliments

Definitions

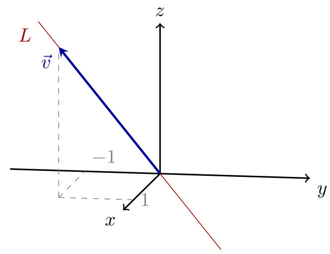
Let W be a subspace of $\mathbb{R}^n$. A vector $\vec{z} \in \mathbb{R}^n$ is said to be **orthogonal** to W if $\vec{z}$ is orthogonal to each vector in W.

The set of all vectors orthogonal to W is a subspace, the **orthogonal compliment** of W, or $W^\perp$ or 'W perp.'

$$W^\perp = \{\vec{z} \in \mathbb{R}^n \ : \ \vec{z} \cdot \qquad \}$$

Example

Line L is a subspace of $\mathbb{R}^3$ spanned by $\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$. Then the space $L^\perp$ is a plane. Construct an equation of the plane $L^\perp$.



Can also visualise line and plane with CalcPlot3D: web.monroecc.edu/calcNSF

Row$A$

Definition

Row$A$ is the space spanned by the rows of matrix A.

We can show that

- dim(Row(A)) = dim(Col(A))
- a basis for Row$A$ is the pivot rows of A

Example

Describe the Null(A) in terms of an orthogonal subspace.

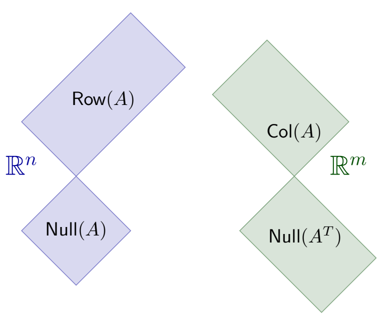
A vector $\vec{x}$ is in $\mathrm{Null}\,A$ if and only if

1. $A\vec{x} =$
2. This means that $\vec{x}$ is ______ to each row of A.
3. $\mathrm{Row}\,A$ is ______ to $\mathrm{Null}\,A$.
4. The dimension of $\mathrm{Row}\,A$ plus the dimension of $\mathrm{Null}\,A$ equals ______

Theorem (The Four Subspaces)

For any $A \in \mathbb{R}^{m\times n}$, the orthogonal complement of $\mathrm{Row}\,A$ is $\mathrm{Null}\,A$, and the orthogonal complement of $\mathrm{Col}\,A$ is $\mathrm{Null}\,A^T$.

The idea behind this theorem is described in the diagram below.



Additional Example (if time permits)

A has the LU factorization:

$$A = LU = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

a) Construct a basis for $(\mathrm{Row}A)^\perp$
b) Construct a basis for $(\mathrm{Col}A)^\perp$

Hint: it is not necessary to compute A. Recall that $A^T = U^T L^T$, matrix L^T is invertible, and U^T has a non-empty nullspace.

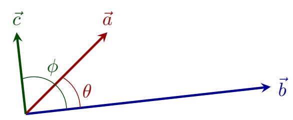
Angles

Theorem

$\vec{a} \cdot \vec{b} = |\vec{a}|\,|\vec{b}|\cos\theta$. Thus, if $\vec{a} \cdot \vec{b} = 0$, then:

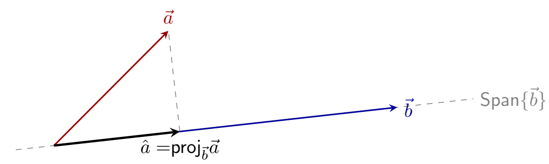
- $\vec{a}$ and/or $\vec{b}$ are _______ vectors, or
- $\vec{a}$ and $\vec{b}$ are _______.

For example, consider the vectors below.



Looking Ahead - Projections

Suppose we want to find the closed vector in Span$\{\vec{b}\}$ to $\vec{a}$.



- Later in this Chapter, we will make connections between dot products and **projections**.
- Projections are also used throughout multivariable calculus courses.

Section 6.2 : Orthogonal Sets

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Orthogonal Sets of Vectors
2. Orthogonal Bases and Projections.

Learning Objectives

1. Apply the concepts of orthogonality to
   a) compute orthogonal projections and distances,
   b) express a vector as a linear combination of orthogonal vectors,
   c) characterize bases for subspaces of $\mathbb{R}^n$, and
   d) construct orthonormal bases.

