Contents

Reference: Lay, Linear Algebra.

Math 3108 - Fall 2019
Chapter 1: Linear Equations in Linear Algebra

A **system of linear equations** is a collection of one more linear equations in the same variables. For example,

$$\begin{aligned} 2x_1 - x_2 + \tfrac{3}{2}x_3 &= 8 \\ x_1 - 4x_3 &= -7. \end{aligned}$$

is a system of two equations in the three unknowns x_1, x_2, x_3.

A **solution** to this system is given by $(5, \frac{13}{2}, 3)$.

The set of all possible solutions is the **solution set.** Two systems are **equivalent** if they have the same solution set.

Consistency

The special case of 2×2 systems corresponds to finding the points of intersection of two lines. In this case we find that linear system has

- no solution,
- exactly one solution, or
- infinitely many solutions.

In fact, this is true of **all linear systems**.

Definition

A linear system is **consistent** if it has a solution; it is **inconsistent** if it has no solutions.

Matrix Notation

We may rewrite linear systems in matrix form:

Example

The system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0, \\ 2x_2 - 8x_3 &= 8, \\ 5x_1 - 5x_3 &= 10 \end{aligned}$$

corresponds to the 3×4 **augmented matrix**

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}.$$

Removing the final column gives the 3×3 **coefficient matrix**.

Row Operations

We solve a linear system by performing **row operations** to replace it with equivalent systems that are progressively easier to solve. The three types of row operations are the following:

Definition (Elementary row operations)

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply the entries of a row by a nonzero constant.

If we can obtain a matrix B from a matrix A by a sequence of row operations, we say that A and B are **row equivalent.**

Row equivalent matrices have the **same** solution set.

Section 1.2 - Row Reduction and Echelon Forms

In this section we discuss the **row reduction algorithm** for solving linear systems.

The key observation is that triangular linear systems are straightforward to solve. So, given a linear system, we should perform row operations to obtain a triangular matrix. This will be called **echelon form**.

In fact, once you have a matrix in echelon form, you can perform further operations to make the system even simpler to solve. This will be called **reduced echelon form**.

Echelon and Reduced Echelon Forms

Definition (Echelon and Reduced Echelon Form)

A matrix is in **echelon form** if:

1. Nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

A matrix in echelon form is in **reduced echelon form** if additionally

4. The leading entry in each nonzero row is 1.
5. Each leading entry is the only nonzero entry in its column.

The Matlab command to compute the reduced echelon form of a matrix A is rref(A).

Examples: Echelon Form

- Not in echelon form:

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}.$$

- Echelon form, but not reduced echelon form:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{bmatrix}.$$

- Reduced echelon form:

$$\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}.$$

Theorem (Theorem 1)

Any nonzero matrix is row equivalent to a unique reduced echeleon form matrix.

On the other hand, matrices can be reduced to many different matrices in echelon form.

Definition

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position.

Roughly speaking, the first several weeks of this class could be described as ‘pivot counting’.

Row Reduction Algorithm

The following algorithm describes how to put a matrix in reduced echelon form:

1. Start with the leftmost nonzero column. The pivot position is at the top.
2. Choose a nonzero entry in the pivot column to be the pivot (using interchange to move this entry into the pivot position).
3. Use row replacement to create zeros in all positions below the pivot.
4. Repeat steps 1–3 on the sub-matrix that remains when you ignore the row containing the pivot position (and any rows above it). Repeat this until there are no more nonzero rows to modify.
5. Start with the rightmost pivot and work upward and to the left, making zeros above each pivot. Make each pivot have the value 1.

Example

Example

Reduce the matrix

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

to echelon form

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

and then to reduced echelon form

$$\begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}.$$

Variables corresponding to pivot columns are called **basic variables**, while the remaining variables are called **free variables**.

Example

Suppose the matrix of a linear system has reduced echelon form

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The associated system equations is

$$x_1 - 5x_3 = 1, \quad x_2 + x_3 = 4.$$

Then x_1, x_2 are basic and x_3 is free. The solution set can be written

$$x_1 = 1 + 5x_3, \quad x_2 = 4 - x_3, \quad x_3 \text{ is free.}$$

Another Example

Example

Find the solution set for a linear system whose augmented matrix has been reduced to

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}.$$

This is in echelon form. Let's put it in reduced echelon form:

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

Another Example (continued)

Example

The associated system is

$$x_1 + 6x_2 + 3x_4 = 0, \quad x_3 - 4x_4 = 5, \quad x_5 = 7.$$

The free variables are x_2 and x_4. The solution set is:

$$x_1 = -6x_2 - 3x_4, \quad x_3 = 5 + 4x_4, \quad x_5 = 7,$$

with x_2 and x_4 free variables.

Existence and uniqueness

Theorem (Theorem 2)

(i) *A linear system is consistent if and only if the rightmost column of the augmented matrix is* **not** *a pivot column.*

(ii) *If a linear system is consistent and has no free variables, then it has a unique solution.*

(iii) *If a linear system is consistent and has at least one free variable, then it has infinitely many slutions.*

Row Reduction Method

To summarize, here is how to use row reduction to solve a linear system:

1. Write down the augmented matrix A for the system.
2. Use row reduction to reduce the matrix to echelon form. If the system is inconsistent, stop.
3. If the system is consistent, put the matrix in reduced echelon form U.
4. Write down the linear system corresponding to the reduced matrix U.
5. Express each basic variable in terms of free variables to describe the solution set.

A final example

Example

Find the general solution of the linear system whose augmented matrix is

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

A **vector in $\mathbb{R}^n$** is an ordered list of n real numbers, usually written as an $n \times 1$ column matrix. For example,

$$\begin{bmatrix} 3 \\ 5 \\ -2 \\ 1 \end{bmatrix}$$

is a vector in $\mathbb{R}^4$. A general vector in $\mathbb{R}^n$ will be written

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix},$$

where $u_1, u_2, \cdots, u_n$ are the entries or components of the vector $\mathbf{u}$.

Algebraic Properties of Vectors

Algebraic Properties of $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$,
where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

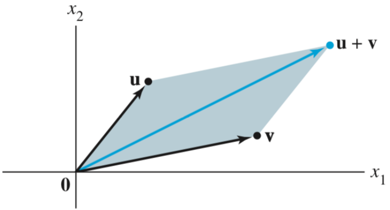
- Addition/scalar multiplication are performed component-wise.
- $\mathbf{0}$ denotes the **zero vector** (all entries equal to zero), while a **scalar** refers to a real number.

Points in the plane

We identify a point (a, b) in the plane with the vector

$$\begin{bmatrix} a \\ b \end{bmatrix}$$

in $\mathbb{R}^2$. We can then add vectors according to the 'parallelogram rule':



Linear Combinations

Definition

The **linear combination** of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ in $\mathbb{R}^n$ with **weights** $c_1, \ldots, c_p$ is the vector

$$\boldsymbol{y} = c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p.$$

Example

The linear combination of

$$\boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

with weights $c_1 = 4$ and $c_2 = 2$ is

$$\begin{bmatrix} 10 \\ 10 \end{bmatrix}.$$

Example

Example

Determine whether $\boldsymbol{b}$ can be written as a linear combination of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, where

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}.$$

To solve this, we try to solve the system $x_1\boldsymbol{a_1} + x_2\boldsymbol{a}_2 = \boldsymbol{b}$. This leads to the augmented matrix

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & -5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Solution: Use weights $x_1 = 3$ and $x_2 = 2$.

Fundamental fact

A vector equation

$$x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$$

has the same solution set as the linear system with augmented matrix

$$\begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n & \boldsymbol{b} \end{bmatrix}.$$

In particular: $\boldsymbol{b}$ can be written as a linear combination of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$ if and only if the linear system above is consistent.

Span

Definition

The **span** of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ is the set of all linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$. This set (which is a subset of $\mathbb{R}^n$) is denoted

$$\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}.$$

The following statements are equivalent:

- The vector $\boldsymbol{b}$ belongs to $\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$.
- The vector equation $x_1\boldsymbol{v}_1 + \cdots + x_p\boldsymbol{v}_p = \boldsymbol{b}$ has a solution.
- The vector $\boldsymbol{b}$ can be written as a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$.
- The linear system with augmented matrix $[\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_p \ \boldsymbol{b}]$ has a solution.

Span (Geometric Description)

If $\boldsymbol{v}$ is a nonzero vector in $\mathbb{R}^3$, then Span$\{\boldsymbol{v}\}$ is the set of points on the line in $\mathbb{R}^3$ passing through $\boldsymbol{v}$ and $\mathbf{0}$.

If $\{\boldsymbol{v}, \boldsymbol{u}\}$ are nonzero vectors in $\mathbb{R}^3$, then Span$\{\boldsymbol{v}, \boldsymbol{u}\}$ is the plane in $\mathbb{R}^3$ contaning $\mathbf{0}$, $\boldsymbol{v}$, and $\boldsymbol{u}$.

Definition

Let A be an $m \times n$ matrix with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Let $\boldsymbol{x} \in \mathbb{R}^n$.
The **product** of A and $\boldsymbol{x}$, denoted $A\boldsymbol{x}$, is the linear combination of the columns of A using the entries of $\boldsymbol{x}$ as the weights:

$$A\boldsymbol{x} = [\boldsymbol{a}_1 \;\; \boldsymbol{a}_2 \;\; \cdots \;\; \boldsymbol{a}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n.$$

Example

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

Linear Systems as Matrix Equations

Linear systems can be rewritten in the form $A\boldsymbol{x} = \boldsymbol{b}$.

Example

The system

$$\begin{aligned} x_1 + 2x_2 - 3x_3 &= 4, \\ -5x_2 + 3x_3 &= 1 \end{aligned}$$

can be written $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & 2 & -3 \\ 0 & -5 & 3 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 1 \end{bmatrix},$$

and $\boldsymbol{x} \in \mathbb{R}^3$.

Theorem 3

Theorem (Theorem 3)

Let A be an $m \times n$ matrix with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$. Let $\mathbf{b} \in \mathbb{R}^m$. The matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

which has the same solution set as the system of linear equations with augmented matrix

$$[\mathbf{a}_1 \ \cdots \ \mathbf{a}_n \ \mathbf{b}].$$

In particular, we see that $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Existence of Solutions

Example

Let

$$A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}.$$

Determine whether $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every choice of $\boldsymbol{b}$.

Solution:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_1 - \frac{1}{2}b_2 + b_3 \end{bmatrix}.$$

The answer is no. The system is consistent if and only if $b_1 - \frac{1}{2}b_2 + b_3 = 0$.

Theorem 4

Theorem (Theorem 4)

Let A be an $m \times n$ matrix. The following are equivalent:

a. *For every $\boldsymbol{b} \in \mathbb{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution.*
b. *Every $\boldsymbol{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.*
c. *The columns of A span $\mathbb{R}^m$.*
d. *A has a pivot position in every row.*

Alternate view of the product Ax

We can view the j^{th} entry of $A\boldsymbol{x}$ as the dot product between the j^{th} row of A and the vector $\boldsymbol{x}$.

Example

$$\begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ -x_1 + 5x_2 - 3x_3 \\ 6x_1 - 2x_2 + 8x_3 \end{bmatrix}.$$

Algebraic Properties

Theorem (Theorem 5)

Let A be an $m \times n$ matrix, $\boldsymbol{u}$ and $\boldsymbol{v}$ vectors in $\mathbb{R}^n$, and c a scalar. Then

$$A(\boldsymbol{u} + \boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v}, \quad A(c\boldsymbol{u}) = c(A\boldsymbol{u}).$$

A linear system is **homogeneous** if it is of the form $A\boldsymbol{x} = \mathbf{0}$, where A is $m \times n$, $\boldsymbol{x} \in \mathbb{R}^n$, and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$.

Homogeneous systems always the solution $\boldsymbol{x} = \mathbf{0}$ (the zero vector in $\mathbb{R}^m$). This is called the **trivial solution**, whereas a nonzero solution would be called a **nontrivial solution**.

The homogeneous equation $A\boldsymbol{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

An example

Example

Describe the solution set for the following homogeneous system:

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0. \end{aligned}$$

Does the system have a nontrivial solution?

Solution: We form the augmented matrix. We can omit the final column.

$$\begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Example (Example, continued)

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0. \end{aligned} \rightarrow \begin{bmatrix} 3 & 5 & -4 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution set is

$$x_2 = 0, \quad x_3 \text{ free}, \quad x_1 = \tfrac{4}{3}x_3, \quad \text{i.e.} \quad \boldsymbol{x} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}.$$

It has a non-trivial solution, e.g. $(4, 0, 3)$.

Parametric Vector Form

The solution set of a homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$ can always be expressed in the form

$$\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$$

for some collection of vectors. Equivalently, we may write the general solution as

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p \tag{1}$$

for arbitrary $c_1, \ldots, c_p \in \mathbb{R}$.

We call (1) the **parametric vector form** of the solution.

Solutions of Nonhomogeneous Systems

Theorem (Theorem 6)

The general solution to $A\boldsymbol{x} = \boldsymbol{b}$ is of the form

$$\boldsymbol{x} = \boldsymbol{x}_h + \boldsymbol{x}_p,$$

where $\boldsymbol{x}_h$ is the general solution to the homogeneous equation $A\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{x}_p$ is any particular solution to $A\boldsymbol{x} = \boldsymbol{b}$.

Example

Describe all solutions to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & 8 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

Example (continued)

Example

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & 8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{array}{r} x_1 - \frac{4}{3}x_3 = -1 \\ x_2 = 2. \end{array}.$$

So the general solution is

$$\boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}, \quad x_3 \in \mathbb{R}.$$

Remark. The solution set is a line through the origin in $\mathbb{R}^3$ translated by a fixed vector.

Practice Problem

Example

Write the general solution of

$$10x_1 - 3x_2 - 2x_3 = 7$$

in parametric vector form.

Linear systems have many applications. For example, the book discusses examples related to:

- A homogeneous system in economics.
- Balancing chemical equations.
- Network flow.

Chemical equation example

Example

Propane (C_3H_8) combines with oxygen (O_2) to form carbon dioxide (CO_2) and water (H_2O). We want to balance the equation

$$x_1 \cdot C_3H_8 + x_2 \cdot O_2 \to x_3 \cdot CO_2 + x_4 \cdot H_2O.$$

We write three equations, one for C, H, and O respectively:

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Example (continued)

Example (Continued)

Equivalently, we need to solve the homogeneous system with matrix

$$\begin{bmatrix} 3 & 0 & -1 & 0 \\ 8 & 0 & 0 & -2 \\ 0 & 2 & -2 & -1 \end{bmatrix}.$$

The general solution is

$$x_1 = \tfrac{1}{4}x_1, \quad x_2 = \tfrac{5}{4}x_4, \quad x_3 = \tfrac{3}{4}x_4, \quad x_4 \text{ free.}$$

The balanced equation is

$$C_3H_8 + 5O_2 \to 3CO_2 + 4H_2O.$$

Definition

A set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is **(linearly) independent** if the vector equation

$$x_1\boldsymbol{v}_1 + \cdots + x_p\boldsymbol{v}_p = \mathbf{0}$$

has *only* the trivial solution $\boldsymbol{x} = \mathbf{0}$.

The set is **(linearly) dependent** if there exist weights $c_1, \ldots, c_p$ not all zero such that

$$c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}.$$

Example

Example

Determine whether $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is independent, where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

If not, find a dependence relation between $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

Solution: We write

$$\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

This shows that the set is *dependent*.

Example (continued)

Example (continued)

To find a dependence relation, continue reducing:

$$\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

This has the solution set

$$x_1 = 2x_3, \quad x_2 = -x_3, \quad x_3 \text{ free.}$$

So we can write (choosing $x_3 = 1$, say) the dependence relation

$$2\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3 = \mathbf{0}.$$

Matrix Columns

Applying the above definition to the columns of a matrix A, we find:

The columns of a matrix A are linearly independent if and only if the equation $A\boldsymbol{x} = \mathbf{0}$ has *only* the trivial solution.

Example

Are the columns of the matrix

$$A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 5 & 8 & 0 \end{bmatrix}$$

independent?

Solution: Yes:

$$A \sim \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 4 \\ 0 & 0 & 13 \end{bmatrix}.$$

Some simple cases.

- A set $\{\boldsymbol{v}\}$ is independent if and only if $\boldsymbol{v}$ is not the zero vector.
- A set $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is independent if and only if neither vector is a multiple of the other.

This generalizes to:

Theorem (Theorem 9)

If a set contains the zero vector, then it is linearly dependent.

Theorem (Theorem 7)

A set S is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others.

A final theorem

Theorem (Theorem 8)

Any set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ *in* $\mathbb{R}^n$ *is linearly dependent if* $p > n$.

Proof.

Let A be the matrix with $\mathbf{v}_1, \ldots, \mathbf{v}_p$ as its columns. Then the system $A\mathbf{x} = \mathbf{0}$ has more variables than equations, and hence has a nontrivial solution. $\square$

Definition (Transformation)

A **transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\boldsymbol{x} \in \mathbb{R}^n$ a vector $T(\boldsymbol{x}) \in \mathbb{R}^m$. We write

$$T : \mathbb{R}^n \to \mathbb{R}^m.$$

We call $\mathbb{R}^n$ the **domain** of T and $\mathbb{R}^m$ the **codomain**.

We call $T(\boldsymbol{x})$ the **image** of $\boldsymbol{x}$. The set of all images is the **range** of T.

Given an $m \times n$ matrix A, we may define the transformation

$$T : \mathbb{R}^n \to \mathbb{R}^m, \quad T(\boldsymbol{x}) = A\boldsymbol{x}.$$

The range of T is the span of the columns of A, i.e. the set of all linear combinations of the columns of A.