Motivating Question

What are the special properties of this basis for $\mathbb{R}^3$?

$$\begin{bmatrix}3\\1\\1\end{bmatrix}/\sqrt{11},\quad \begin{bmatrix}-1\\2\\1\end{bmatrix}/\sqrt{6},\quad \begin{bmatrix}-1\\-4\\7\end{bmatrix}/\sqrt{66}$$

Orthogonal Vector Sets

Definition

A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_p\}$ are an **orthogonal set** of vectors if for each $j \neq k$, $\vec{u}_j \perp \vec{u}_k$.

Example: Fill in the missing entries to make $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ an orthogonal set of vectors.

$$\vec{u}_1 = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -2 \\ 1 \\ \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} \\ \\ \end{bmatrix}$$

Linear Independence

Theorem (Linear Independence for Orthogonal Sets)

Let $\{\vec{u}_1, \ldots, \vec{u}_p\}$ be an orthogonal set of vectors. Then, for scalars $c_1, \ldots, c_p$,

$$\left\| c_1\vec{u}_1 + \cdots + c_p\vec{u}_p \right\|^2 = c_1^2\|\vec{u}_1\|^2 + \cdots + c_p^2\|\vec{u}_p\|^2.$$

In particular, if all the vectors $\vec{u}_r$ are non-zero, the set of vectors $\{\vec{u}_1, \ldots, \vec{u}_p\}$ are linearly independent.

Orthogonal Bases

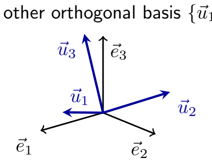
Theorem (Expansion in Orthogonal Basis)

Let $\{\vec{u}_1, \ldots, \vec{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then, for any vector $\vec{w} \in W$,

$$\vec{w} = c_1\vec{u}_1 + \cdots + c_p\vec{u}_p.$$

Above, the scalars are $c_q = \dfrac{\vec{w} \cdot \vec{u}_q}{\vec{u}_q \cdot \vec{u}_q}$.

For example, any vector $\vec{w} \in \mathbb{R}^3$ can be written as a linear combination of $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, or some other orthogonal basis $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$.



Example

$$\vec{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{s} = \begin{pmatrix} 3 \\ -4 \\ 1 \end{pmatrix}$$

Let W be the subspace of $\mathbb{R}^3$ that is orthogonal to $\vec{x}$.

a) Check that an orthogonal basis for W is given by $\vec{u}$ and $\vec{v}$.
b) Compute the expansion of $\vec{s}$ in basis W.

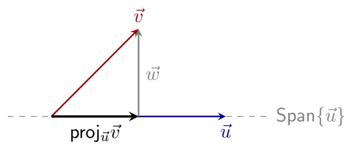
Projections

Let $\vec{u}$ be a non-zero vector, and let $\vec{v}$ be some other vector. The **orthogonal projection of $\vec{v}$ onto the direction of $\vec{u}$** is the vector in the span of $\vec{u}$ that is closest to $\vec{v}$.

$$\boxed{\text{proj}_{\vec{u}}\vec{v} = \frac{\vec{v}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{u}.}$$

The vector $\vec{w} = \vec{v} - \text{proj}_{\vec{u}}\vec{v}$ is orthogonal to $\vec{u}$, so that

$$\vec{v} = \text{proj}_{\vec{u}}\vec{v} + \vec{w}$$
$$\|\vec{v}\|^2 = \|\text{proj}_{\vec{u}}\vec{v}\|^2 + \|\vec{w}\|^2$$



Example

Let L be spanned by $(1,1,1,1)$ in $\mathbb{R}^4$.

1. Find the projection of $\vec{v} = (-3,5,6,-4)$ onto the line L.
2. How close is $\vec{v}$ to the line L?

Definition

Definition (Orthonormal Basis)

An **orthonormal basis** for a subspace W is an orthogonal basis $\{\vec{u}_1,\ldots,\vec{u}_p\}$ in which every vector $\vec{u}_q$ has unit length. In this case, for each $\vec{w} \in W$,

$$\vec{w} = [(\vec{w})\cdot\vec{u}_1]\vec{u}_1 + \cdots + [(\vec{w})\cdot\vec{u}_p]\vec{u}_p$$
$$\|\vec{w}\| = \sqrt{[(\vec{w})\cdot\vec{u}_1]^2 + \cdots + [(\vec{w})\cdot\vec{u}_p]^2}$$

Example

The subspace W is a subspace of $\mathbb{R}^3$ perpendicular to $(1,1,1)$. Calculate the missing coefficients in the orthonormal basis for W.

$$\begin{bmatrix} 1 \\ 0 \\ \end{bmatrix} / \sqrt{} \qquad \begin{bmatrix} \\ \\ \end{bmatrix} / \sqrt{}$$

Orthogonal Matrices

An **orthogonal matrix** is a square matrix whose columns are orthonormal.

Theorem

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I_n$.

Note that this theorem does not apply when $n > m$. Why?

Theorem

Theorem (Mapping Properties of Orthogonal Matrices)

Assume $m \times m$ matrix U has orthonormal columns. Then

1. (Preserves length) $\|U\vec{x}\| = \boxed{}$
2. (Preserves angles) $(U\vec{x}) \cdot (U\vec{y}) = \boxed{}$
3. (Preserves orthogonality)

Example

Compute the length of the vector below.

$$\begin{bmatrix} 1/2 & 2/\sqrt{14} \\ 1/2 & 1/\sqrt{14} \\ 1/2 & -3/\sqrt{14} \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ -3 \end{bmatrix}$$

Additional Example (if time permits)

A 4×4 orthonormal matrix is below. It's columns are orthonormal.

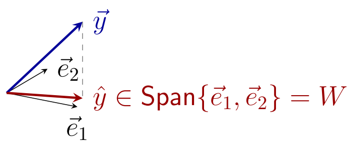
$$A = \begin{bmatrix} 1/2 & 2/\sqrt{10} & 1/2 & 1/\sqrt{10} \\ 1/2 & 1/\sqrt{10} & -1/2 & -2/\sqrt{10} \\ 1/2 & -1/\sqrt{10} & -1/2 & 2/\sqrt{10} \\ 1/2 & -2/\sqrt{10} & 1/2 & -1/\sqrt{10} \end{bmatrix}$$

Verify that the rows also form an orthonormal basis.

Section 6.3 : Orthogonal Projections

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors $\vec{e}_1$ and $\vec{e}_2$ form an orthonormal basis for subspace W.
Vector $\vec{y}$ is not in W.
The orthogonal projection of $\vec{y}$ onto W =Span$\{\vec{e}_1, \vec{e}_2\}$ is $\hat{y}$.

Topics and Objectives

Topics

1. Orthogonal projections and their basic properties
2. Best approximations

Learning Objectives

1. Apply concepts of orthogonality and projections to
   a) compute orthogonal projections and distances,
   b) express a vector as a linear combination of orthogonal vectors,
   c) construct vector approximations using projections,
   d) characterize bases for subspaces of $\mathbb{R}^n$, and
   e) construct orthonormal bases.

Motivating Question For the matrix A and vector $\vec{b}$, which vector $\widehat{b}$ in column space of A, is closest to $\vec{b}$?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -4 & -2 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Example 1

Let $\vec{u}_1, \ldots, \vec{u}_5$ be an orthonormal basis for $\mathbb{R}^5$. Let $W = \text{Span}\{\vec{u}_1, \vec{u}_2\}$.
For a vector $\vec{y} \in \mathbb{R}^5$, write $\vec{y} = \widehat{y} + w^\perp$, where $\widehat{y} \in W$ and $w^\perp \in W^\perp$.

Orthogonal Decomposition Theorem

Theorem

Let W be a subspace of $\mathbb{R}^n$. Then, each vector $\vec{y} \in \mathbb{R}^n$ has the **unique** decomposition

$$\vec{y} = \widehat{y} + w^\perp, \quad \widehat{y} \in W, \quad w^\perp \in W^\perp.$$

And, if $\vec{u}_1, \ldots, \vec{u}_p$ is any orthogonal basis for W,

$$\hat{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{y} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p}\vec{u}_p.$$

We say that $\widehat{y}$ is the **orthogonal projection of** $\vec{y}$ **onto** W**.**

If time permits, we will prove this theorem on the next slide.

Proof (if time permits)

We can write

$$\widehat{y} =$$

Then, $w^{\perp} = \vec{y} - \widehat{y}$ is in $W^{\perp}$ because

Uniqueness:

Example 2a

$$\vec{y} = \begin{pmatrix} -1 \\ 2 \\ 6 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$$

Construct the decomposition $\vec{y} = \widehat{y} + w^{\perp}$, where $\widehat{y}$ is the orthogonal projection of $\vec{y}$ onto the subspace $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$.