Example

Consider the matrix transformation given by

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}.$$

Set

$$\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}.$$

Example (continued)

Example (Continued)

a. Find $T(\boldsymbol{u})$. *Answer:* $\begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$.

b. Find $\boldsymbol{x} \in \mathbb{R}^2$ such that $T(\boldsymbol{x}) = \boldsymbol{b}$. *Answer:* $\begin{bmatrix} 3/2 \\ -1/2 \end{bmatrix}$.

c. Is $\boldsymbol{c}$ in the range of T? *Answer:* No.

More examples

Example

The matrix transformation with

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

is a projection of $\mathbb{R}^3$ onto the xy plane.

Example

A matrix of the form

$$A = \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix}$$

gives rise to a **shear transformation** of the plane $\mathbb{R}^2$.

Definition

A transformation T is **linear** if

(i) $T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v}$ in the domain of T, and

(ii) $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all scalars c and all $\boldsymbol{u}$ in the domain of T.

- Every matrix transformation is linear.
- Linear transformations satisfy $T(\mathbf{0}) = \mathbf{0}$.
- (i) and (ii) can be combined to $T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v})$.
- More generally,

$$T(c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p) = c_1T(\boldsymbol{v}_1) + \cdots + c_pT(\boldsymbol{v}_p).$$

Example

Example

Describe the geometric effect of the linear transformation corresponding to the matrix

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

Solution. The transformation is a counterclockwise rotation by 90 degrees.

In the case that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ arises geometrically, we would like to write down an explicit formula for the matrix giving rise to T. Here's how to do it:

Theorem (Theorem 10)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then

$$T(\boldsymbol{x}) = A\boldsymbol{x},$$

where A is the $m \times n$ matrix whose j^{th} column is the vector $T(\boldsymbol{e}_j)$:

$$A = [\, T(\boldsymbol{e}_1) \cdots T(\boldsymbol{e}_n) \,].$$

Here $\boldsymbol{e}_j$ is the j^{th} column of the identity matrix in $\mathbb{R}^n$.

Example

Example

Suppose that T is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -3 \\ -8 \\ 0 \end{bmatrix}.$$

Then $T(\boldsymbol{x}) = A\boldsymbol{x}$, where

$$A = \begin{bmatrix} 5 & -3 \\ -7 & -8 \\ 2 & 0 \end{bmatrix}.$$

We call A the **standard matrix** of T.

Rotations of the Plane

Example

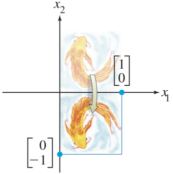
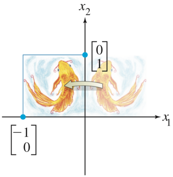
Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation of the plane through the origin by angle ϕ. Since

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix},$$

the standard matrix of T is

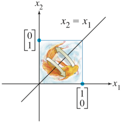
$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}.$$

TABLE 1 Reflections

Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

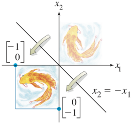
Transformations of the Plane

Reflection through the line $x_2 = x_1$



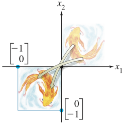
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Reflection through the line $x_2 = -x_1$



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

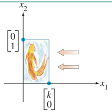
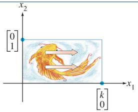
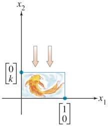
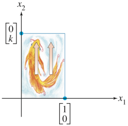
Reflection through the origin



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

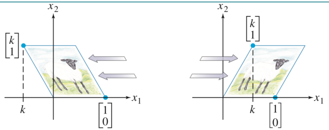
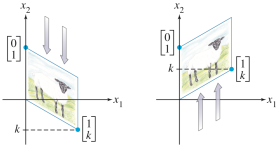
Transformations of the Plane

TABLE 2 Contractions and Expansions

Transformation	Image of the Unit Square		Standard Matrix
Horizontal contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

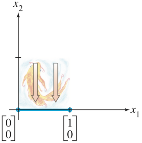
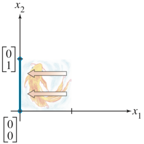
Transformations of the Plane

TABLE 3 Shears

Transformation	Image of the Unit Square	Standard Matrix
Horizontal shear	 $k < 0$ $k > 0$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear	 $k < 0$ $k > 0$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

Transformations of the Plane

TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$

More Definitions

Definition (Onto and one-to-one)

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **onto** if each $\boldsymbol{b} \in \mathbb{R}^m$ is the image of *at least one* $\boldsymbol{x} \in \mathbb{R}^n$.

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **one-to-one** if each $\boldsymbol{b} \in \mathbb{R}^m$ is the image of *at most one* $\boldsymbol{x} \in \mathbb{R}^n$.

Example

Example

Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be given by $T(\boldsymbol{x}) = A\boldsymbol{x}$, where

$$A = \begin{bmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

Then (by considering the equation $A\boldsymbol{x} = \boldsymbol{b}$):

- T is onto.
- T is *not* one-to-one.

Theorems

Theorem (Theorem 11)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then* T *is one-to-one if and only if the equation* $T(\boldsymbol{x}) = \mathbf{0}$ *has only the trivial solution* $\boldsymbol{x} = \mathbf{0}$*.*

Theorem (Theorem 12)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation with standard matrix* A*. Then*

a. T *is onto if and only if the columns of* A *span* $\mathbb{R}^m$*.*

b. T *is one-to-one if the columns of* A *are linearly independent.*

Example

Let

$$T(x_1, x_2) = (3x_1 + x_2, 5x_1 + 7x_2, x_1 + 3x_2).$$

Show that T is a one-to-one linear transformation that is not onto.

Solution. We write $T : \mathbb{R}^2 \to \mathbb{R}^3$ as $T(\boldsymbol{x}) = A\boldsymbol{x}$, with

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

The columns are independent, but cannot span $\mathbb{R}^3$.

Section 1.10 - Linear Models in Business, Science, and Engineering

We focus on one example, namely, linear equations and electrical networks.

Ohm's law models the passage of current through a resistor by

$$V = RI,$$

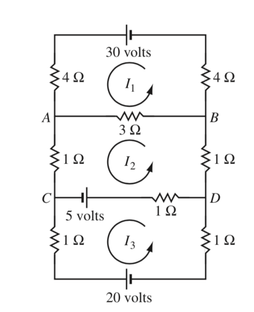
where

- V (voltage) is measured in volts,
- R (resistance) is measured in ohms,
- I (current flow) is measured in amps.

Example

Example

Determine the loop currents in the following circuit.



Example (continued)

Example (continued)

We need to use **Kirchhoff's voltage law:** the sum of the *RI* voltage drops in one direction around a loop equals the sum of the voltage sources in the same direction around the loop.

- For loop 1, we get
$$11I_1 - 3I_2 = 30.$$
- For loop 2, we get
$$-3I_1 + 6I_2 - I_3 = 5.$$
- For loop 3, we get
$$-I_2 + 3I_3 = -25.$$

Example (continued)

We get a linear system for I_1, I_2, I_3, which we can solve for

$$I_1 = 3, \quad I_2 = 1, \quad I_3 = -8.$$

We can use this to determine the current in each branch.

Math 3108 - Fall 2019
Chapter 2: Matrix Algebra

- The entries of an $m \times n$ matrix A are denoted a_{ij}.
- The diagonal entries are $a_{11}, a_{22}, \ldots$.
- The $n \times n$ identity matrix (denoted I_n or just I) is the diagonal matrix with 1s along the diagonals.
- The zero matrix (denoted by 0) has all $a_{ij} = 0$.

Sums and Scalar Multiples

Sums and scalar multiples of matrices are defined similarly to the case of vectors.

Example

Set

$$A = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix},\ B = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix},\ C = \begin{bmatrix} 2 & -3 \\ 0 & 1 \end{bmatrix}.$$

Then

$$A + B = \begin{bmatrix} 5 & 1 & 6 \\ 2 & 8 & 9 \end{bmatrix},$$

while $A + C$ is not defined. We also have

$$2B = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix}.$$

Matrix Algebra - Summary

THEOREM 1 Let A, B, and C be matrices of the same size, and let r and s be scalars.

a. $A + B = B + A$
b. $(A + B) + C = A + (B + C)$
c. $A + 0 = A$
d. $r(A + B) = rA + rB$
e. $(r + s)A = rA + sA$
f. $r(sA) = (rs)A$

- In other words, there is nothing unexpected when dealing with matrix addition and *scalar* multiplication.

DEFINITION If A is an $m \times n$ matrix, and if B is an $n \times p$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \ldots, A\mathbf{b}_p$. That is,

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

- Here $A\boldsymbol{b}_1, \ldots, A\boldsymbol{b}_p$ denote the matrix-vector multiplication we studied in Chapter 1.
- If A is the standard matrix of a transformation T and B is the standard matrix of a transformation S, then AB is the standard matrix of the **composition** $T \circ S$. This follows from the fact that

$$A(B\boldsymbol{x}) = (AB)\boldsymbol{x}.$$

Examples

Example

With

$$A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix},$$

we have

$$AB = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}.$$

Note that each column of AB is a linear combination of the columns of A.

Example

If A is 3×5 and B is 5×2, what are the sizes of AB and BA (if they are defined)?

Solution: AB is 3×2; BA is not defined.

Another Method to Compute AB

The ij^{th} entry of AB (if it is defined) is the 'dot product' between the i^{th} row of A and the j^{th} column of B:

$$(AB)_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}$$

when A has n columns and B has n rows.

Example

$$\begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}.$$

Properties of Matrix Multiplication

THEOREM 2

Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left distributive law)

c. $(B + C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_m A = A = A I_n$ (identity for matrix multiplication)

Matrix Multipication is Not Commutative

- We say A and B **commute** if $AB = BA$.
- In general, matrix multiplication is *not* commutative.

Example

$$\begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix},$$
$$\begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix}.$$

Other Differences

- In general $AB \neq BA$.
- If $AB = AC$, we cannot conclude $B = C$.
- If $AB = 0$, we cannot conclude that $A = 0$ or $B = 0$.

Other operations

- If A is an $n \times n$ matrix and k a positive integer, then A^k denotes $A \cdots A$ (k times).
- If A is an $m \times n$ matrix, then the **transpose** of A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A.

Example

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & 5 & -2 & 7 \end{bmatrix}^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix}.$$

- A convenient way to write a column vector is in the form $\boldsymbol{x} = [1, 2, 3]^T$.

Theorem about Transposes

THEOREM 3 Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$
b. $(A + B)^T = A^T + B^T$
c. For any scalar r, $(rA)^T = rA^T$
d. $(AB)^T = B^T A^T$

Pay special attention to the order of multiplication in part d.

Practice Problems

- Compute $\boldsymbol{x}\boldsymbol{x}^T$ and $\boldsymbol{x}^T\boldsymbol{x}$, where

$$\boldsymbol{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

- Let A be a 4×4 matrix and $\boldsymbol{x} \in \mathbb{R}^4$. What is the fastest way to compute $A^2\boldsymbol{x}$?

Definition (Inverse)

An $n \times n$ matrix A is **invertible** if there is an $n \times n$ matrix C such that

$$CA = AC = I_n.$$

In this case, C is an **inverse** of A.

Inverses are necessarily unique, and so we call C *the* inverse of A and write $C = A^{-1}$. Thus,

$$AA^{-1} = A^{-1}A = I_n.$$

A non-invertible matrix is called **singular**. An invertible matrix is called **nonsingular**.

Example

Example

If

$$A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix},$$

then A is invertible and

$$A^{-1} = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}.$$

The 2×2 case

Theorem (Theorem 4)

Let

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

If $ad - bc \neq 0$, *then* A *is invertible and*

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

If $ad - bc = 0$, *then* A *is not invertible.*

- The quantity $ad - bc$ is called the **determinant** of A.

Example

Example

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}^{-1} = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix}.$$

Usefulness of Matrix Inverses

Theorem (Theorem 5)

If A is an invertible $n \times n$ matrix, then for each $\boldsymbol{b} \in \mathbb{R}^n$ the equation $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

- To verify this, note $AA^{-1}\boldsymbol{b} = I_n\boldsymbol{b} = \boldsymbol{b}$.
- For uniqueness: if $A\boldsymbol{u} = \boldsymbol{b}$, then we apply A^{-1} to get $\boldsymbol{u} = A^{-1}\boldsymbol{b}$.

Example

Example

Solve the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

Solution: The system is equivalent to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

The solution is given by

$$\boldsymbol{x} = A^{-1}\boldsymbol{b} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

Properties of Matrix Inverses

THEOREM 6

a. If A is an invertible matrix, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

b. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

$$(AB)^{-1} = B^{-1}A^{-1}$$

c. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is,

$$(A^T)^{-1} = (A^{-1})^T$$

In general, the product of invertible matrices is invertible, with

$$[A_1 \cdots A_k]^{-1} = A_k^{-1} \cdots A_1^{-1}.$$

Elementary Matrices

Definition

An **elementary matrix** is a matrix obtained by performing a single elementary row operation on the identity matrix.

Example

Let E correspond to a row replacement, e.g.

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

Then

$$E \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \\ e-4a & f-4b \end{bmatrix}.$$

Elementary Matrices (Continued)

- Elementary matrices correspond to row replacement, row interchange, or scaling.
- If E is an elementary matrix corresponding to a row operation, then the product EA equals the matrix obtained by performing the same row operation on A.
- Every elementary matrix is invertible. To compute the inverse, just 'undo' the corresponding row operation.

Inverting Elementary Matrices

Example

Find the inverse of

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

We transform E back into I_3 by the row operation

$$R_3 \mapsto R_3 + 4R_1,$$

which corresponds to

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

Computing Matrix Inverses

Our method to compute matrix inverses is based off of the following theorem:

Theorem (Theorem 7)

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n.

In this case, if the row operations $E_1, \ldots, E_k$ reduce A to I_n, then the same row operations transform I_n into A^{-1}. In particular,

$$A^{-1} = E_k \cdots E_1.$$

- A invertible $\implies$ $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b}$.
- n pivots $\iff$ invertible.
- $E_k \cdots E_1 A = I_n \implies A^{-1} = E_k \cdots E_1$.

Computing A^{-1}

- Row reduce $[A \;\; I]$.
  - If $A \sim I$, then $[A \;\; I] \sim [I \;\; A^{-1}]$.
  - Otherwise, A is not invertible.

Example

Determine whether

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$$

is invertible. If so, compute its inverse.

Solution:

$$A^{-1} = \begin{bmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}.$$

Another Viewpoint

- When we row reduce $[A\, I]$, we are simultaneously solving $A\boldsymbol{x} = \boldsymbol{e}_j$ for each $j = 1, \ldots, n$.
- The columns of A^{-1} are then the solutions to each of these equations.

THEOREM 8

The Invertible Matrix Theorem

Let A be a square $n \times n$ matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix.
b. A is row equivalent to the $n \times n$ identity matrix.
c. A has n pivot positions.
d. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
e. The columns of A form a linearly independent set.
f. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
g. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.
h. The columns of A span $\mathbb{R}^n$.
i. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
j. There is an $n \times n$ matrix C such that $CA = I$.
k. There is an $n \times n$ matrix D such that $AD = I$.
l. A^T is an invertible matrix.

Strategy of Proof

$$\begin{tikzcd}
& \text{(a)} \arrow[dr, Rightarrow] & \\
\text{(b)} \arrow[ur, Rightarrow] & & \text{(j)} \arrow[dl, Rightarrow] \\
\text{(c)} \arrow[u, Rightarrow] & \text{(d)} \arrow[l, Rightarrow] &
\end{tikzcd}$$

FIGURE 1

$$\begin{tikzcd}
& \text{(k)} \arrow[dr, Rightarrow] & \\
\text{(a)} \arrow[ur, Rightarrow] & & \text{(g)} \arrow[ll, Rightarrow]
\end{tikzcd}$$

$$\text{(g)} \Leftrightarrow \text{(h)} \Leftrightarrow \text{(i)}$$

$$\text{(d)} \Leftrightarrow \text{(e)} \Leftrightarrow \text{(f)}$$

$$\text{(a)} \Leftrightarrow \text{(l)}$$

A Useful Fact

- Let A, B be square matrices. If $AB = I$, then A and B are both invertible, with $A = B^{-1}$ and $B = A^{-1}$.

This uses items j. and k. from the invertible matrix theorem, along with the uniqueness of inverses.

Example

Determine whether the matrix A invertible, where

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix}.$$

Solution: Perform row reduction to get

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

As A has three pivots, it is invertible.

Invertible Linear Transformations

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **invertible** if there exists a transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ so that

$$S(T(\boldsymbol{x})) = T(S(\boldsymbol{x})) = \boldsymbol{x} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Theorem (Theorem 9)

A linear transformation T is invertible if and only if its standard matrix A is invertible. In this case, $S(\boldsymbol{x}) := A^{-1}\boldsymbol{x}$ is the inverse of T; in particular, S is also a linear transformation.

- Show that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one linear transformation, then T is invertible.
- Determine whether or not

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

is invertible.

- A **factorization** of a matrix A is an equation that expresses A as a product of two or more matrices.
- Matrix factorizations play an important role in applications, e.g. the singular value decomposition in machine learning (to be discussed later).
- In this section we focus on the LU factorization, which is used to efficiently solve sequences of equations all with the same coefficient matrix.

LU Factorization

- Suppose A is an $m \times n$ matrix that can be reduced to echelon form **without row interchanges**.
- This means A can be written in the form $A = LU$, where
  - L is an $m \times m$ **unit lower triangular** matrix.
  - U is an $m \times n$ echelon form of A, which is upper triangular.
- To solve $A\boldsymbol{x} = \boldsymbol{b}$, we can equivalently solve the pair of equations

$$L\boldsymbol{y} = \boldsymbol{b}, \quad U\boldsymbol{x} = \boldsymbol{y}.$$

Each equation can be solved quickly because L and U are triangular.