Best Approximation Theorem

Theorem

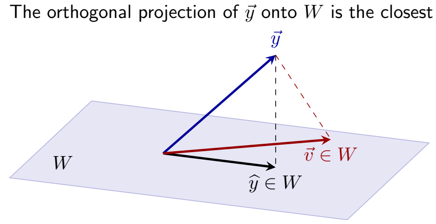
Let W be a subspace of $\mathbb{R}^n$, $\vec{y} \in \mathbb{R}^n$, and $\widehat{y}$ is the orthogonal projection of $\vec{y}$ onto W. Then for **any** $\vec{w} \neq \hat{y} \in W$, we have

$$\|\vec{y} - \widehat{y}\| < \|\vec{y} - \vec{w}\|$$

That is, $\widehat{y}$ is the unique vector in W that is closest to $\vec{y}$.

Proof (if time permits)

The orthogonal projection of $\vec{y}$ onto W is the closest point in W to $\vec{y}$.



Example 2b

$$\vec{y} = \begin{pmatrix} -1 \\ 2 \\ 6 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$$

What is the distance between $\vec{y}$ and subspace $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$? Note that these vectors are the same vectors that we used in Example 2a.

Additional Example (if time permits)

Indicate whether each statement is true or false. If true, explain why in one or two sentences. If false, give a counter example or explain why in one or two sentences.

a) If $\vec{x}$ is orthogonal to $\vec{v}$ and $\vec{w}$, then $\vec{x}$ is also orthogonal to $\vec{v} - \vec{w}$.

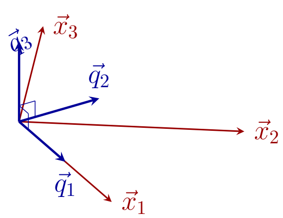
b) If $\mathrm{proj}_W \vec{y} = \vec{y}$, then $\vec{y} \in W$.

c) If $\vec{y} = \vec{u}_1 + \vec{v}_1$, where $\vec{u}_1 \in W$ and $\vec{v}_1 \in W^\perp$, then $\vec{u}_1$ is the orthogonal projection of $\vec{y}$ onto W.

Section 6.4 : The Gram-Schmidt Process

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are given linearly independent vectors. We wish to construct an orthonormal basis $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$ for the space that they span.

Topics and Objectives

Topics

1. Gram Schmidt Process
2. The QR decomposition of matrices and its properties

Learning Objectives

1. Apply the iterative Gram Schmidt Process, and the QR decomposition, to construct an orthogonal basis.
2. Compute the QR factorization of a matrix.

Motivating Question The vectors below span a subspace W of $\mathbb{R}^4$. Identify an orthogonal basis for W.

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Example

The vectors below span a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

$$\vec{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

The Gram-Schmidt Process

Given a basis $\{\vec{x}_1, \dots, \vec{x}_p\}$ for a subspace W of $\mathbb{R}^n$, iteratively define

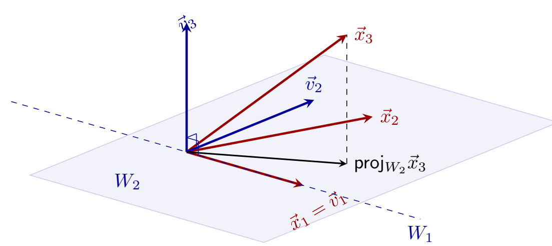
$$\begin{aligned}
\vec{v}_1 &= \vec{x}_1 \\
\vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \\
\vec{v}_3 &= \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 \\
&\vdots \\
\vec{v}_p &= \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}}\vec{v}_{p-1}
\end{aligned}$$

Then, $\{\vec{v}_1, \dots, \vec{v}_p\}$ is an orthogonal basis for W.

Proof

Geometric Interpretation

Suppose $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are linearly independent vectors in $\mathbb{R}^3$. We wish to construct an orthogonal basis for the space that they span.



We construct vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, which form our **orthogonal** basis.
$W_1 = \text{Span}\{\vec{v}_1\}$, $W_2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}$.

Orthonormal Bases

Definition

A set of vectors form an **orthonormal basis** if the vectors are mutually orthogonal and have unit length.

Example

The two vectors below form an orthogonal basis for a subspace W. Obtain an orthonormal basis for W.

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

QR Factorization

Theorem

Any $m \times n$ matrix A with linearly independent columns has the **QR factorization**

$$A = QR$$

where

1. Q is $m \times n$, its columns are an orthonormal basis for $\mathrm{Col}\, A$.
2. R is $n \times n$, upper triangular, with positive entries on its diagonal, and the length of the j^{th} column of R is equal to the length of the j^{th} column of A.

In the interest of time:

- we will not consider the case where A has linearly dependent columns
- students are not expected to know the conditions for which A has a QR factorization

Proof

Examples (if time permits)

Construct the QR decomposition for A.

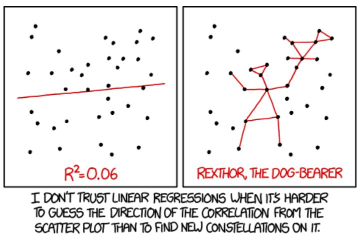
a) $A = \begin{bmatrix} 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$

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

Section 6.5 : Least-Squares Problems

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



https://xkcd.com/1725

Topics and Objectives

Topics

1. Least Squares Problems
2. Different methods to solve Least Squares Problems

Learning Objectives

1. Compute general solutions, and least squares errors, to least squares problems using the normal equations and the QR decomposition.

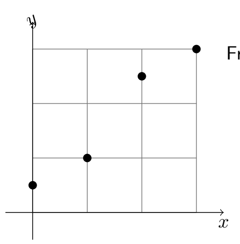
Motivating Question A series of measurements are corrupted by random errors. How can the dominant trend be extracted from the measurements with random error?

Inconsistent Systems

Suppose we want to construct a line of the form

$$y = mx + b$$

that best fits the data below.



From the data, we can construct the system:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

Can we 'solve' this inconsistent system?

The Least Squares Solution to a Linear System

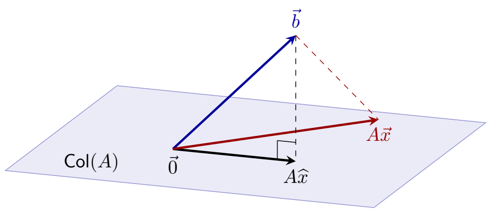
Definition: Least Squares Solution

Let A be a $m \times n$ matrix. A **least squares solution to** $A\vec{x} = \vec{b}$ is the solution $\widehat{x}$ for which

$$\| \vec{b} - A\widehat{x} \| \leq \| \vec{b} - A\vec{x} \|$$

for all $\vec{x} \in \mathbb{R}^n$.

A Geometric Interpretation

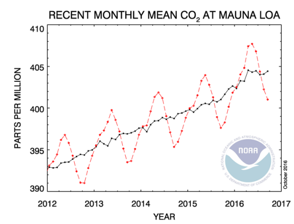


The vector $\vec{b}$ is closer to $A\hat{x}$ than to $A\vec{x}$ for all other $\vec{x} \in \text{Col}A$.

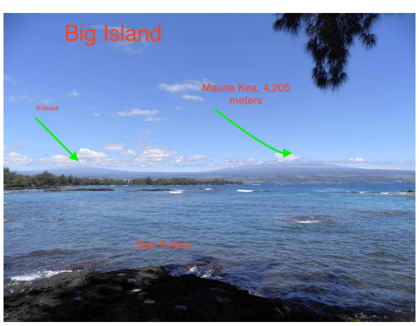
1. If $\vec{b} \in \text{Col}\,A$, then $\widehat{x}$ is ...
2. Seek $\widehat{x}$ so that $A\widehat{x}$ is as close to $\vec{b}$ as possible. That is, $\widehat{x}$ should solve $A\widehat{x} = \widehat{b}$ where $\widehat{b}$ is ...

Important Examples: Overdetermined Systems (Tall/Thin Matrices)

A variety of factors impact the measured quantity.



In the above figure, the dashed red line with diamond symbols represents the monthly mean values, centered on the middle of each month. The black line with the square symbols represents the same, after correction for the average seasonal cycle. (NOAA graph.)



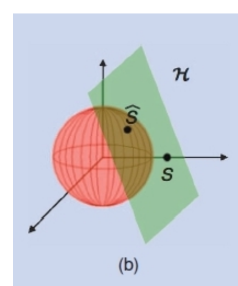
Previous data is the important time series of mean CO_2 in the atmosphere. The data is collected at the Mauna Loa observatory on the island of Hawaii (The Big Island). One of the most important observatories in the world, it is located at the top of the Mauna Kea volcano, 4,205 meters altitude.