Example

Example

Use the LU factorization

$$\underbrace{\begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}}_{U}$$

to solve $A\boldsymbol{x} = \boldsymbol{b}$, where

$$b = [-9\ 5\ 7\ 11]^T.$$

Example (continued)

Example (Continued)

- First solve $L\boldsymbol{y} = \boldsymbol{b}$:

$$[L\ \boldsymbol{b}] \sim [I\ \ \boldsymbol{y}], \quad \boldsymbol{y} = \begin{bmatrix} -9 \\ -4 \\ 5 \\ 1 \end{bmatrix}.$$

- Then solve $U\boldsymbol{x} = \boldsymbol{y}$:

$$[U\,\boldsymbol{y}] \sim [I\,\boldsymbol{x}], \quad \boldsymbol{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \\ -1 \end{bmatrix}.$$

Computational Efficiency

- In the previous example, once we have determined L and U, it takes 12 arithmetic operations to find $\boldsymbol{y}$, followed by 28 arithmetic operations to find $\boldsymbol{x}$.
- By contrast, direct row reduction of $[A\ \boldsymbol{b}]$ to $[I\ \boldsymbol{x}]$ requires 62 operations.
- Thus, LU decomposition can increase computational efficiency in cases in which one needs to solve $A\boldsymbol{x} = \boldsymbol{b}$ for a fixed A but many different choices of $\boldsymbol{b}$.

LU Algorithm

- Suppose A can be reduced to an echelon form U using only replacements that add a multiple of one row to another row below it.
- Then there exist unit lower triangular elementary matrices $E_1, \dots E_p$ so that
$$E_p \cdots E_1 A = U.$$
This gives us a choice of U, and we may take
$$L = [E_p \cdots E_1]^{-1}.$$
[*Remark:* Why is L unit lower triangular?]
- These same row operations reduce L to I.

Example

Example

Find an LU factorization of

$$A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & -1 \end{bmatrix}.$$

Solution:

$$\begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & -9 & -3 & -4 & 10 \\ 0 & 12 & 4 & 12 & -5 \end{bmatrix}$$

Example (Continued)

$$\sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \sim \underbrace{\begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}}_{U}.$$

We take

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

General Case

- In general, one needs to use row interchange when performing row reduction.
- In this case, the 'L' that one produces is a permutation of a lower triangular matrix.

Sample Problem

Example

Find an LU factorization of

$$A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix}.$$

Note: A has only three pivots; the final two columns of L will come from I_5.

- Suppose nation's economy has n sectors.
- $\boldsymbol{x} \in \mathbb{R}^n$: production vector
- $\boldsymbol{d} \in \mathbb{R}^n$: final demand vector
- C: $n \times n$ consumption matrix. [*For each sector, how many units of each other sector are consumed per unit of output?*]
- $C\boldsymbol{x} \in \mathbb{R}^n$: intermediate demand vector
- Leontief Input-Output Model:

$$\boldsymbol{x} = C\boldsymbol{x} + \boldsymbol{d}$$

A relevant theorem

Theorem (Theorem 11)

If C and $\boldsymbol{d}$ have nonnegative entries and each column sum of C is less than 1, then

$$\boldsymbol{x} = (I - C)^{-1}\boldsymbol{d}$$

has nonnegative entries and is the unique solution to $\boldsymbol{x} = C\boldsymbol{x} + \boldsymbol{d}$.

- To approximate $(I - C)^{-1}$, use a Taylor series expansion:

$$(I - C)^{-1} = I + C + C^2 + C^3 + \dots$$

- The entries in $(I - C)^{-1}$ can be used to predict how the production $\boldsymbol{x}$ must change in response to a change in the final demand $\boldsymbol{d}$.

Example

Example

An economy has three sectors: manufacturing, agriculture, and services, with consumption matrix

$$C = \begin{bmatrix} .5 & .4 & .2 \\ .2 & .3 & .1 \\ .1 & .1 & .3 \end{bmatrix}.$$

Suppose the final demand is $\boldsymbol{d} = [50\ 30\ 20]^T$. Find the production level $\boldsymbol{x}$ that satisfies this demand.

Solution: We solve $(I - C)\boldsymbol{x} = \boldsymbol{d}$ by row reduction to deduce

$$\boldsymbol{x} = [226\ 119\ 78]^T.$$

- In this section we describe some basic applications of linear algebra to $2D$ computer graphics.

Basic Example

Example

We can represent a letter (say N) by using eight points in the plane. We store this in a data matrix D, say

$$D = \begin{bmatrix} 0 & .5 & .5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix}.$$

Each column corresponds to a vertex in the $2D$ plane.

By applying the shear transformation

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

we can shear the N.

Matlab Code

```
% Plot the letter N with successive shear transformations applied to it.

% Data Matrices
D = [0, .5, .5, 6, 6, 5.5, 5.5, 0; 0, 0, 6.42, 0, 8, 8, 1.58, 8];
DD = [D(:,8),D(:,1)];

% Shear transformation of the plane
A = [1,.1;0,1];

figure

for m=1:75
    clf                         %clear the figure
    D = A*D;                    %update the data matrix
    DD = [D(:,8),D(:,1)];
    plot(D(1,:),D(2,:))
    hold on
    plot(DD(1,:),DD(2,:))
    axis([-2 70 -2 10])
    drawnow
end

```

Homogeneous Coordinates

- Translation is **not** a linear transformation of the plane — indeed, it does not send $\mathbf{0}$ to $\mathbf{0}$.
- However, we can model translation of the $2D$ plane using a $3D$ linear transformation together with **homogeneous coordinates**.
- In particular, we associate a point $(x, y) \in \mathbb{R}^2$ with the point $(x, y, 1) \in \mathbb{R}^3$.
- Then translation by the vector $[h, k]^T$ is represented by the matrix

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix},$$

which sends

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} x+h \\ y+k \\ 1 \end{bmatrix}.$$

$2D$ Transformations in Homogeneous Coordinates

- We can still model a $2D$ linear transformation using homogeneous coordinates. In particular, if the transformation has the 2×2 standard matrix A, then we apply the matrix

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

to the homogeneous coordinates, sending

$$\begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} A\boldsymbol{x} \\ 1 \end{bmatrix}.$$

- Composition of transformations corresponds to matrix multiplication (even in the setting of homogeneous coordinates).

Example: rotation about a point

Example

Find the matrix that performs rotation by angle ϕ about a $\boldsymbol{p}$ in $\mathbb{R}^2$.

Solution. We use homogeneous coordinates $[x\ y\ 1]^T$.

- We first translate by $-\boldsymbol{p}$ via

$$T_- = \begin{bmatrix} 1 & 0 & -p_1 \\ 0 & 1 & -p_2 \\ 0 & 0 & 1 \end{bmatrix}$$

- Now perform a rotation by angle ϕ about the origin:

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

- Translate back to $\boldsymbol{p}$ via T_+.
- The transformation is given by the product T_+RT_-.

Matlab Code

```matlab
%each column gives the position of an object in homogeneous coordinates
D=[0,-5;0,0;1,1];

%object one will translate along the line y=x
%object two will orbit object one

%translation by (.1,.1)
T = [1, 0, .1; 0, 1, .1; 0, 0, 1];

%rotation by angle .1
R=[cos(.1),-sin(.1),0;sin(.1),cos(.1),0;0,0,1];

figure

for m=1:250
    clf
    D(:,1)=T*D(:,1);    %update object one

    %object two should be translated, then
    %rotated about object one's current position
    Tminus = [1,0,-D(1,1);0,1,-D(2,1);0,0,1];
    Tplus = [1,0,D(1,1);0,1,D(2,1);0,0,1];

    D(:,2)=Tplus*R*Tminus*T*D(:,2); %update object two

    scatter(D(1,1:2),D(2,1:2),50) %when plotting, omit final row of 1s
    axis([-50 50 -50 50])
    drawnow

end
```

Further Topics

See the textbook for further discussion of $3D$ graphics, homogeneous coordinates in $3D$, and perspective projections.

Definition (Subspace)

A **subspace** of $\mathbb{R}^n$ is a set H in $\mathbb{R}^n$ satisfying the following three properties:

a. H contains the zero vector $\mathbf{0}$.
b. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are in H, then $\boldsymbol{u} + \boldsymbol{v}$ is in H. [*Closed under addition.*]
c. If $\boldsymbol{u}$ is in H and c is a scalar, then $c\boldsymbol{u}$ is in H. [*Closed under scalar multiplication.*]

A Key Example

Example (Span)

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors in $\mathbb{R}^n$ and

$$H = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\},$$

then H is a subspace of $\mathbb{R}^n$.

The same is true for any finite collection of vectors in $\mathbb{R}^n$.

Later we will see that *every* subspace is of this form!

- In the previous example, H is either a line (if $\{\mathbf{v}_1, \mathbf{v}_2\}$ are dependent) or a plane (if $\{\mathbf{v}_1, \mathbf{v}_2\}$ are independent).
- A line or a plane that does not pass through the origin is not a subspace.

Column Space and Null Space

Definition (Column Space)

The **column space** of a matrix A, denoted $\mathrm{Col}\,(A)$, is the set of all linear combinations of the columns of A.

- If A is $m \times n$, then the column space of A is the span of the columns of A and hence is a subspace of $\mathbb{R}^n$.

Definition (Null Space)

The **null space** of a matrix A, denoted $\mathrm{Nul}\,(A)$, is the set of all solutions $\boldsymbol{x}$ to the homogeneous equation $A\boldsymbol{x} = \mathbf{0}$.

Theorem (Theorem 12)

If A is an $m \times n$ matrix then $Nul\,(A)$ is a subspace of $\mathbb{R}^n$.

Example

Determine whether $\boldsymbol{b}$ is in $\operatorname{Col}(A)$, where

$$A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}.$$

Solution: We must determine whether $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. As

$$[A\ \boldsymbol{b}] \sim \begin{bmatrix} 1 & -3 & -4 & -3 \\ 0 & -6 & -18 & 15 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we see that $\boldsymbol{b} \in \operatorname{Col}(A)$.

Bases

Definition

A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.

Example

The **standard basis** for $\mathbb{R}^n$ consists of the vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Example

The columns of any invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

Basis for the Null Space

Example

Find a basis for the null space of

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Solution: Write the solution to $A\mathbf{x} = \mathbf{0}$ in parametric vector form:

$$A \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

so x_2, x_4, x_5 are free, with

$$x_1 = 2x_2 + x_4 - 3x_5, \quad x_3 = -2x_4 + 2x_5.$$

Basis for the Null Space (Continued)

Example (Continued)

The general solution is

$$\mathbf{x} = x_2 \underbrace{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{:=\mathbf{u}} + x_4 \underbrace{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}}_{:=\mathbf{v}} + x_5 \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}_{:=\mathbf{w}},$$

from which we can deduce $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis for $\mathrm{Nul}\,(A)$.

Basis for the Column Space

Example

Find a basis for the column space of

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}.$$

Solution. The columns of A span $\mathrm{Col}\,(A)$, but they are not independent.

$$A \sim \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Basis for the Column Space (continued)

Example (Continued)

Keeping the pivot columns of A, we obtain the basis

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}.$$

Pivot Columns; Practice Problems

Theorem (Theorem 13)

The pivot columns of a matrix A form a basis for the column space of A.

- Given
$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$
find a vector in $\operatorname{Nul}(A)$ and a vector in $\operatorname{Col}(A)$.
- Suppose an $n \times n$ matrix A is invertible. What can you say about $\operatorname{Col}(A)$? What can you say about $\operatorname{Nul}(A)$?

Definition (Coordinates)

If $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p\}$ is a basis for a subspace H in $\mathbb{R}^n$, then any $\boldsymbol{x}$ in H may be written *uniquely* in the form

$$\boldsymbol{x} = c_1\boldsymbol{b}_1 + \cdots + c_p\boldsymbol{b}_p$$

for some weights $c_1, \ldots, c_p$. We define the **coordinates of x relative to the basis B** by

$$[x]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

- Uniqueness is due to linear independence.

Example

Let

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}.$$

Then $B = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

(i) Show that x belongs to H.

(ii) Find $[x]_B$ (the coordinates of $\mathbf{x}$ relative to B).

Example (continued)

Example

Solution: We solve

$$\begin{bmatrix} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix},$$

which shows that $\boldsymbol{x} \in \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ with

$$[\boldsymbol{x}]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

- If a subspace of H has a basis consisting of p vectors, then *every* basis of H must have exactly p vectors.

Definition (Dimension)

The **dimension** of a nonzero subspace H, denoted by $\dim H$, is the number of vectors in any basis for H. The dimension of the subspace $\{\mathbf{0}\}$ is defined to be zero.

Definition (Rank)

The **rank** of a matrix A, denoted $\operatorname{rank} A$, is the dimension of the column space of A.

Example

The dimension of the null space of a matrix A is the number of free variables in the equation $A\boldsymbol{x} = \mathbf{0}$.

Rank Theorem; Basis Theorem

Theorem (The Rank Theorem)

If a matrix A has n columns, then

$$\text{rank}\, A + \dim \text{Nul}\, A = n.$$

Proof: Every column is either a pivot column or leads to a free variable in the equation $A\mathbf{x} = \mathbf{0}$.

Theorem (The Basis Theorem)

Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of p elements of H is a basis for H; any set of p elements of H that spans H is a basis for H.

Continuation of Invertible Matrix Theorem

Theorem (The Invertible Matrix Theorem (continued))

Let A be an $n \times n$ matrix. The following are equivalent to the statement that A is invertible:

m. *The columns of A form a basis for $\mathbb{R}^n$.*

n. $\mathit{Col}\, A = \mathbb{R}^n$.

o. $\dim \mathit{Col}\, A = n$.

p. $\mathit{rank}\, A = n$.

q. $\mathit{Nul}\, A = \{\mathbf{0}\}$.

r. $\dim \mathit{Nul}\, A = 0$.

Practice Problems

- Is $\mathbb{R}^3$ a subspace of $\mathbb{R}^4$?
- What is the basis of the subspace of $\mathbb{R}^3$ spanned by

$$\begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ -7 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 6 \\ 7 \end{bmatrix}.$$

- Let B be the basis for $\mathbb{R}^2$ with elements $[1\ 2]^T$ and $[2\ 1]^T$. If $[\boldsymbol{x}]_B = [3\ 2]^T$, then what is $\boldsymbol{x}$?

Math 3108 - Fall 2019
Chapter 3: Determinants

We encountered the determinant of a 2×2 matrix when discussing invertibility. We now extend this notion to higher order matrices.

- The determinant of a 1×1 matrix $A = [a_{11}]$ is simply

$$\det A = a_{11}.$$

- The determinant of a 2×2 matrix $A = [a_{ij}]$ is

$$\det A = a_{11}a_{22} - a_{12}a_{21}.$$

- To describe the determinant of higher order (square) matrices, we need to introduce the notion of a **submatrix**.

Higher Order Determinants

- Given an $n \times n$ matrix, the **submatrix** A_{ij} is the $(n-1) \times (n-1)$ matrix obtained by removing row i and column j from A.
- The determinant of a 3×3 matrix A is

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}.$$

Example

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

Determinants - General Definition

The general definition of the determinant is 'inductive':

Definition (Determinant)

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is given by the alternating sum

$$\begin{aligned} \det A &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \\ &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n}. \end{aligned}$$

Here A_{ij} denotes the $(n-1) \times (n-1)$ submatrix of A obtained by removing row i and column j.

- We may also write $|A|$ for $\det A$.

Cofactor Expansions

- There are more ways to compute the determinant.
- The (i,j) **cofactor** of A is defined by

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

- The definition of determinant uses a 'cofactor expansion across the first row'.

Theorem (Theorem 1)

The determinant of an $n \times n$ matrix A can be computed using the cofactor expansion across any row or column. That is:

$$\begin{aligned} \det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \quad \text{for any} \quad i \\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \quad \text{for any} \quad j. \end{aligned}$$

Example

Compute $\det A$, where

$$A = \begin{bmatrix} 3 & -7 & 8 & 9 & -6 \\ 0 & 2 & -5 & 7 & 3 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 0 \end{bmatrix}.$$

Solution. Choose the most convenient cofactor expansions:

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

$$= 6 \cdot (-1) \cdot (-2) \cdot \det \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} = 12.$$

Special Case: Triangular Matrices

Theorem (Theorem 2)

If A is a triangular matrix, then $\det A$ is the product of the entries along the diagonal of A.

- In general, cofactor expansion of an $n \times n$ matrix requires more than $n!$ multiplications.
- This means that even for a 25×25 matrix (say), with a calculator performing one trillion multiplications per second, computing the determinant would take several hundred thousand years...

Practice Problem

- Compute

$$\det \begin{bmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{bmatrix}.$$

- If two matrices are connected by row operations, their determinants are related as well.

Theorem (Theorem 3 - Row Operations and Determinants)

Let A be a square matrix.

- *If B is obtained from A by a row replacement, then* $\det A = \det B$.
- *If B is obtained from A by a row interchange, then* $\det B = -\det A$.
- *If B is obtained by scaling a row of A by k, then* $\det B = k \cdot \det A$.

- This means we can use row reduction to efficiently compute determinants!

Example

Compute $\det A$, where

$$A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}.$$

Solution: Using two replacements and one interchange,

$$\begin{aligned} A \sim \begin{bmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{bmatrix} &\sim \begin{bmatrix} 1 & -4 & 2 \\ -1 & 7 & 0 \\ 0 & 0 & -5 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & -5 \end{bmatrix}. \end{aligned}$$

Thus $\det A = 15$.

Another Example

Example

Compute

$$\det \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix} = -36.$$

Generalizations

- In general, we deduce that $\det A$ either equals
  - 0, if A is not invertible (not equivalent to I_n), or
  - $\pm$ the product of the pivots in any echelon form of A.

Theorem (Theorem 4)

A square matrix A is invertible if and only if $\det A \neq 0$.