Important Examples: Underdetermined Systems (Short/Fat Matrices)

There are too few measurements, and many solutions to $A\vec{x} = \vec{b}$. Choose $\widehat{x}$ solving the system, with the smallest length.

1. $A\widehat{x} = \vec{b}$.
2. For all $\vec{x}$ with $A\vec{x} = \vec{b}$, $\|\widehat{x}\| \leq \|\vec{x}\|$.

This is the least squares problem of 'Big Data.' (But not addressed in this course.)



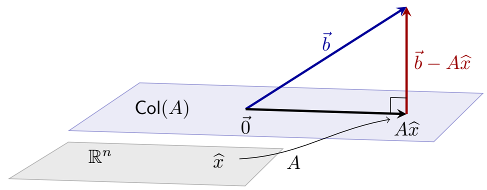
The Normal Equations

Theorem (Normal Equations for Least Squares)

The least squares solutions to $A\vec{x} = \vec{b}$ coincide with the solutions to

$$\underbrace{A^T A\vec{x} = A^T\vec{b}}_{\text{Normal Equations}}$$

Derivation



The least-squares solution $\hat{x}$ is in $\mathbb{R}^n$.

1. $\widehat{x}$ is the least squares solution, is equivalent to $\vec{b} - A\widehat{x}$ is orthogonal to ______ A.
2. A vector $\vec{v}$ is in $\mathrm{Null}\, A^T$ if and only if ______ $\vec{v} = \vec{0}$.
3. So we obtain the Normal Equations:

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Solution:

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

$$A^T\vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} =$$

The normal equations $A^T A\vec{x} = A^T\vec{b}$ become:

Theorem

Theorem (Unique Solutions for Least Squares)

Let A be any $m \times n$ matrix. These statements are equivalent.

1. The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b} \in \mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

And, if these statements hold, the least square solution is

$$\widehat{x} = (A^T A)^{-1} A^T \vec{b}.$$

Useful heuristic: $A^T A$ plays the role of 'length-squared' of the matrix A. (See the sections on symmetric matrices and singular value decomposition.)

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Hint: the columns of A are orthogonal.

Theorem (Least Squares and QR)

Let $m \times n$ matrix A have a QR decomposition. Then for each $\vec{b} \in \mathbb{R}^m$ the equation $A\vec{x} = \vec{b}$ has the unique least squares solution

$$R\widehat{x} = Q^T \vec{b}.$$

(Remember, R is upper triangular, so the equation above is solved by back-substitution.)

Example 3. Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

Solution. The QR decomposition of A is

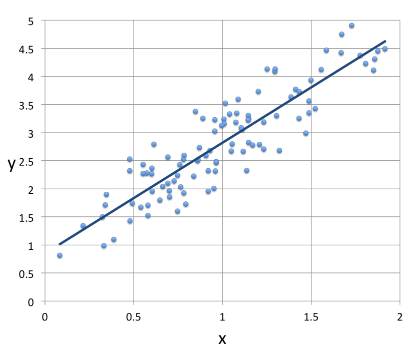
$$A = QR = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

$$Q^T\vec{b} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} \\ -6 \\ 4 \end{bmatrix}$$

And then we solve by backwards substitution $R\vec{x} = Q^T\vec{b}$

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \\ -6 \\ 4 \end{bmatrix}$$

Chapter 6 : Orthogonality and Least Squares
6.6 : Applications to Linear Models



Topics and Objectives

Topics

1. Least Squares Lines
2. Linear and more complicated models

Learning Objectives
For the topics covered in this section, students are expected to be able to do the following.

1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. Apply least-squares to fit polynomials and other curves to data.

Motivating Question
Compute the equation of the line $y = \beta_0 + \beta_1 x$ that best fits the data

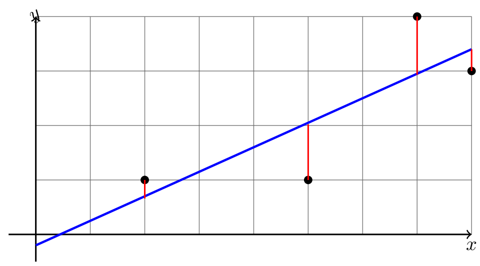
x	2	5	7	8
y	1	1	4	3

The Least Squares Line

Graph below gives an approximate linear relationship between x and y.

1. Black circles are data.
2. Blue line is the **least squares** line.
3. Lengths of red lines are the ___________.

The least squares line minimizes the sum of squares of the ___________.