Example

Compute det A, where

$$A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}.$$

Solution: Adding 2 times row 1 to row 3 yields the matrix

$$\begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix}.$$

Thus det $A = 0$.

- Computer programs use this 'row reduction' method to compute $\det A$. This requires about $2n^3/3$ operations. Thus only 10,000 operations are required for a 25×25 matrix, which takes a fraction of a second.
- Cofactor expansion can be used together with row reduction.

Example

Compute the determinant of

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

Example (continued)

Example

Compute the determinant of

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

Answer: -30.

Column Operations

Theorem (Theorem 5)

If A is an $n \times n$ matrix, then $\det A^T = \det A$.

(Recall that A^T is the transpose of A, obtained by interchanging the rows and columns of A.)

- The proof is by induction and cofactor expansion.
- This theorem shows that 'column operations' have the same effect on determinants as row operations.
- We focus on row operations.

Determinants and Matrix Products

Theorem (Theorem 6)

If A and B are $n \times n$ matrices, then

$$\det AB = (\det A) \cdot (\det B).$$

- We won't prove this, but at least let's see it in action!

Example

First, compute

$$\begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 14 & 13 \end{bmatrix}.$$

Next, observe

$$9 \cdot 5 = 45 = 325 - 280.$$

Practice Problems

- Use a determinant to determine if the following three vectors are independent:

$$\begin{bmatrix} 5 \\ -7 \\ 9 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 3 \\ -5 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -7 \\ 5 \end{bmatrix}.$$

- Suppose A is $n \times n$ and $A^2 = I$. Show that $\det A$ equals 1 or -1.

Section 3.3 - Cramer's Rule, Volume, and Linear Transformations

- In this section we will briefly mention some further applications of determinants.

Theorem (Theorem 7 - Cramer's Rule)

Let A be an $n \times n$ invertible matrix and $\boldsymbol{b} \in \mathbb{R}^n$. Then unique solution of $A\boldsymbol{x} = \boldsymbol{b}$ has entries given by

$$x_i = \frac{\det A_i(\boldsymbol{b})}{\det A},$$

where $A_i(\boldsymbol{b})$ is the matrix obtained from A by replacing column i with the vector $\boldsymbol{b}$.

- *Application:* In engineering, systems of differential equations are converted to systems of algebraic equations by the Laplace transform. These systems may then be solved by Cramer's rule.

A Formula for Matrix Inverses

- Since the j^{th} column of A^{-1} is the solution to $A\boldsymbol{x} = \boldsymbol{e}_j$, Cramer's rule implies
$$A^{-1}_{ij} = \frac{\det A_i(e_j)}{\det A},$$
where the notation $A_i(\cdot)$ is as in Cramer's theorem.
- By cofactor expansion, we have
$$\det A_i(\boldsymbol{e}_j) = C_{ji},$$
where C_{ji} is the cofactor introduced above. So we can also write
$$A^{-1}_{ij} = \tfrac{1}{\det A} C_{ji}.$$

Geometric Interpretation of Determinant

Theorem (Theorem 9)

- *If A is a 2×2 matrix, then the area of the parallelogram determined by the columns of A is equal to $|\det A|$.*
- *If A is a 3×3 matrix, then the volume of the parallelepiped determined by the columns of A is $|\det A|$.*

- Proof (sketch): It is true for diagonal matrices, and so you need to check what happens under row operations.

Linear Transformations

Theorem

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$Area\{T(S)\} = |\det A| \cdot Area(S).$$

If instead $T : \mathbb{R}^3 \to \mathbb{R}^3$ has standard matrix A and S is a parallelepiped in $\mathbb{R}^3$, then

$$Volume\{T(S)\} = |\det A| \cdot Volume(S).$$

- This generalizes to any region S in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Math 3108 - Fall 2019
Chapter 4: Vector Spaces

DEFINITION

A **vector space** is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms (or rules) listed below.[1] The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u}+\mathbf{v}$, is in V.
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.
4. There is a **zero** vector $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.
6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.
9. $c(d\mathbf{u})=(cd)\mathbf{u}$.
10. $1\mathbf{u}=\mathbf{u}$.

- We may also use complex vectors and complex scalars.

Examples

- The fundamental example in this class is $V = \mathbb{R}^n$.
- Let $\mathbb{S}$ be the space of all doubly infinite sequences of numbers

$$\{y_k\} = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots).$$

- For $n \geq 0$, let $\mathbb{P}_n$ be the set of all polynomials of the form

$$\boldsymbol{p}(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n.$$

- Let $F(\mathbb{R})$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$.
- Let $C(\mathbb{R})$ be the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$.
- And so on...

An Important Question

What is a vector?

Subspaces

DEFINITION A **subspace** of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.[2]

b. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.

c. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

- Note that any subspace is itself a vector space.

Examples

- The **zero subspace** is the subspace $\{0\}$.
- For any n, $\mathbb{P}_n$ is a subspace of the vector space $\mathbb{P}$ of all polynomials, which is in turn a subspace of $C(\mathbb{R})$, which is a subspace of $F(\mathbb{R})$.
- $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$, but the set

$$H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s, t \in \mathbb{R} \right\}$$

  is a subspace of $\mathbb{R}^3$.
- A plane in $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$ if and only if it contains the zero vector.

Subspaces Spanned by a Set

In the setting of a general vector space, we still have the notions of **linear combination** and **span**.

Theorem (Theorem 1)

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in a vector space V, then Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

- To prove this, you must check the definition of subspace.

Example

The set of all vectors of the form $(a-3b, b-a, a, b)$ is a subspace, since it is equal to the span of

$$\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Practice Problems

- The set of points of the form $(3s, 2+5s)$ is not a vector space.
- Show that the set of symmetric $n \times n$ matrices is a subspace of the vector space of all $n \times n$ matrices.

MyLab Problems

Determine if the given set is a subspace of $\mathbb{P}_7$. Justify your answer.

The set of all polynomials of the form $\mathbf{p}(t) = at^7$, where a is in $\mathbb{R}$.

Choose the correct answer below.

- **A.** The set is not a subspace of $\mathbb{P}_7$. The set is not closed under multiplication by scalars when the scalar is not an integer.
- **B.** The set is a subspace of $\mathbb{P}_7$. The set contains the zero vector of $\mathbb{P}_7$, the set is closed under vector addition, and the set is closed under multiplication by scalars.
- **C.** The set is not a subspace of $\mathbb{P}_7$. The set does not contain the zero vector of $\mathbb{P}_7$.
- **D.** The set is a subspace of $\mathbb{P}_7$. The set contains the zero vector of $\mathbb{P}_7$, the set is closed under vector addition, and the set is closed under multiplication on the left by $m \times 7$ matrices where m is any positive integer.

MyLab Problems

Let W be the set of all vectors of the form shown on the right, where a and b represent arbitrary real numbers. Find a set S of vectors that spans W, or give an example or an explanation showing why W is not a vector space.

$$\begin{bmatrix} -a+8 \\ a-5b \\ 2b+a \end{bmatrix}$$

Select the correct choice below and, if necessary, fill in the answer box to complete your choice.

- **A.** The set W is a vector space and a spanning set is S = { }.
  (Use a comma to separate vectors as needed.)
- **B.** The set W is not a vector space because not all vectors **u** in W have the property 1**u** = **u**.
- **C.** The set W is not a vector space because not all vectors **u**, **v**, and **w** in W have the property that **u** + **v** = **v** + **u** and (**u** + **v**) + **w** = **u** + (**v** + **w**).
- **D.** The set W is not a vector space because the zero vector is not in W.

Section 4.2 - Null Spaces, Column Spaces, and Linear Transformations

- Recall that we studied null spaces and column spaces of matrices in Chapter 2.

Definition

The **null space** of an $m \times n$ matrix A is

$$\text{Nul}\, A = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \mathbf{0}\}.$$

Theorem (Theorem 2)

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$.

Column Space

Definition

The **column space** of an $m \times n$ matrix

$$A = [\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_n]$$

is defined by

$$\operatorname{Col} A = \operatorname{Span}\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}.$$

Theorem (Theorem 3)

The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$.

Note that

$$\operatorname{Col} A = \{\boldsymbol{b} \in \mathbb{R}^m : A\boldsymbol{x} = \boldsymbol{b} \quad \text{for some} \quad \boldsymbol{x} \in \mathbb{R}^n\}.$$

Examples

Example

Show that the set of vectors in $\mathbb{R}^4$ whose coordinates a, b, c, d satisfy

$$a - 2b + 5c = d \quad \text{and} \quad c - a = b \quad \text{is a subspace.}$$

Solution. The set is the same as the null space of

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

Example

Write the set

$$\begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix}, \quad a, b \in \mathbb{R} \quad \text{as the column space of a matrix.}$$

Solution. $A = \begin{bmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{bmatrix}.$

Null Space Versus Column Space

Contrast Between Nul A and Col A for an m x n Matrix A

Nul A	Col A
1. Nul A is a subspace of $\mathbb{R}^n$.	**1.** Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2.** Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	**3.** It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4.** There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5.** A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6.** Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7.** Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8.** Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

Linear Transformations (General Case)

Definition

A **linear transformation** T from a vector space V to a vector space W is a function $T : V \to W$ such that

(i) $T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for all vectors $\boldsymbol{u}, \boldsymbol{v} \in V$,

(ii) $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all vectors $\boldsymbol{u} \in V$ and scalars c.

Here are some examples:

- If A is an $m \times n$ matrix, then $T(\boldsymbol{x}) = A\boldsymbol{x}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- If V is the vector space of differentiable functions, then $Tf = \frac{d}{dx} f$ is a linear transformation from V to $F(\mathbb{R})$.

Kernel, Range

Definition

Let $T : V \to W$ be a linear transformation.

The **kernel** (also called **null space**) of a linear transformation T is the set of all vectors $\boldsymbol{u}$ such that $T(\boldsymbol{u}) = 0$.

The **range** of T is the set of all vectors of the form $T(\boldsymbol{x})$ for $\boldsymbol{x} \in V$.

- Note that if $T : V \to W$ is a linear transformation, then the kernel of T is a subspace of V and the range of T is a subspace of W.

Example

Example

Let $\omega \in \mathbb{R}$ and let T be the linear transformation

$$T = \tfrac{d^2}{dx^2} + \omega^2.$$

Then the kernel of T is the set of solutions to the differential equation

$$y'' + \omega^2 y = 0.$$

In particular, the set of solutions forms a vector space.

(In fact, this is a two-dimensional vector space, and a basis is given by the functions $\{\cos(\omega t), \sin(\omega t)\}$.)

Practice Problems

- Let A be an $n \times n$ matrix. Suppose $\operatorname{Col} A = \operatorname{Nul} A$. Show that $\operatorname{Nul} A^2 = \mathbb{R}^n$.

Solution. For any $\boldsymbol{x} \in \mathbb{R}^n$, $A\boldsymbol{x}$ belongs to the column space, and hence the null space of A. Thus

$$A^2\boldsymbol{x} = A(A\boldsymbol{x}) = 0.$$

This means A^2 is the zero matrix, so $\operatorname{Nul} A^2 = \mathbb{R}^n$.

Consider the following two systems of equations.

$$\begin{aligned} 5x_1 + 2x_2 - 3x_3 &= 0 \\ -9x_1 + 5x_2 + 7x_3 &= -4 \\ 4x_1 + 2x_2 - 8x_3 &= 14 \end{aligned} \qquad\qquad \begin{aligned} 5x_1 + 2x_2 - 3x_3 &= 0 \\ -9x_1 + 5x_2 + 7x_3 &= -12 \\ 4x_1 + 2x_2 - 8x_3 &= 42 \end{aligned}$$

It can be shown that the first system has a solution. Use this fact and the theory of null spaces and column spaces of matrices to explain why the second system must also have a solution. (Make no row operations.)

MyLab Problems

Define a linear transformation $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(0) \end{bmatrix}$. Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T, and describe the range of T.

Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T. Choose the correct answer below.

- **A.** $\mathbf{p}_1(t) = t$ and $\mathbf{p}_2(t) = t^2 - 1$
- **B.** $\mathbf{p}_1(t) = t$ and $\mathbf{p}_2(t) = t^3$
- **C.** $\mathbf{p}_1(t) = t$ and $\mathbf{p}_2(t) = t^2$
- **D.** $\mathbf{p}_1(t) = 1$ and $\mathbf{p}_2(t) = t^2$
- **E.** $\mathbf{p}_1(t) = 3t^2 + 5t$ and $\mathbf{p}_2(t) = 3t^2 - 5t + 7$
- **F.** $\mathbf{p}_1(t) = t + 1$ and $\mathbf{p}_2(t) = t^2$
- **G.** $\mathbf{p}_1(t) = t^2$ and $\mathbf{p}_2(t) = -t^2$

The definition of linear independence in a general vector space is identical to the definition in $\mathbb{R}^n$:

Definition (Linearly Independent)

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in a vector space V is **linearly independent** if the equation

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution $c_1 = \cdots = c_p = 0$.

Otherwise, we call the set **linearly dependent**.

Theorem (Theorem 4)

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors with $\mathbf{v}_1 \neq 0$ is linearly dependent if and only if some $\mathbf{v}_j$ (with $j > 1$) can be written as a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

Examples

- In a general vector space, the equation

$$c_1 \boldsymbol{v}_1 + \cdots + c_p \boldsymbol{v}_p = \mathbf{0}$$

cannot generally be written as a matrix vector equation.

Example

The polynomials $\boldsymbol{p}_1(t) = 1$, $\boldsymbol{p}_2(t) = t$, and $\boldsymbol{p}_3(t) = 4 - t$ are linearly dependent in $\mathbb{P}$ since $\boldsymbol{p}_3 = 4\boldsymbol{p}_1 - \boldsymbol{p}_2$.

Example

The set $\{\sin t, \cos t\}$ is linearly independent in $F(\mathbb{R})$. The set $\{\sin t \cos t, \sin 2t\}$ is linearly dependent.

Bases

The definition of a basis in a general vector space is also the same as in the setting of $\mathbb{R}^n$:

Definition (Basis)

Let H be a subspace of a vector space V. A set $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p\}$ in V is a **basis** for H if:

(i) B is a linearly independent set, and

(ii) $H = \text{Span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p\}$.

Examples

Example

All of the old examples from $\mathbb{R}^n$ are pertinent.

Example

The set $S = \{1, t, t^2, \ldots, t^n\}$ is a basis for $\mathbb{P}_n$. This is the **standard basis** for $\mathbb{P}_n$.

Example (Fourier Series)

The set containing $\{\sin(nt), \cos(nt)\}$, where $n = 0, 1, 2, \ldots$ is a basis for square-integrable periodic functions on $[-\pi, \pi]$ (written $L^2(\mathbb{T})$).

A More Familiar Example

Example

Let

$$\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}$$

and set $H = \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$.

Since $\boldsymbol{v}_3 = 5\boldsymbol{v}_1 + 3\boldsymbol{v}_2$, we may actually write

$$H = \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}.$$

In particular, $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a basis for H.

Spanning Set Theorem

Theorem (Theorem 5)

Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a set in a vector space V and let $H = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

a. *If a vector $\mathbf{v}_k$ in S is a linear combination of the other vectors in S, then the set obtained by removing $\mathbf{v}_k$ from S still spans H.*

b. *If $H \neq \{\mathbf{0}\}$, then some subset of S is a basis for H.*

Bases for Nul A and Col A.

- Recall that to find a basis for the null space of a matrix A, we write the general solution to $A\boldsymbol{x} = \mathbf{0}$ in parametric vector form. This writes the general solution as a linear combination of the basis vectors.
- To find a basis for the column space of a matrix A, we put the matrix in echelon form to identify the pivot columns. We then keep the pivot columns in the original matrix.

Practice Problems

Example

Let V and W be vector spaces.

Suppose $T : V \to W$ and $U : V \to W$ are linear transformations.

Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ be a basis for V.

Show that if $T(\boldsymbol{v}_j) = U(\boldsymbol{v}_j)$ for every $j = 1, \ldots, p$, then $T(\boldsymbol{x}) = U(\boldsymbol{x})$ for every vector $\boldsymbol{x}$ in V.

MyLab Problems

Find a basis for the null space of the matrix $\begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & -2 & 4 \\ 3 & -5 & 1 & -14 \end{bmatrix}$.

Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subset of V and T is a one-to-one linear transformation, so that an equation $T(\mathbf{u}) = T(\mathbf{v})$ always implies $\mathbf{u} = \mathbf{v}$. Show that if the set of images $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)\}$ is linearly dependent, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly dependent.

Theorem (Unique Representation)

Let $B = \{\boldsymbol{b}_1, \ldots, b_n\}$ be a basis for a vector space V. Then for each $\boldsymbol{x} \in V$, there exist unique scalars $c_1, \ldots, c_n$ such that

$$\boldsymbol{x} = c_1 \boldsymbol{b}_1 + \cdots + c_n \boldsymbol{b}_n.$$

Definition

We define the **coordinates of x relative to B** to be the vector

$$[\boldsymbol{x}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

The mapping

$$\boldsymbol{x} \mapsto [\boldsymbol{x}]_B$$

is called the **coordinate mapping**.

Examples ($\mathbb{R}^n$ Case)

- All of the old examples from $\mathbb{R}^n$ are relevant here:
  - Finding the coordinates of $\boldsymbol{x}$ with respect to a basis B is equivalent to solving $A\boldsymbol{c} = \boldsymbol{x}$, where the columns of A are given by the vectors in B.
  - If $[\boldsymbol{x}]_B = \boldsymbol{c}$, then $\boldsymbol{x} = A\boldsymbol{c}$, where the columns of A are given by the vectors in B.
- Given a basis $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ for $\mathbb{R}^n$, we set

$$P_B = [\boldsymbol{b}_1\ \boldsymbol{b}_2\ \cdots\ \boldsymbol{b}_n].$$

  We call this the **change-of-coordinates matrix** from B to the standard basis. We have

$$\boldsymbol{x} = P_B[\boldsymbol{x}]_B.$$

- The inverse of P_B is precisely the coordinate mapping:

$$P_B^{-1}\boldsymbol{x} = [\boldsymbol{x}]_B.$$

Example

Example

It can be shown that

$$B = \{1 + t, 1 + t^2, t + t^2\}$$

is a basis for $\mathbb{P}_2$. Find the coordinates of $\boldsymbol{p}(t) = 6 + 3t - t^2$ relative to B.

Solution: $[\boldsymbol{p}]_B = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}$.

The Coordinate Mapping

Theorem (Theorem 8)

Suppose B is a basis for a vector space V. Then the coordinate map $\boldsymbol{x} \mapsto [\boldsymbol{x}]_B$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.

- We say that the coordinate map is an **isomorphism** between V and $\mathbb{R}^n$ (i.e. one-to-one and onto).
- This tells us that any vector space with a basis consisting of n elements is essentially 'the same' as $\mathbb{R}^n$.

Useful Fact: If V is isomorphic to $\mathbb{R}^n$ and has a basis $\{\boldsymbol{b}_1, \ldots \boldsymbol{b}_n\}$, then a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ in V is independent if and only $\{[\boldsymbol{v}_1]_B, \ldots, [\boldsymbol{v}_p]_B\}$ is independent.

Example: Polynomials

Example

The basis $B = \{1, t, \ldots, t^n\}$ shows that $\mathbb{P}_n$ is isomorphic to $\mathbb{R}^{n+1}$. In particular, we naturally identify a polynomial

$$\boldsymbol{p}(t) = c_0 + c_1 t + \cdots + c_n t^n$$

with its coordinates

$$[\boldsymbol{p}]_B = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

However, we can use a different basis for $\mathbb{P}_n$; then the coordinates would change...

Theorem (Theorem 10)

If a basis V has a basis with n vectors, then every basis of V has exactly n vectors.

Proof.

Suppose B is a basis with n elements and C is a basis with m elements. Passing through the coordinate map, we can construct an isomorphism between $\mathbb{R}^n$ and $\mathbb{R}^m$. Thus $n = m$. $\square$

Definition

If V is spanned by a finite set, we call V **finite-dimensional**. Then (by the Spanning Set Theorem), V has a basis. We define $\dim V$ to be the number of elements in this (and any) basis.

The dimension of the vector space $\{0\}$ is zero by definition.

If V is not spanned by a finite set then V is **infinite dimensional**.

Examples

- The dimension of $\mathbb{R}^n$ is n.
- Subspaces of $\mathbb{R}^3$ have dimension $0, 1, 2$, or 3.
- The dimension of $\mathbb{P}_n$ is $n+1$.
- The dimension of $\mathbb{P}$ is infinite.
- The dimension of $\mathbb{S}$ (the sequence space) is infinite.
- The dimension of the kernel of

  $$T = \tfrac{d^2}{dx^2} + \omega^2$$

  is two.
- The dimension of the range of $\frac{d}{dx}$ is infinite.

Another example

Example

Find the dimension of the subspace

$$H = \left\{ \begin{bmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{bmatrix} : a, b, c, d \in \mathbb{R} \right\}.$$

We write this as the span of

$$\begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}.$$

Using this, we may deduce $\underline{\dim H = 3}$.

Subspaces and the Basis Theorem

Theorem (Theorem 11)

Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be extended (if necessary) to a basis of H. Furthermore, H is finite-dimensional and

$$\dim H \leq \dim V.$$

Theorem (Theorem 12 - The Basis Theorem)

Suppose V is a p-dimensional vector space with $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

Dimensions of Familiar Subspaces

For the null space and column space of a matrix A we have the following:

- The dimension of $\operatorname{Nul} A$ is the number of free variables in the equation $A\boldsymbol{x} = 0$.
- The dimension of $\operatorname{Col} A$ is the number of pivot columns in A.

We discussed this in Chapter 2. **You will work out numerical examples in the MyLab homework.**

A linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ can be expanded to a basis for $\mathbb{R}^n$. One way to do this is to create $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k & \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}$ with $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the columns of the identity matrix; the pivot columns of A form a basis for $\mathbb{R}^n$. Complete parts (a) and (b) below.

a. Use the method described to extend the following vectors to a basis for $\mathbb{R}^5$. Choose the correct answer below.

$$\mathbf{v}_1 = \begin{bmatrix} -6 \\ -7 \\ 6 \\ -5 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 5 \\ 5 \\ 2 \\ 6 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 8 \\ 7 \\ -6 \\ 5 \\ -7 \end{bmatrix}$$

- A. $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_1, \mathbf{e}_2\}$
- B. $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_2, \mathbf{e}_3\}$
- C. $\{\mathbf{e}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_4, \mathbf{e}_5\}$
- D. $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_1, \mathbf{e}_3\}$

b. Explain why the method works in general. Why are the original vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ included in the basis found for Col A?

The original vectors are the first k columns of A. Since the set of original vectors is assumed to be linearly independent, these columns of A will be pivot columns and the original set of vectors will be included in the basis.

Why is Col $A = \mathbb{R}^n$?

Since all of the columns of the $n \times n$ identity matrix are columns of A, every vector in $\mathbb{R}^n$ is in Col A. Since every column of A is in $\mathbb{R}^n$, every vector in Col A is in $\mathbb{R}^n$. This shows that Col A and $\mathbb{R}^n$ are equivalent.

MyLab Problems

The first four Hermite polynomials are 1, $2t$, $-2+4t^2$, and $-12t+8t^3$. These polynomials arise naturally in the study of certain important differential equations in mathematical physics. Show that the first four Hermite polynomials form a basis of $\mathbb{P}_3$.

To show that the first four Hermite polynomials form a basis of $\mathbb{P}_3$, what theorem should be used?

- **A.** If a vector space V has a basis $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.
- **B.** Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H.
- **C.** If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.
- **D.** Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V.

Section 4.6 - Rank

- For an $m \times n$ matrix A, we define the **column space** $\operatorname{Col} A$ to be the span of the columns of A. It is a subspace of $\mathbb{R}^m$.
- We define the **row space** Row A to be the span of the rows of A. It is a subspace of $\mathbb{R}^n$.
- The **null space** $\operatorname{Nul} A$ is the set of solutions to $A\boldsymbol{x} = \mathbf{0}$. It is a subspace of $\mathbb{R}^n$.

Definition (Rank)

The **rank** of A is the dimension of the column space of A.

Theorem (Theorem 14 - The Rank Theorem)

Let A be $m \times n$. We have

$$rank\, A = \dim Col\, A = \dim Row\, A = \#\ \textit{of pivot positions in } A.$$

Furthermore,

$$rank\, A + \dim Nul\, A = n.$$

Row Space

- The only new part in the Rank Theorem is the part about the row space. We need the following:

Theorem (Theorem 13)

If A and B are row equivalent, then $\text{Row}\, A = \text{Row}\ B$.

If B is in echelon form, then the nonzero rows of B form a basis for $\text{Row}\, A$.

- *Key Observation:* if B is obtained from A by row operations, then the rows of B are linear combinations of the rows of A.
- With this theorem in place, we can see that the column and row spaces have the same dimension.

Example

Find bases for the row space, column space, and null space of

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix}.$$

Solution. Reduce A to an echelon form:

$$A \sim \begin{bmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & -4 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Example (Continued)

Example

- Basis for Row A:

$$\{(1,3,-5,1,5),(0,1,-2,2,-7),(0,0,0,-4,20)\}.$$

- Basis for Col A:

$$\left\{\begin{bmatrix} -2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -5 \\ 3 \\ 11 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 7 \\ 5 \end{bmatrix}\right\}.$$

- Basis for Nul A: We should put the matrix in reduced echelon form.

Example (Continued)

Example

$$A \sim B \sim \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus a basis for $\operatorname{Nul} A$ is

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 0 \\ 5 \\ 1 \end{bmatrix} \right\}.$$

More Examples; MyLab Problem

- If A is 7×9 and $\dim \operatorname{Nul} A = 2$, what is the rank of A?
- Can a 6×9 matrix have a two-dimensional null space?

Is it possible that all solutions of a homogeneous system of thirteen linear equations in seventeen variables are multiples of one fixed nonzero solution? Discuss.

Consider the system as $A\mathbf{x} = \mathbf{0}$, where A is a 13×17 matrix. Choose the correct answer below.

- **A.** Yes. Since A has at most 13 pivot positions, rank $A \le 13$. By the Rank Theorem, dim Nul $A = 17 - \text{rank } A \ge 4$. Since there is at least one free variable in the system, all solutions are multiples of one fixed nonzero solution.
- **B.** Yes. Since A has 13 pivot positions, rank $A = 13$. By the Rank Theorem, dim Nul $A = 13 - \text{rank } A = 0$. Thus, all solutions are multiples of one fixed nonzero solution.
- **C.** No. Since A has at most 13 pivot positions, rank $A \le 13$. By the Rank Theorem, dim Nul $A = 17 - \text{rank } A \ge 4$. Thus, it is impossible to find a single vector in Nul A that spans Nul A.
- **D.** No. Since A has 13 pivot positions, rank $A = 13$. By the Rank Theorem, dim Nul $A = 13 - \text{rank } A = 0$. Since Nul $A = 0$, it is impossible to find a single vector in Nul A that spans Nul A.

MyLab Problem

Let A be an $m \times n$ matrix. Explain why the equation $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbb{R}^m$ if and only if the equation $A^T\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Choose the correct answer below.

- **A.** The system $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$, or $\dim \operatorname{Col} A = m$. The equation $A^T\mathbf{x} = \mathbf{0}$ has only the trivial solution if and only if $\dim \operatorname{Nul} A = 0$. By the Rank Theorem, $\dim \operatorname{Col} A = \operatorname{rank} A = m - \dim \operatorname{Nul} A$. Thus, $\dim \operatorname{Col} A = m$ if and only if $\dim \operatorname{Nul} A = 0$.
- **B.** The system $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$, or $\dim \operatorname{Row} A = m$. The equation $A^T\mathbf{x} = \mathbf{0}$ has only the trivial solution if and only if $\dim \operatorname{Nul} A^T = 0$. Since $\operatorname{Row} A = \operatorname{Col} A^T$, $\dim \operatorname{Row} A = \dim \operatorname{Col} A^T = m - \dim \operatorname{Nul} A^T$ by the Rank Theorem. Thus, $\dim \operatorname{Row} A = m$ if and only if $\dim \operatorname{Nul} A^T = 0$.
- **C.** The system $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$, or $\dim \operatorname{Col} A = m$. The equation $A^T\mathbf{x} = \mathbf{0}$ has only the trivial solution if and only if $\dim \operatorname{Nul} A^T = 0$. Since $\operatorname{Col} A = \operatorname{Row} A^T$, $\dim \operatorname{Col} A = \dim \operatorname{Row} A^T = \operatorname{rank} A^T = m - \dim \operatorname{Nul} A^T$ by the Rank Theorem. Thus, $\dim \operatorname{Col} A = m$ if and only if $\dim \operatorname{Nul} A^T = 0$.

Section 4.7 - Change of Basis

Theorem (Theorem 15)

Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases for a vector space V. There exists a unique $n \times n$ matrix $P_{C \leftarrow B}$ (the **change-of-coordinates matrix***) such that*

$$[\mathbf{x}]_C = P_{C \leftarrow B}[\mathbf{x}]_B$$

for every $\mathbf{x}$ in V.

The columns of $P_{C \leftarrow B}$ are given by

$$P_{C \leftarrow B} = \begin{bmatrix} [\mathbf{b}_1]_C & [\mathbf{b}_2]_C & \cdots & [\mathbf{b}_n]_C \end{bmatrix}.$$

Proof.

Writing $\mathbf{e}_k$ for the standard basis vectors, we have $\mathbf{e}_k = [\mathbf{b}_k]_B$. □

Visualizing $P_{C \leftarrow B}$

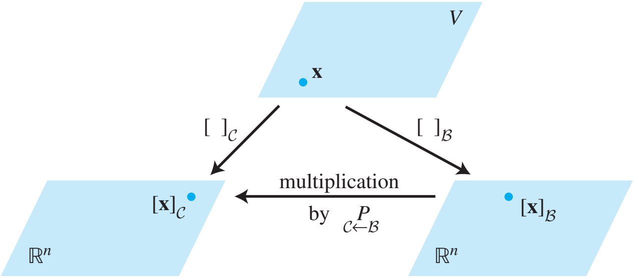


FIGURE 2 Two coordinate systems for V.

Example

Example

Let

$$\boldsymbol{b}_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}, \quad \boldsymbol{b}_2 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \quad \boldsymbol{c}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \quad \boldsymbol{c}_2 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}.$$

To compute $P_{C\leftarrow B}$ we solve

$$[\boldsymbol{c}_1 \ \boldsymbol{c}_2 \ | \boldsymbol{b}_1 \ \boldsymbol{b}_2 \ ] \sim \begin{bmatrix} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{bmatrix},$$

yielding

$$P_{C\leftarrow B} = \begin{bmatrix} 6 & 4 \\ -5 & -3 \end{bmatrix}.$$

Remarks

- The previous example generalizes: in the case of $V = \mathbb{R}^n$, we may compute $P_{C \leftarrow B}$ by

$$[\ \boldsymbol{c}_1\ \boldsymbol{c}_2\ \cdots\ \boldsymbol{c}_n \mid \boldsymbol{b}_1\ \boldsymbol{b}_2\ \cdots\ \boldsymbol{b}_n] \sim [I_n \mid P_{C \leftarrow B}].$$

- $(P_{C \leftarrow B})^{-1} = P_{B \leftarrow C}$
- If $V = \mathbb{R}^n$ and E denotes the standard basis, then $P_{E \leftarrow B}$ is the same as the change of coordinates matrix P_B from Section 4.4.
- Using the previous observation, we deduce

$$P_{C \leftarrow B} = P_C^{-1} P_B.$$

Practice Problems

Example (MyLab Problem)

If $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ and $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2\}$ are bases for V and

$$\boldsymbol{b}_1 = -4\boldsymbol{c}_1 + 2\boldsymbol{c}_2 \quad \text{and} \quad \boldsymbol{b}_2 = 8\boldsymbol{c}_1 - 6\boldsymbol{c}_2,$$

then

$$P_{C \leftarrow B} = \begin{bmatrix} -4 & 8 \\ 2 & -6 \end{bmatrix}.$$

If $\boldsymbol{x} = 7b_1 - 6\boldsymbol{b}_2$, then to find $[\boldsymbol{x}]_C$ we apply the matrix above to $\begin{bmatrix} 7 \\ -6 \end{bmatrix}$.

MyLab Problem

In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $B = \{1 - 2t + t^2, 4 - 7t + 5t^2, 2 - 2t + 5t^2\}$ to the standard basis $C = \{1, t, t^2\}$. Then find the B-coordinate vector for $-4 + 7t - 4t^2$.

In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $B = \{1 - 2t + t^2, 4 - 7t + 5t^2, 2 - 2t + 5t^2\}$ to the standard basis $C = \{1, t, t^2\}$.

$$\underset{C \leftarrow B}{P} = \begin{bmatrix} 1 & 4 & 2 \\ -2 & -7 & -2 \\ 1 & 5 & 5 \end{bmatrix}$$ (Simplify your answers.)

Find the B-coordinate vector for $-4 + 7t - 4t^2$.

$$[\mathbf{x}]_B = \begin{bmatrix} 6 \\ -3 \\ 1 \end{bmatrix}$$ (Simplify your answers.)

- A vector with nonnegative entries that add up to 1 is called a **probability vector**.
- A **stochastic matrix** is a square matrix whose columns are probability vectors.
- A **Markov chain** is a sequence of probability vectors $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2$, together with a stochastic matrix P such that

$$\boldsymbol{x}_{k+1} = P\boldsymbol{x}_k \quad \text{for} \quad k = 0, 1, 2, \ldots$$

- We call each $\boldsymbol{x}_k$ a **state vector**.

Example

EXAMPLE 1 Section 1.10 examined a model for population movement between a city and its suburbs. See Figure 1. The annual migration between these two parts of the metropolitan region was governed by the *migration matrix* M:

$$\begin{array}{cccc} & \multicolumn{2}{c}{\text{From:}} & \\ & \text{City} & \text{Suburbs} & \text{To:} \\ M = & \left[\begin{array}{cc} .95 & .03 \\ .05 & .97 \end{array}\right] & & \begin{array}{l} \text{City} \\ \text{Suburbs} \end{array} \end{array}$$

That is, each year 5% of the city population moves to the suburbs, and 3% of the suburban population moves to the city. The columns of M are probability vectors, so M is a stochastic matrix. Suppose the 2014 population of the region is 600,000 in the city and 400,000 in the suburbs. Then the initial distribution of the population in the region is given by $\mathbf{x}_0$ in (1) above. What is the distribution of the population in 2015? In 2016?

Steady-State Vectors

A **steady-state vector** for a stochastic matrix P is a probability vector $\boldsymbol{q}$ so that

$$P\boldsymbol{q} = \boldsymbol{q}.$$

Theorem (Theorem 18)

If P is a 'regular' stochastic matrix, then P has a unique steady-state vector $\boldsymbol{q}$. Furthermore, if $\boldsymbol{x}_0$ is any initial state and $\boldsymbol{x}_{k+1} := P\boldsymbol{x}_k$ for $k \geq 0$, then the Markov chain $\boldsymbol{x}_k$ converges to $\boldsymbol{q}$ as $k \to \infty$.

- To find a steady state vector, we should find a basis for the null space of $P - I$, which is evidently one-dimensional. Then 'normalize' to produce a probability vector.