Example 1 Compute the least squares line $y = \beta_0 + \beta_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

We want to solve

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

This is a least-squares problem : $X\vec{\beta} = \vec{y}$.

The normal equations are

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \end{bmatrix} \begin{bmatrix} 1 & \\ 1 & \\ 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x = \frac{-5}{21} + \frac{19}{42} x$$

As we may have guessed, β_0 is negative, and β_1 is positive.

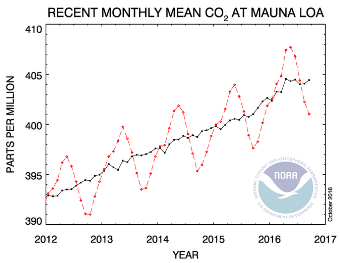
Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

$$y = \beta_0 + \beta_1 f_1(x) + \beta_1 f_2(x) + \cdots + \beta_k f_k(x).$$

where the functions f_j are known. Should have only a few functions!
Keep in mind this is a **linear problem in the** β **variables.**

Least Squares Fitting for Other Curves



Black line is yearly CO_2 levels, and the monthly is the red line. To capture seasonality, would need a curve

$$\text{daily } CO_2 = \beta_0 + \beta_1 t + \beta_2 \sin\left(2\pi \tfrac{t}{12}\right) + \beta_3 \cos\left(2\pi \tfrac{t}{12}\right)$$

Above, t is time, measured in months.

WolframAlpha and Mathematica Syntax

Least squares problems can be computed with WolframAlpha, Mathematica, and many other software.

WolframAlpha

`linear fit` $\{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\}$

Mathematica

`LeastSquares`$[\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \dots, \{x_n, x_n, y_n\}\}]$

Almost any spreadsheet program does this as a function as well.

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
3. Spectral decomposition

Learning Objectives

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

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.

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

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

A^TA is Symmetric

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

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

$$= \underbrace{\begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix}}_{\text{Entries are the dot products of columns of } A}$$

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:

Example 1

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

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

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 ___________.
2. The dimenison of each eigenspace is full, that it's dimension is **equal to** it's algebraic multiplicity.
3. The eigenspaces are mutually orthogonal.
4. A can be diagonalized: $A = PDP^T$, where D is diagonal and P is ___________.

Proof (if time permits):

Spectral Decomposition of a Matrix

Spectral Decomposition

Suppose A can be orthogonally diagonalized as

$$A = PDP^T = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix}$$

Then A has the decomposition

$$A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \cdots + \lambda_n \vec{u}_n \vec{u}_n^T = \sum_{i=1}^{n} \lambda_i \vec{u}_i \vec{u}_i^T$$

Each term in the sum, $\lambda_i \vec{u}_i \vec{u}_i^T$, is an $n \times n$ matrix with rank ________.

Example 2

Construct a spectral decomposition for A whose orthogonal diagonalization is given.

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = PDP^T$$

$$= \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

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

Definition

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

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

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

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

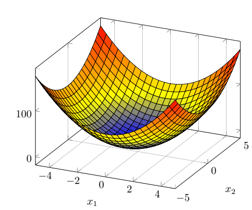
Example 1

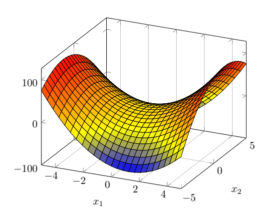
Compute the quadratic form $\vec{x}^T A\vec{x}$ for the matrices below.

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

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

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

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:

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.

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

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

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

Principle Axes

Suppose $Q(\vec{x}) = \vec{x}^T A\vec{x}$, where $A \in \mathbb{R}^2$ is symmetric and invertible. Then the set of $\vec{x}$ that satisfies

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

gives a **curve** in $\mathbb{R}^2$.

Principle Axes Theorem

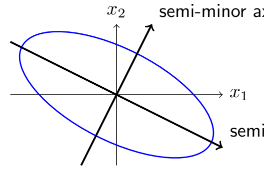
Theorem

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

Proof (if time permits):

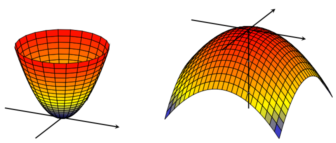
Example 5

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



Classifying Quadratic Forms

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



Definition

A quadratic form Q is

1. **positive definite** if ___________ for all $\vec{x} \neq \vec{0}$.
2. **negative definite** if ___________ for all $\vec{x} \neq \vec{0}$.
3. **positive semidefinite** if ___________ for all $\vec{x}$.
4. **negative semidefinite** if ___________ for all $\vec{x}$.
5. **indefinite** if ______________________