Math 3108 - Fall 2019
Chapter 5: Eigenvalues and Eigenvectors

Definition

An **eigenvalue** of an $n \times n$ matrix A is a scalar λ such that the equation

$$A\mathbf{x} = \lambda\mathbf{x} \quad \text{has a nontrivial solution.}$$

A nonzero solution to $A\mathbf{x} = \lambda\mathbf{x}$ is called an **eigenvector** of A (corresponding to λ).

- **Warning!** Although we primarily consider matrices with real-valued entries, the eigenvalues of A may be **complex-valued**, and the entries of the eigenvectors may also be **complex-valued**!
- By definition, eigenvectors must be **nonzero**. Why is this reasonable?

Examples

Example

Let

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$

Are $\boldsymbol{u}$ and $\boldsymbol{v}$ eigenvectors of A?

Solution. Compute

$$A\boldsymbol{u} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\boldsymbol{u},$$

so $\boldsymbol{u}$ is an eigenvector, but

$$A\boldsymbol{v} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\boldsymbol{v} \quad \text{for any} \quad \lambda.$$

Example

Show that 7 is an eigenvalue of

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

Solution. We need to find a nontrivial solution to

$$A\mathbf{x} = 7\mathbf{x}, \quad \text{i.e.} \quad (A - 7I)\mathbf{x} = \mathbf{0}.$$

Since

$$A - 7I = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

we find that 7 is an eigenvalue. Any multiple of $[1\ 1]^T$ is an eigenvector.

Eigenspaces

- If λ is an eigenvalue of A, then the **eigenspace** E_λ of A is defined to be the null space of $A - \lambda I$.
- In particular, E_λ consists of all eigenvectors of A corresponding to eigenvalue λ, together with the zero vector.
- In the preceding example, we saw that E_7 is the line through the origin in $\mathbb{R}^2$ spanned by $[1\ 1]^T$.

Example

Let

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

Given that $\lambda = 2$ is an eigenvalue, find a basis for the eigenspace E_2.

Solution. Note that

$$A - 2I = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus a basis is given by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Special Cases

Theorem (Theorem 1)

The eigenvalues of a triangular matrix are given by its diagonal entries.

Proof.

If λ equals one of the diagonal entries, then $A - \lambda I$ will not have a pivot in every column. $\square$

- A matrix A has eigenvalue $\lambda = 0$ if and only if A is **not** invertible.
- Indeed, both are equivalent to the fact that $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.

Independence of Eigenvectors

Theorem (Theorem 2)

Suppose $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors corresponding to **distinct** *eigenvalues $\lambda_1, \dots, \lambda_r$ of a matrix A. Then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent.*

Proof.

Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is independent. Now suppose

$$c_1\mathbf{v}_1 + \cdots + c_{k+1}\mathbf{v}_{k+1} = \mathbf{0}. \quad (1)$$

Apply A to get

$$\lambda_1 c_1 \mathbf{v}_1 + \cdots + \lambda_{k+1} c_{k+1} \mathbf{v}_{k+1} = \mathbf{0}.(2)$$

Multiply (1) by λ_{k+1} and subtract from (2) to get

$$(\lambda_1 - \lambda_{k+1})c_1\mathbf{v}_1 + \cdots + (\lambda_k - \lambda_{k+1})c_k\mathbf{v}_k = 0.$$

Thus... □

Practice Problems

- If $\boldsymbol{x}$ is an eigenvector of A corresponding to eigenvalue λ, what is $A^3\boldsymbol{x}$?
- If λ is an eigenvector of an invertible matrix A, then λ^{-1} is an eigenvalue of A^{-1}.
- Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T.

MyLab Problems

For $A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -2 & 4 \\ 1 & -2 & 4 \end{bmatrix}$, find one eigenvalue, with no calculation. Justify your answer.

Choose the correct answer below.

- **A.** One eigenvalue of A is $\lambda = -2$. This is because each column of A is equal to the product of -2 and the column to the left of it.
- **B.** One eigenvalue of A is $\lambda = 1$. This is because 1 is one of the entries on the main diagonal of A, which are the eigenvalues of A.
- **C.** One eigenvalue of A is $\lambda = 1$. This is because each row of A is equal to the product of 1 and the row above it.
- **D.** One eigenvalue of A is $\lambda = 0$. This is because the columns of A are linearly dependent, so the matrix is not invertible.

A is an $n \times n$ matrix. Mark each statement below True or False. Justify each answer.

a. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A. Choose the correct answer below.

- **A.** True. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A because the only solution to this equation is the trivial solution.
- **B.** True. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A because λ is an inverse of A.
- **C.** False. The condition that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ is not sufficient to determine if $\mathbf{x}$ is an eigenvector of A. The vector $\mathbf{x}$ must be nonzero.
- **D.** False. The equation $A\mathbf{x} = \lambda\mathbf{x}$ is not used to determine eigenvectors. If $\lambda A\mathbf{x} = \mathbf{0}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A.

- We need a systematic way of determining the eigenvalues λ of a matrix A. (Once we have done so, we can find eigenvectors by solving the homogeneous equation $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$.)
- Finding λ such that $A-\lambda I$ is *not* invertible is equivalent to finding λ such that

$$\det[A-\lambda I]=0. \quad (*)$$

The equation $(*)$ is called the **characteristic equation**.

Example

Example

Find the characteristic equation of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solution. As the matrix is triangular we deduce

$$\det[A - \lambda I] = (5-\lambda)^2(3-\lambda)(1-\lambda).$$

In particular, the eigenvalues are $\lambda = 5, 3, 1$. We say that $\lambda = 5$ has **multiplicity 2**.

The Characteristic Polynomial

- Given an $n \times n$ matrix A, we may define $p : \mathbb{C} \to \mathbb{C}$ by

$$p(\lambda) = \det[A - \lambda I].$$

  Then the characteristic equation becomes $p(\lambda) = 0$.
- In fact, it turns out that $p(\lambda)$ is a degree n polynomial in λ, called the **characteristic polynomial** of A.

Example

Example

If the characteristic polynomial of a 6×6 matrix is $\lambda^6 - 4\lambda^5 - 12\lambda^4$, find the eigenvalues and multiplicities.

Solution. Factor the polynomial as

$$\lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2).$$

The eigenvalues are $\lambda = 0$ (with multiplicity 4), $\lambda = 6$, and $\lambda = -2$.

Similarity of Matrices

Definition

Two $n \times n$ matrices A and B are **similar** if there exists an invertible matrix P such that $A = PBP^{-1}$.

- Similarity of matrices is an equivalence relation.
- Similarity is not related to row equivalence.

Theorem (Theorem 4)

If A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (including multiplicities).

- Matrices can have the same eigenvalues without being similar.

Dynamical Systems

A **dynamical system** is given by an initial state vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ and an $n \times n$ matrix A through the recursion relation

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k.$$

For example, Markov chains are examples of dynamical systems; steady state vectors are eigenvectors with eigenvalue $\lambda = 1$.

The eigenvalues/eigenvectors of A may allow us to determine the 'long-time behavior' of the dynamical system.

For example, if $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ were a **basis** of eigenvectors for A with eigenvalues $\lambda_1, \ldots, \lambda_n$ and

$$\boldsymbol{x}_0 = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n,$$

then we would have

$$\boldsymbol{x}_k = c_1\lambda_1^k\boldsymbol{v}_1 + \cdots + c_n\lambda_n^k\boldsymbol{v}_n.$$

- In applications, it is desirable to construct a **basis** of eigenvectors for a given matrix A.
- Finding a basis of eigenvectors is equivalent to **diagonalizing** the matrix A.

Definition

A square matrix A is **diagonalizable** if it is similar to a diagonal matrix, that is, if

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

for some invertible matrix P and diagonal matrix D.

The Diagonalization Theorem

Theorem (Theorem 5 - The Diagonalization Theorem)

- *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.*
- *In fact, $A = PDP^{-1}$ if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the entries of D are the corresponding eigenvalues.*

- *Proof Sketch:* Observe that $A = PDP^{-1}$ is equivalent to $AP = PD$, which in turn is equivalent to

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i,$$

where $\boldsymbol{v}_i$ is the i^{th} column of P and λ_i is the i^{th} entry of D along the diagonal.

Example 1

Example

If possible, diagonalize the following matrix:

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

The steps are as follows:

- Find the eigenvalues of A.
- Find three linearly independent eigenvectors of A.
- Construct P and D so that $A = PDP^{-1}$.

Example 1 (Continued)

Example (Continued)

To find the eigenvalues of A, we solve the characteristic equation:

$$0 = \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2.$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$ (multiplicity 2).

Example 1 (Continued)

Example (Continued)

We next find a basis for each eigenspace E_λ, i.e. the null space of $A - \lambda I$.

- Basis for $\lambda = 1$ is given by $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.
- Basis for $\lambda = -2$ is given by $\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Example 1 (Continued)

Example (Continued)

Now we form the matrices P and D:

$$P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

We can then verify that $A = PDP^{-1}$.

Example 2

Example

Diagonalize the following matrix, if possible:

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

- The characteristic equation is the same as in Example 1, and so the eigenvalues are

$$\lambda = 1 \quad \text{and} \quad \lambda = -2 \quad \text{(with multiplicity 2).}$$

Example 2 (Continued)

Example (Continued)

We next find bases for the eigenspaces E_λ:

- Basis for $\lambda = 1$ is given by $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.
- Basis for $\lambda = -2$ is given by $\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$.

Conclusion: The matrix A is **not** diagonalizable.

Diagonalizability

Theorem (Theorem 6)

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Proof.

Eigenvectors corresponding to distinct eigenvalues are independent. $\square$

- This gives a sufficient condition for diagonalizability, although it is not necessary (cf. Example 1 above).

Example

Example

Determine whether or not the following matrix is diagonalizable:

$$A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}.$$

Solution. The matrix is triangular and has eigenvalues $\lambda = 5, 0, -2$. Thus A is diagonalizable.

Powers of a Diagonalizable Matrix

- Diagonalizing a matrix A is useful if you need to compute powers of A, since
$$A = PDP^{-1} \implies A^k = PD^kP^{-1},$$
and computing powers of a diagonal matrix is straightforward, cf.
$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}^k = \begin{bmatrix} a^k & 0 & 0 \\ 0 & b^k & 0 \\ 0 & 0 & c^k \end{bmatrix}.$$
- *Application:* Computing matrix exponentials to solve linear systems of differential equations.

Repeated Eigenvalues

THEOREM 7 Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p$.

a. For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (*i*) the characteristic polynomial factors completely into linear factors and (*ii*) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \ldots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

Example

Example

Diagonalize the following matrix if possible:

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

Solution: The matrix is triangular and has eigenvalues $\lambda = 5, -3$, each with multiplicity 2.

We look for bases for each eigenspace.

Example (Continued)

- Basis for $\lambda = 5$: $\boldsymbol{v}_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$.
- Basis for $\lambda = -3$: $\boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

The matrix is diagonalizable, with

$$P = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \boldsymbol{v}_3 \ \boldsymbol{v}_4] \quad \text{and} \quad D = \text{diag}\{5, 5, -3, -3\}.$$

Practice Problems

- Suppose A is 4×4 and has eigenvalues $5, 3, -2$. Suppose E_3 is two-dimensional. Is A diagonalizable?
- How would you compute A^8 if $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$?

MyLab Problems

Matrix A is factored in the form PDP^{-1}. Use the Diagonalization Theorem to find the eigenvalues of A and a basis for each eigenspace.

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

Let A, P, and D be $n \times n$ matrices. Mark each statement true or false. Justify each answer. Complete parts (a) through (d) below.

a. A is diagonalizable if $A = PDP^{-1}$ for some matrix D and some invertible matrix P. Choose the correct answer below.

b. If $\mathbb{R}^n$ has a basis of eigenvectors of A, then A is diagonalizable. Choose the correct answer below.

c. A is diagonalizable if and only if A has n eigenvalues, counting multiplicities. Choose the correct answer below.

d. If A is diagonalizable, then A is invertible. Choose the correct answer below.

MyLab Problems

Identify a nonzero 2×2 matrix that is invertible but not diagonalizable.

Choose the correct answer below.

- A. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$
- B. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
- C. $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$
- D. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
- E. $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$
- F. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- Recall that any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ may be represented by an $m \times n$ matrix A (the *standard matrix* of T), i.e.

$$T(\boldsymbol{x}) = A\boldsymbol{x} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^n.$$

- More generally, suppose $T : V \to W$ is a linear transformation with $\dim V = n$ and $\dim W = m$. Let B, C be bases for V, W, respectively. Now define the $m \times n$ matrix M by

$$M = \big[\, [T(\boldsymbol{b}_1)]_C \ \cdots \ [T(\boldsymbol{b}_n)]_C \,\big].$$

It follows that

$$[T(\boldsymbol{x})]_C = M[\boldsymbol{x}]_B \quad \text{for all} \quad \boldsymbol{x} \in V.$$

We call M the **matrix for T relative to the bases B and C**.

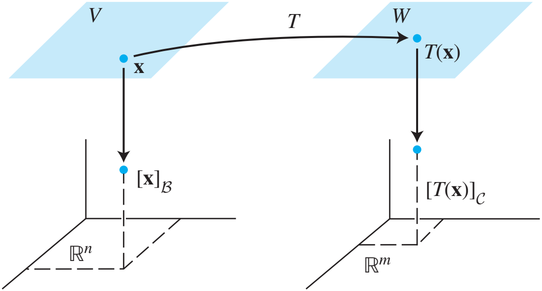


FIGURE 1 A linear transformation from V to W.

Example

Example

If $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ and $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$ are bases for V, W, and $T : V \to W$ is a linear transformation such that

$$T(\boldsymbol{b}_1) = 3\boldsymbol{c}_1 - 2\boldsymbol{c}_2 + 5\boldsymbol{c}_3 \quad \text{and} \quad T(\boldsymbol{b}_2) = 4\boldsymbol{c}_1 + 7\boldsymbol{c}_2 - \boldsymbol{c}_3,$$

then

$$M = \begin{bmatrix} 3 & 4 \\ -2 & 7 \\ 5 & -1 \end{bmatrix}.$$

Linear Transformations on V

- Often, we take $V = W$ and $C = B$, in which case the matrix M is called the **matrix for T relative to B**, or the **B-matrix for T**, denoted by $[T]_B$. In particular,

$$[T(\boldsymbol{x})]_B = [T]_B[\boldsymbol{x}]_B \quad \text{for all} \quad \boldsymbol{x} \in V.$$

Example

Let $T : \mathbb{P}_2 \to \mathbb{P}_2$ be given by $T(\boldsymbol{p}) = \boldsymbol{p}'$.

(i) Find the B matrix for T, where $B = \{1, t, t^2\}$. (ii) Check that $[T(\boldsymbol{p})]_B = [T]_B[\boldsymbol{p}]_B$.

Solution. (i) Since

$$T(1) = 0, \quad T(t) = 1, \quad T(t^2) = 2t,$$

we get

$$[T]_B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Example (Continued)

Example (Continued)

(ii) Note that

$$T(a_0 + a_1 t + a_2 t^2) = a_1 + 2a_2 t,$$

so

$$[T(\boldsymbol{p})]_B = [a_1 + 2a_2 t]_B = \begin{bmatrix} a_1 \\ 2a_2 \\ 0 \end{bmatrix},$$

while

$$[T]_B[\boldsymbol{p}]_B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ 2a_2 \\ 0 \end{bmatrix}.$$

Theorem (Theorem 8 - Diagonal Matrix Representation)

Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If B is the basis of $\mathbb{R}^n$ formed from the columns of P, then D is the B-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Proof.

The essential facts are

$$P[\mathbf{x}]_B = \mathbf{x} \quad \text{and} \quad [\mathbf{x}]_B = P^{-1}\mathbf{x}.$$

$\square$

Example

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(\boldsymbol{x}) = A\boldsymbol{x}$, with

$$A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}.$$

Find a basis B for $\mathbb{R}^2$ such that the B-matrix for T is diagonal.

Solution. Diagonalize A as $A = PDP^{-1}$, where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

Let $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ be the basis consisting of the columns of P.
Then D is the B-matrix of T.

Similarity of Matrix Representations

• More generally, if A and C are similar $n \times n$ matrices, then they represent the same linear transformation.

Indeed: if $T(\boldsymbol{x}) = A\boldsymbol{x}$ and $A = PCP^{-1}$, then $C = [T]_B$, where B is the basis consisting of the columns of P.

• In fact, if B is any basis for $\mathbb{R}^n$, then $[T]_B$ is similar to A.

To see this, define P to have columns given by the vectors in B. Then

$$A = P[T]_B P^{-1}.$$

• As before, the essential facts are $P[\boldsymbol{x}]_B = \boldsymbol{x}$ and $[\boldsymbol{x}]_B = P^{-1}\boldsymbol{x}$.

Define $T:\mathbb{P}_2 \to \mathbb{R}^3$ as shown to the right.

$$T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(-3) \\ \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix}$$

a. Find the image under T of $\mathbf{p}(t) = 1 - 2t$.

b. Show that T is a linear transformation.

c. Find the matrix for T relative to the basis $B = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\} = \{1, t, t^2\}$ for $\mathbb{P}_2$ and the standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for $\mathbb{R}^3$.

Find the *B*-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$, where $B = \{\mathbf{b}_1, \mathbf{b}_2\}$.

$$A = \begin{bmatrix} -5 & -1 \\ 4 & 1 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

The *B*-matrix of the given transformation is $\begin{bmatrix} -2 & -5 \\ -1 & -2 \end{bmatrix}$.

- The characteristic polynomial of a (real-valued) $n \times n$ matrix A is a degree n polynomial, and hence it has n roots ("fundamental theorem of algebra").
- Roots may be repeated (as we have seen), but they may also be **complex**.
- A **complex number** has the form

$$z = x + iy$$

where x, y are real numbers and i satisfies $i^2 = -1$. We write $z \in \mathbb{C}$.
- The **magnitude** of z is $|z| = \sqrt{x^2 + y^2}$.
- A **complex vector** is a vector with *complex* entries. We write $\boldsymbol{x} \in \mathbb{C}^n$.

Complex Eigenvalues and Eigenvectors

- A **complex eigenvalue/eigenvector pair** for a matrix A is a complex number λ and a non-zero complex vector $\boldsymbol{x}$ satisfying

$$A\boldsymbol{x} = \lambda\boldsymbol{x}.$$

- The method for finding complex eigenvalues/eigenvectors is the same as the real case; however, now we have to work with complex numbers.
- Real matrices may have complex eigenvalues/eigenvectors.

Example

Let

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

This is counterclockwise rotation by 90°. There are no real eigenvalues/eigenvectors.

The characteristic equation is $\lambda^2 + 1 = 0$, which has roots $\lambda = \pm i$.

Eigenvectors corresponding to $\lambda = \pm i$ are given by

$$\begin{bmatrix} 1 \\ -i \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Example

Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}.$$

Solution. The characteristic equation is

$$0 = \det \begin{bmatrix} .5 - \lambda & -.6 \\ .75 & 1.1 - \lambda \end{bmatrix} = \lambda^2 - 1.6\lambda + 1.$$

By the quadratic formula, the eigenvalues are

$$\lambda = .8 \pm .6i.$$

Example (Continued)

Example (Continued)

Let $\lambda = .8 - .6i$. We look for the eigenspace E_λ:

$$A - (.8 - .6i)I = \begin{bmatrix} -.3 + .6i & -.6 \\ .75 & .3 + .6i \end{bmatrix}.$$

We use either row. We need to solve

$$.75x_1 + (.3 + .6i)x_2 = 0,$$

which we may solve with

$$\boldsymbol{x} = \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}.$$

Example (Continued)

Example (Continued)

Similarly, we can find an eigenvector for $\lambda = .8 + .6i$ is given by

$$\boldsymbol{x}_2 = \begin{bmatrix} -2+4i \\ 5 \end{bmatrix}.$$

Real and Imaginary Parts of Vectors

- If $z = x + iy$ is a complex number, then we write $x = \text{ Re } z$ and $y = \text{ Im } z$ for the **real** and **imaginary parts** of z.
- Similarly, a complex vector can be written as $\boldsymbol{v} = [\text{Re } \boldsymbol{v}] + i[\text{Im } \boldsymbol{v}]$, where we take the real and imaginary part of each entry.
- The **complex conjugate** of $z = x + iy$ is given by $\bar{z} = x - iy$.
- Similarly, the complex conjugate of a complex vector $\boldsymbol{v}$ is given by

$$\bar{\boldsymbol{v}} = \text{ Re}\,\boldsymbol{v} - i\,\text{Im}\,\boldsymbol{v}.$$

- Example:

$$\begin{bmatrix} 3-i \\ i \\ 2+5i \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + i \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}.$$

Some Algebraic Properties

- Let r be a scalar, $\boldsymbol{x}$ a vector, and B, C matrices. Then

$$\overline{r\boldsymbol{x}} = \bar{r}\bar{\boldsymbol{x}}, \quad \overline{B\boldsymbol{x}} = \bar{B}\bar{\boldsymbol{x}}, \quad \overline{BC} = \bar{B}\,\bar{C}, \quad \overline{rB} = \bar{r}\bar{B}.$$

- Suppose A is a *real* matrix. Then

$$\overline{A\boldsymbol{x}} = \bar{A}\bar{\boldsymbol{x}} = A\bar{\boldsymbol{x}}.$$

  If λ is an eigenvalue with eigenvector $\boldsymbol{x} \in \mathbb{C}^n$, then

$$A\boldsymbol{x} = \overline{A\boldsymbol{x}} = \overline{\lambda\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}}.$$

  Thus $\bar{\lambda}$ is an eigenvalue, with $\bar{\boldsymbol{x}}$ an eigenvector.

- *Conclusion: When A is real, complex eigenvalues and eigenvectors occur in conjugate pairs.*

Example

Let

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad a, b \in \mathbb{R}, \quad a, b \neq 0.$$

The eigenvalues of C are $\lambda = a \pm bi$.

Define $r = |\lambda| = \sqrt{a^2 + b^2}$. Then, writing $\frac{a}{r} = \cos\varphi$, we can factor

$$C = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}.$$

Then C consists of a rotation by φ and a scaling by $|\lambda|$.

General Case - Complex Eigenvalues

Theorem (Theorem 9)

Let A be a real 2×2 matrix with complex eigenvalue $\lambda = a - bi$ (with $b \neq 0$) and associated eigenvector $\boldsymbol{v} \in \mathbb{C}^2$. Then

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

where

$$P = [\, Re(\boldsymbol{v}) \;\; Im(\boldsymbol{v}) \,] \quad \text{and} \quad C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

- In higher dimensions, a complex conjugate eigenvalue pair for A corresponds to a plane on which A acts as a rotation combined with a scaling.

Example

Example

Return to the matrix

$$A = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}, \quad \text{with} \quad \lambda = -8 - 6i \quad \text{and} \quad \boldsymbol{v}_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix}.$$

Set

$$P = \begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} .8 & -.6 \\ .6 & .8 \end{bmatrix}.$$

Then $A = PCP^{-1}$. Note that C is a pure rotation.

List the eigenvalues of A. The transformation $\mathbf{x} \mapsto A\mathbf{x}$ is the composition of a rotation and a scaling. Give the angle φ of the rotation, where $-\pi < \varphi \le \pi$, and give the scale factor r.

$$A = \begin{bmatrix} -6 & 6 \\ -6 & -6 \end{bmatrix}$$

The eigenvalues of A are $\lambda = \boxed{-6-6i, -6+6i}$.

(Use a comma to separate answers as needed. Type an exact answer, using radicals and i as needed.)

$$\varphi = \boxed{-\frac{3\pi}{4}}$$

(Type an exact answer, using π as needed.)

$$r = \boxed{6\sqrt{2}}$$

(Type an exact answer, using radicals as needed.)

- *Remark.* My presentation deviates from the book significantly.
- A system of linear differential equations takes the form

$$\boldsymbol{x}'(t) = A\boldsymbol{x}(t),$$

where $\boldsymbol{x}(t)$ is a function of t taking values in $\mathbb{R}^n$, A is an $n \times n$ matrix, and $\boldsymbol{x}'(t)$ is the component-wise derivative of $\boldsymbol{x}(t)$.

Example

A second order ODE of the form

$$y'' + by' + cy = 0$$

may be rewritten as

$$\boldsymbol{x}' = A\boldsymbol{x}, \quad \boldsymbol{x} = \begin{bmatrix} y \\ y' \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ -c & -b \end{bmatrix}.$$

The Matrix Exponential

- When $n = 1$, solutions to $x' = Ax$ are of the form $x(t) = e^{At}c$.
- The same will be true for $n > 1$.

Definition (Matrix Exponential)

For an $n \times n$ matrix A, we define

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

Theorem

Solutions to $\mathbf{x}' = A\mathbf{x}$ are of the form $\mathbf{x}(t) = e^{At}\mathbf{c}$, where $\mathbf{c}$ is a fixed vector in $\mathbb{R}^n$.

(In fact, $\mathbf{c} = \mathbf{x}(0)$, called the initial condition.)

Examples

Example

If $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then

$$e^A = \mathrm{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}).$$

Example

If $A = PDP^{-1}$, then

$$e^A = Pe^D P^{-1}.$$

Combining these two examples, we find that if A is diagonalizable, then we can compute its matrix exponential.

Examples

Example

We have $e^0 = I$.

More generally, if A is 'nilpotent' (meaning $A^p = 0$ for some p), then

$$e^A = \sum_{k=0}^{p-1} \frac{1}{k!} A^k.$$

Example

If $AB = BA$, then

$$e^{A+B} = e^A e^B = e^B e^A.$$

In particular, e^A is always invertible, with

$$(e^A)^{-1} = e^{-A}.$$

Numerical Example

Example

Consider

$$y'' - 4y' + 3y = 0 \implies \boldsymbol{x}' = A\boldsymbol{x}, \quad A = \begin{bmatrix} 0 & 1 \\ -3 & 4 \end{bmatrix}.$$

To solve to ODE $\boldsymbol{x}' = A\boldsymbol{x}$, we diagonalize A:

$$A = PDP^{-1}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}, \quad D = \mathrm{diag}(1,3).$$

Then

$$e^{tA} = P[\mathrm{diag}(e^t, e^{3t})]P^{-1}.$$

Numerical Example (continued)

Example (Continued)

Computing $e^{tA}\boldsymbol{p}_j$ for $\boldsymbol{p}_j$ equal to the columns of P, we get the solutions

$$e^{tA}\boldsymbol{p}_j = P\text{diag}(e^t, e^{3t})\boldsymbol{e}_j \quad \rightsquigarrow \quad \boldsymbol{x}(t) = \begin{bmatrix} e^t \\ e^t \end{bmatrix}, \begin{bmatrix} e^{3t} \\ 3e^{3t} \end{bmatrix}.$$

In terms of the original ODE, this gives the solutions $y(t) = e^t$ and $y(t) = e^{3t}$.

In fact, any solution is a linear combination of the two (independent) solutions above, because the set of solutions is a vector space with dimension two.

Complex Eigenvalues

- If A is diagonalizable with complex eigenvalues, then the method above will yield complex-valued solutions to a real-valued ODE. ☹
- Instead, recall that if A is a real-valued 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$, then we can write

$$A = PCP^{-1}, \quad C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

where the columns of P are given by $\mathrm{Re}(\boldsymbol{v})$ and $\mathrm{Im}(\boldsymbol{v})$ for an eigenvector $\boldsymbol{v}$ corresponding to $\lambda = a - bi$.

Complex Eigenvalues (continued)

- To compute e^{Ct}, write

$$C = aI + b\sigma, \quad \sigma = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

  and note that $I\sigma = \sigma I = \sigma$.

- Now compute

$$\sigma^2 = -I, \quad \sigma^3 = -\sigma, \quad \sigma^4 = I, \ldots$$

  from which we deduce

$$e^{bt\sigma} = \begin{bmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{bmatrix}.$$

Complex Eigenvalues (Continued)

- Finally (recalling $A = PCP^{-1}$),

$$e^{At} = e^{at} P e^{bt\sigma} P^{-1} = e^{at} P \begin{bmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{bmatrix} P^{-1}.$$

- As before, to solve the ODE we would use the vectors comprising the columns of P. This leads to the following solutions:

$$\boldsymbol{x}_1(t) = e^{at} P \begin{bmatrix} \cos(bt) \\ \sin(bt) \end{bmatrix} = e^{at}[\cos(bt)\,\mathrm{Re}(\boldsymbol{v}) + \sin(bt)\,\mathrm{Im}(\boldsymbol{v})],$$

$$\boldsymbol{x}_2(t) = e^{at} P \begin{bmatrix} -\sin(bt) \\ \cos(bt) \end{bmatrix} = e^{at}[-\sin(bt)\,\mathrm{Re}(\boldsymbol{v}) + \cos(bt)\,\mathrm{Im}(\boldsymbol{v})].$$

Complex Eigenvalues (conclusion)

Theorem

Consider the real-valued 2×2 *ODE system*

$$\boldsymbol{x}' = A\boldsymbol{x}.$$

Suppose A *has eigenvalues* $\lambda = a \pm ib$ *and that* $\boldsymbol{v}$ *is an eigenvector corresponding to eigenvalue* $\lambda = a - ib$. *A basis of solutions is given by*

$$\boldsymbol{x}_1(t) = e^{at}[\cos(bt)\, Re(\boldsymbol{v}) + \sin(bt)\, Im(\boldsymbol{v})],$$
$$\boldsymbol{x}_2(t) = e^{at}[-\sin(bt)\, Re(\boldsymbol{v}) + \cos(bt)\, Im(\boldsymbol{v})].$$

Phenomenology in Planar Systems

In the case of diagonalizable, invertible 2×2 matrices, we can find the following behaviors of solutions to the corresponding ODE systems (characterized by the eigenvalues):

- Source: $++$
- Sink: $--$
- Saddle point: $+-$
- Center: complex, $a = 0$
- Spiral source: complex, $a > 0$
- Spiral sink: complex, $a < 0$

ODE Trajectories

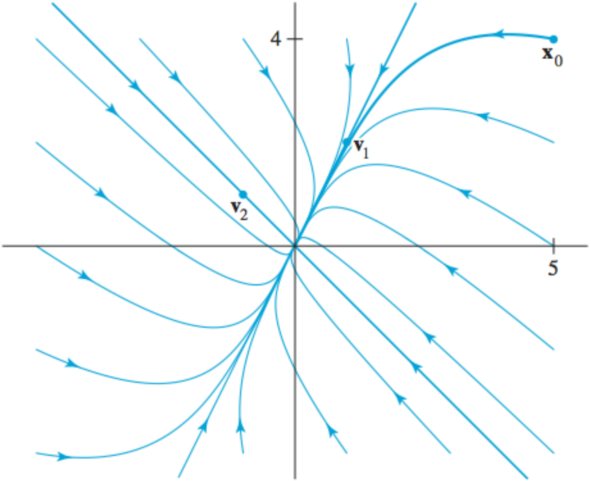


FIGURE 2 The origin as an attractor.

ODE Trajectories

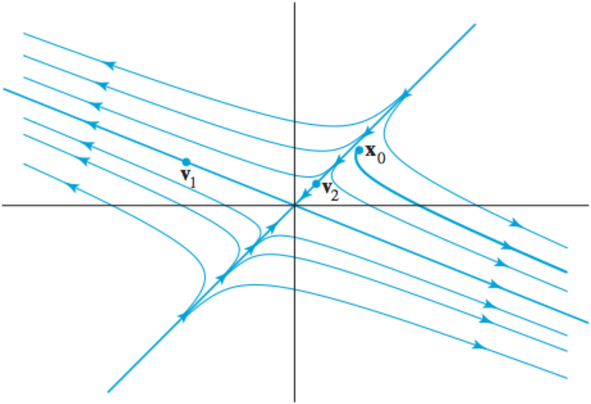


FIGURE 3 The origin as a saddle point.

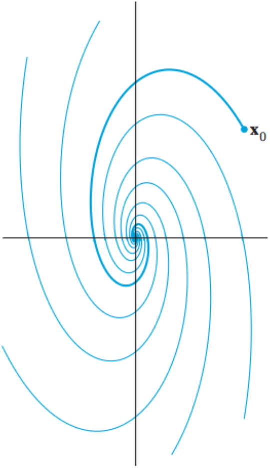


FIGURE 5
The origin as a spiral point.

Non-diagonalizable Matrices

- While not every matrix is diagonalizable, every matrix can be put in *Jordan canonical form* (which is closely related).
- This form can be used to compute the matrix exponential.
- We will not pursue the general theory, but let us consider one example.

Example

Consider

$$y'' - 2y' + y = 0 \implies \boldsymbol{x}' = A\boldsymbol{x}, \quad A = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$$

The eigenvalues of A are $\lambda = 1$ (multiplicity 2); however,

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

has one-dimensional null space, spanned by $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

So A is not diagonalizable!

Observe, however, that $(A - I)^2 = 0$.

Example (Continued)

So, we may write

$$e^{At} = e^{It}e^{(A-I)t} = e^t\{I + (A-I)t\} = e^t \begin{bmatrix} 1-t & t \\ -t & 1+t \end{bmatrix}.$$

To find solutions, we first use the eigenvector, yielding the solution

$$\boldsymbol{x}_1(t) = e^{At}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = e^t \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We next choose a vector independent of the eigenvector, e.g.

$$\boldsymbol{x}_2(t) = e^{At}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = e^t \begin{bmatrix} t \\ 1+t \end{bmatrix}.$$

In terms of the original ODE, we get the solutions

$$y_1(t) = e^t, \quad y_2(t) = te^t.$$

A particle moving in a planar force field has a position vector $\mathbf{x}$ that satisfies $\mathbf{x}' = A\mathbf{x}$. The 2×2 matrix A has eigenvalues 3 and 2, with corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$. Find the position of the particle at time t assuming that $\mathbf{x}(0) = \begin{bmatrix} -30 \\ -2 \end{bmatrix}$.

Select the correct choice below and fill in the answer boxes to complete your choice.

(Type integers or simplified fractions.)

A. $\mathbf{x}(t) = (\ \square\ ) \begin{bmatrix} -6 \\ 1 \end{bmatrix} e^{2t} + (\ \square\ ) \begin{bmatrix} 6 \\ 1 \end{bmatrix} e^{3t}$

B. $\mathbf{x}(t) = \left(\frac{3}{2}\right) \begin{bmatrix} -6 \\ 1 \end{bmatrix} e^{3t} + \left(-\frac{7}{2}\right) \begin{bmatrix} 6 \\ 1 \end{bmatrix} e^{2t}$

MyLab Problems

Solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ for $t \geq 0$, with $\mathbf{x}(0) = (5,2)$. Classify the nature of the origin as an attractor, repeller, or saddle point of the dynamical system described by $\mathbf{x}' = A\mathbf{x}$. Find the directions of greatest attraction and/or repulsion.

$$A = \begin{bmatrix} -7 & -1 \\ 3 & -11 \end{bmatrix}$$

Solve the initial value problem.

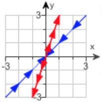
$$\mathbf{x}(t) = -\frac{3}{2}\begin{bmatrix} 1 \\ 3 \end{bmatrix} e^{-10t} + \frac{13}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{-8t}$$

Classify the nature of the origin as an attractor, repeller, or saddle point. Choose the correct answer below.

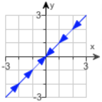
- Attractor
- Saddle Point
- Repeller

Choose the correct graph below that represents the direction(s) of greatest attraction and/or repulsion.

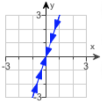
A.



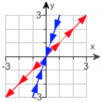
B.



C.



D.