Quadratic Forms and Eigenvalues

Theorem

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

1. **positive definite** iff λ_i ___________
2. **negative definite** iff λ_i ___________
3. **indefinite** iff λ_i ___________

Proof (if time permits):

Example 6

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

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

Section 7.3 : Constrained Optimization

Chapter 7: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

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

Learning Objectives

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

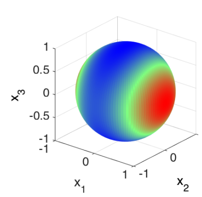
Example 1

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

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

Q is a quantity we want to optimize

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



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

A Constrained Optimization Problem

Suppose we wish to find the maximum or minimum values of

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

subject to

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

That is, we want to find

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

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

Constrained Optimization and Eigenvalues

Theorem

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

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

and associated normalized eigenvectors

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

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

Proof:

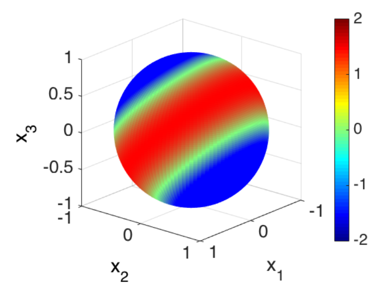
Example 2

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

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

Example 2

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



An Orthogonality Constraint

Theorem

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

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

and associated eigenvectors

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

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

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

Note that λ_2 is the second largest eigenvalue of A.

Example 3

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

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

Example 4 (if time permits)

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

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

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

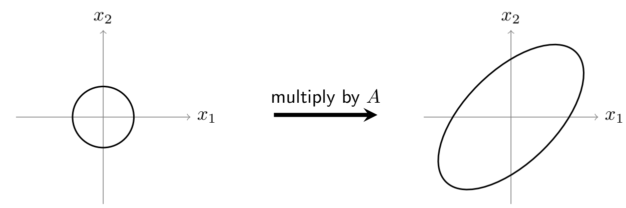
1. Compute the SVD for a rectangular matrix.
2. Apply the SVD to
   - estimate the rank and condition number of a matrix,
   - construct a basis for the four fundamental spaces of a matrix, and
   - construct a spectral decomposition of a matrix.

Example 1

The linear transform whose standard matrix is

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

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



Example 1 - Solution

Singular Values

The matrix A^TA 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 Col$A$: For $1 \leq j < k \leq r$:

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

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

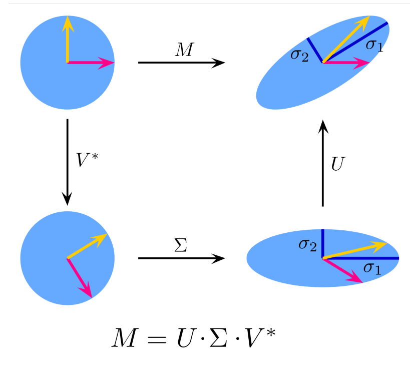
The SVD

Theorem: Singular Value Decomposition

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

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

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



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^TA, σ_i^2, and construct Σ.

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

3. Compute an orthonormal basis for 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 orthonomal basis for $\mathbb{R}^m$, use the basis for form U.

Example 2: Write down the singular value decomposition for

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

Example 3: Construct the singular value decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$.
(It has rank 1.)

Applications of the SVD

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

- Estimating the rank and condition number of a matrix
- Constructing bases for the four fundamental spaces
- Computing the pseudoinverse of a matrix
- Linear least squares problems
- Non-linear least-squares
  https://en.wikipedia.org/wiki/Non-linear_least_squares
- Machine learning and data mining
  https://en.wikipedia.org/wiki/K-SVD
- Facial recognition
  https://en.wikipedia.org/wiki/Eigenface
- Principle component analysis
  https://en.wikipedia.org/wiki/Principal_component_analysis
- Image compression

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

The Condition Number of a Matrix

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

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

is the **condition number** of A.

Note that:

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

Example 4

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

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

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

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

Example 4 - Solution

The Four Fundamental Spaces

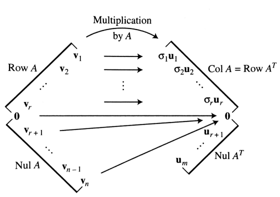


FIGURE 4 The four fundamental subspaces and the action of A.

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