MyLab Problems

For matrix A below, make a change of variables that decouples the equation $\mathbf{x}' = A\mathbf{x}$. Write the equation $\mathbf{x}(t) = P\mathbf{y}(t)$ that leads to the uncoupled system $\mathbf{y}' = D\mathbf{y}$, specifying P and D.

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

Choose the correct values of P and D below that result in the decoupled system $\mathbf{y}' = D\mathbf{y}$ when $\mathbf{x}(t) = P\mathbf{y}(t)$.

A. $P = \begin{bmatrix} -5 & 1 \\ 2 & 1 \end{bmatrix}, D = \begin{bmatrix} -9 & 0 \\ 0 & -2 \end{bmatrix}$

B. $P = \begin{bmatrix} -5 & 1 \\ 2 & 1 \end{bmatrix}, D = \begin{bmatrix} -2 & 0 \\ 0 & -9 \end{bmatrix}$

C. $P = \begin{bmatrix} 5 & 1 \\ 2 & 1 \end{bmatrix}, D = \begin{bmatrix} -2 & 0 \\ 0 & -9 \end{bmatrix}$

D. $P = \begin{bmatrix} 1 & 2 \\ 1 & -5 \end{bmatrix}, D = \begin{bmatrix} -9 & 0 \\ 0 & -2 \end{bmatrix}$

Write the equation $\mathbf{x}(t) = P\mathbf{y}(t)$ using the matrix P found above.

$$\mathbf{x}(t) = \begin{bmatrix} -5 & 1 \\ 2 & 1 \end{bmatrix}\mathbf{y}(t)$$

The power method applies to an $n \times n$ matrix with a **strictly dominant eigenvalue** λ_1, i.e. an eigenvalue larger in absolute value than all others. The power method produces a sequence of scalars approximating λ_1 and a sequence of vectors approximating a corresponding eigenvector.

The Power Method

THE POWER METHOD FOR ESTIMATING A STRICTLY DOMINANT EIGENVALUE

1. Select an initial vector $\mathbf{x}_0$ whose largest entry is 1.

2. For $k = 0, 1, \ldots,$

a. Compute $A\mathbf{x}_k$.

b. Let μ_k be an entry in $A\mathbf{x}_k$ whose absolute value is as large as possible.

c. Compute $\mathbf{x}_{k+1} = (1/\mu_k)A\mathbf{x}_k$.

3. For almost all choices of $\mathbf{x}_0$, the sequence $\{\mu_k\}$ approaches the dominant eigenvalue, and the sequence $\{\mathbf{x}_k\}$ approaches a corresponding eigenvector.

The Inverse Power Method

The inverse power method approximates the value of an arbitrary eigenvalue, provided one has a good initial estimate α of the true eigenvalue λ. It works by applying the power method to $B = (A - \alpha I)^{-1}$, relying on the fact that if the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$, then the eigenvalues of B are

$$\tfrac{1}{\lambda_1 - \alpha}, \ldots, \tfrac{1}{\lambda_n - \alpha}$$

with the same eigenvectors.

Inverse Power Method

THE INVERSE POWER METHOD FOR ESTIMATING AN EIGENVALUE λ OF A

1. Select an initial estimate α sufficiently close to λ.

2. Select an initial vector $\mathbf{x}_0$ whose largest entry is 1.

3. For $k = 0, 1, \ldots,$

a. Solve $(A - \alpha I)\mathbf{y}_k = \mathbf{x}_k$ for $\mathbf{y}_k$.

b. Let μ_k be an entry in $\mathbf{y}_k$ whose absolute value is as large as possible.

c. Compute $\nu_k = \alpha + (1/\mu_k)$.

d. Compute $\mathbf{x}_{k+1} = (1/\mu_k)\mathbf{y}_k$.

4. For almost all choices of $\mathbf{x}_0$, the sequence $\{\nu_k\}$ approaches the eigenvalue λ of A, and the sequence $\{\mathbf{x}_k\}$ approaches a corresponding eigenvector.

Math 3108 - Fall 2019
Chapter 6: Orthogonality and Least Squares

The Real Inner Product

- We previously encountered the **dot product** (also called the **inner product**) between two vectors in $\mathbb{R}^n$, e.g.

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + \cdots + u_n v_n.$$

- We may express this in terms of matrix multiplication by making use of the **transpose** operation:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v}.$$

Example

$$[-2 \ \ -5 \quad 1] \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = -1,$$

$$[3 \ 2 \ \ -3] \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix} = -1.$$

Algebraic Properties of the Real Inner Product

THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$

b. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$

c. $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(c\mathbf{v})$

d. $\mathbf{u}\cdot\mathbf{u} \geq 0$, and $\mathbf{u}\cdot\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

- Here c refers to a *real* scalar.

Conjugate Transpose

- From this point forward, we will regularly consider the case of **complex vectors** $\boldsymbol{u} \in \mathbb{C}^n$.
- For a complex matrix $A \in \mathbb{C}^{m\times n}$, we define the **conjugate transpose** or **adjoint** of A by

$$A^* = (\bar{A})^T \in \mathbb{C}^{n\times m}.$$

- For example,

$$\begin{bmatrix} 1 & 1+i \\ 2 & 3i \end{bmatrix}^* = \begin{bmatrix} 1 & 2 \\ 1-i & -3i \end{bmatrix}.$$

- We have the following algebraic properties:

$$(\alpha A + \beta B)^* = \bar{\alpha}A^* + \bar{\beta}B^*, \quad (AC)^* = C^*A^*.$$

Complex Inner Product

- When $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{C}^n$, we define the **inner product** of $\boldsymbol{u}$ and $\boldsymbol{v}$

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^* \boldsymbol{v}.$$

- This means

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ a_n \end{bmatrix} = \bar{a}_1 b_1 + \cdots + \bar{a}_n b_n \in \mathbb{C}.$$

- For $A = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_k] \in \mathbb{C}^{n \times k}$ and $B = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_\ell] \in \mathbb{C}^{n \times \ell}$, we have

$$(A^* B)_{ij} = \boldsymbol{u}_i \cdot \boldsymbol{v}_j.$$

Algebraic Properties of the Complex Inner Product

Theorem

For $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$:

- $\boldsymbol{u} \cdot \boldsymbol{v} = \overline{\boldsymbol{v} \cdot \boldsymbol{u}}$
- $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$
- $\alpha(\boldsymbol{u} \cdot \boldsymbol{v}) = (\bar{\alpha}\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (\alpha \boldsymbol{v})$
- *If* $\boldsymbol{u} = [a_1 \ \cdots \ a_n]^T$, *then*

$$\boldsymbol{u} \cdot \boldsymbol{u} = |a_1|^2 + \cdots + |a_n|^2 \geq 0,$$

  and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ *if and only if* $\boldsymbol{u} = 0$.

- **Warning!** Some algebraic properties are different in the real and complex cases!

Length, Norm, Distance

Definition

The **length** or **norm** of a vector $\boldsymbol{v}$ is the nonnegative real number $\|\boldsymbol{v}\|$ defined by

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

- Why take the square root?
- The definition is the same whether $\boldsymbol{v} \in \mathbb{R}^n$ or $\boldsymbol{v} \in \mathbb{C}^n$.
- This notion of length agrees with the standard geometric notion.

Example. The length of $\boldsymbol{v} = [1 \ \ -2 \ 2 \ 0]^T$ is $\|\boldsymbol{v}\| = 3$.

- A vector $\boldsymbol{u}$ is a **unit vector** if $\|\boldsymbol{u}\| = 1$.
- We use the norm to measure **distances** between vectors:

$$\text{distance between } \boldsymbol{u} \text{ and } \boldsymbol{v} \quad = \quad \|\boldsymbol{u} - \boldsymbol{v}\|.$$

Properties of Length

Theorem

For vectors $\boldsymbol{u}, \boldsymbol{v}$ *and a scalar* α*:*

- $\|\alpha \boldsymbol{u}\| = |\alpha| \, \|\boldsymbol{u}\|$
- $\|\boldsymbol{u} \cdot \boldsymbol{v}\| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$ *(***Cauchy–Schwarz inequality***)*
- $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ *(***triangle inequality***)*

- This theorem holds in both the real and complex cases.

Proofs of Inequalities (real-valued case)

Proof of Cauchy–Schwarz.

Define $f(\lambda) = \langle \boldsymbol{u} + \lambda \boldsymbol{v}, \boldsymbol{u} + \lambda \boldsymbol{v} \rangle \geq 0$. Now note that

$$f(\lambda) = \lambda^2 \|\boldsymbol{v}\|^2 + 2\lambda \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \|\boldsymbol{u}\|^2 \geq 0$$

is a quadratic polynomial in λ. So its discriminant is ≤ 0. □

Triangle Inequality.

$$\begin{aligned} \|\boldsymbol{u} + \boldsymbol{v}\|^2 &= (\boldsymbol{u} + \boldsymbol{v}) \cdot (\boldsymbol{u} + \boldsymbol{v}) = \|\boldsymbol{u}\|^2 + 2\boldsymbol{u} \cdot \boldsymbol{v} + \|\boldsymbol{v}\|^2 \\ &\leq \|\boldsymbol{u}\|^2 + 2\|\boldsymbol{u}\|\,\|\boldsymbol{v}\| + \|\boldsymbol{v}\|^2 = (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2. \end{aligned}$$

□

Examples

Example

Let $\boldsymbol{u} = [7\ 1]^T$ and $\boldsymbol{v} = [3\ 2]^T$.

(i) Find a unit vector that gives a basis for Span$\{\boldsymbol{u}\}$. (Equivalently: find a unit vector in the same direction as $\boldsymbol{u}$.)

(ii) Compute the distance between $\boldsymbol{u}$ and $\boldsymbol{v}$.

(i): Compute

$$\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} = \frac{1}{\sqrt{50}}\boldsymbol{u} = \begin{bmatrix} 7/\sqrt{50} \\ 1/\sqrt{50} \end{bmatrix}.$$

(ii): Compute

$$\|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}.$$

- By the Cauchy–Schwarz inequality, we always have

$$-1 \leq \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|} \leq 1 \qquad (\text{for} \quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n).$$

- Thus there exists θ such that

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta.$$

  We call θ the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.

Definition

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** (or **perpendicular**) if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

[*Note:* The definition is the same whether the vectors are in $\mathbb{R}^n$ or $\mathbb{C}^n$.]

Orthogonality

Definition

- Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$
- A set S is **orthogonal** if $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal for every distinct $\boldsymbol{u}, \boldsymbol{v} \in S$.
- A set S is **orthonormal** if it is orthogonal and every element of S is a unit vector.

Note: The zero vector is orthogonal to every other vector.

Notation. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal, we write $\boldsymbol{u} \perp \boldsymbol{v}$.

The Pythagorean Theorem

Theorem (Theorem 2)

Two vectors $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are orthogonal if and only if*

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2.$$

Proof (real-valued case).

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\boldsymbol{u}\cdot\boldsymbol{v}.$$

$\square$

Orthogonal Complements

Definition

Let W be a set of vectors. The **orthogonal complement** of W is the set

$$W^{\perp} := \{\text{all vectors } \boldsymbol{v} \text{ such that } \boldsymbol{v} \cdot \boldsymbol{w} = 0 \quad \text{for every} \quad \boldsymbol{w} \in W\}.$$

- The definition is the same for $\mathbb{R}^n$ and $\mathbb{C}^n$.
- For any set W, the set $W^{\perp}$ is a **subspace.**
- Typically, we consider the case when W itself is a subspace.
- The only vector belonging to both W and $W^{\perp}$ is $\mathbf{0}$.
- What is $(W^{\perp})^{\perp}$?

Example

The orthogonal complement of the xy-plane in $\mathbb{R}^3$ is the z-axis.

Orthogonal Complements - Key Theorem

Theorem (Theorem 3 - Real Case)

Let A be a real-valued matrix. Then

$$[Col\,A]^{\perp} = Nul(A^T).$$

This follows from the more general complex-valued case:

Theorem (Theorem 3 - Complex Case)

Let A be a matrix. Then

$$[Col\,A]^{\perp} = Nul(A^*).$$

Orthogonal Complements - Key Theorem

Proof.

Write $A = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_n]$. Then

$$\mathbf{0} = A^*\boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1^* \\ \vdots \\ \boldsymbol{v}_n^* \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1^*\boldsymbol{x} \\ \vdots \\ \boldsymbol{v}_n^*\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{v}_n \cdot \boldsymbol{x} \end{bmatrix}$$

if and only $\boldsymbol{v}_1 \cdot \boldsymbol{x} = \cdots = \boldsymbol{v}_n \cdot \boldsymbol{x} = 0$.

$\square$

- In the real case, this implies $\mathrm{Nul}\,(A) = [\mathrm{Row}\,(A)]^{\perp}$

Example

Find a basis for the orthogonal complement of $W = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$, where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. Let $A = [\boldsymbol{v}_1 \ \boldsymbol{v}_2]$. Then

$$W^\perp = [\text{Col}\, A]^\perp = \text{Nul}(A^T).$$

Example (Continued)

Example (Continued)

Since

$$A^T = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

we find $W^\perp = \text{span}\{\boldsymbol{v}_3, \boldsymbol{v}_4\}$, where

$$\boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_4 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

a. $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$

Choose the correct answer below.

- **A.** The given statement is false. When **u** and **v** are orthogonal, $\mathbf{u} \cdot \mathbf{v} = 0$, so in that case, $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} \neq 0$.
- **B.** The given statement is false. When **u** and **v** are orthogonal, $\mathbf{u} \cdot \mathbf{v} = 1$, so in that case, $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} \neq 0$.
- **C.** The given statement is true. Since the inner product is commutative, $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$. Subtracting $\mathbf{v} \cdot \mathbf{u}$ from each side of this equation gives $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$.
- **D.** The given statement is true. Since the inner product is commutative, $\mathbf{u} \cdot \mathbf{v} = 1 - \mathbf{v} \cdot \mathbf{u}$. Subtracting $\mathbf{v} \cdot \mathbf{u}$ from each side of this equation gives $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$.

- Recall that a set S of vectors is **orthogonal** if every pair of distinct vectors in S is orthogonal.
- If S is orthogonal/orthonormal and linearly independent, then we call S an **orthogonal/orthonormal basis** for Span(S).

Example

Let

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

- $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \mathbf{0}\}$ - orthogonal, not a basis for $\mathbb{R}^3$
- $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ - orthogonal basis for $\mathbb{R}^3$
- $S = \{\frac{1}{\sqrt{2}}\boldsymbol{v}_1, \boldsymbol{v}_2, \frac{1}{\sqrt{2}}\boldsymbol{v}_3\}$ - orthonormal basis for $\mathbb{R}^3$.

A Test for Orthogonality

- *Question.* Given $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ in $\mathbb{C}^m$, how can we determine whether these vectors are orthogonal/orthonormal?
- *Answer.* Form the matrix $A = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_n] \in \mathbb{C}^{m\times n}$, and observe that

$$A^*A = \begin{bmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_1 \cdot \boldsymbol{v}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{v}_n \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \cdot \boldsymbol{v}_n \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

- Thus $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is orthogonal if and only if A^*A is **diagonal**.
- Moreover, $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is orthonormal if and only if $A^*A = I_n$.

In the real-valued case, we replace A^* with A^T.

Unitary Matrices

The previous discussion leads us to the following definition:

Definition

A matrix $A \in \mathbb{C}^{n\times n}$ is **unitary** if $A^*A = I_n$.

In particular, the columns of A are orthonormal if and only if A is unitary.

For the real-valued case, we have the following:

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is **orthogonal** if $A^TA = I_n$.

Unitary/orthogonal matrices preserve **angles** and **lengths**, cf.

$$(A\boldsymbol{x}) \cdot (A\boldsymbol{y}) = (A\boldsymbol{x})^* A\boldsymbol{y} = \boldsymbol{x}^* A^* A\boldsymbol{y} = \boldsymbol{x}^*\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}.$$

Example

Example

The matrix

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

is unitary. The columns form an orthonormal set.

Orthogonality and Independence

Theorem (Theorem 4)

If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal set of non-zero vectors, then S is independent and hence is a basis for span(S).

Proof.

Suppose $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$. Now take an inner product with $\mathbf{v}_1$ to deduce

$$c_1\|\mathbf{v}_1\|^2 = 0 \implies c_1 = 0.$$

And so on... $\square$

This result holds in both the real and complex settings.

Why do we care?

- Given a subspace W with a basis B, finding the B-coordinates of a vector $\boldsymbol{v}$ involves solving a system of linear equations.
- If B is an **orthogonal/orthonormal basis**, then finding the coordinates relative to B becomes very simple.

Theorem (Theorem 5)

If $B = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is an orthogonal basis for W and $y \in W$, then

$$\boldsymbol{y} = \frac{\boldsymbol{u}_1 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{u}_p \cdot \boldsymbol{y}}{\|\boldsymbol{u}_p\|^2}\boldsymbol{u}_p.$$

- To prove it, suppose $\boldsymbol{y} = \alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_p \boldsymbol{u}_p$ and compute $\boldsymbol{u}_j \cdot \boldsymbol{y}$ for each j.
- *Warning:* The order $\boldsymbol{u}_j \cdot \boldsymbol{y}$ (versus $\boldsymbol{y} \cdot \boldsymbol{u}_j$) matters for complex vectors.

Example

The set $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$, where

$$\boldsymbol{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}.$$

Write the vector

$$\boldsymbol{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$$

as a linear combination of $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$.

Example (Continued)

Example (Continued)

Recall

$$\boldsymbol{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}.$$

Thus

$$\begin{aligned} \boldsymbol{y} &= \tfrac{\boldsymbol{u}_1 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1 + \tfrac{\boldsymbol{u}_2 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_2\|^2}\boldsymbol{u}_2 + \tfrac{\boldsymbol{u}_3 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_3\|^2} \\ &= \tfrac{11}{11}\boldsymbol{u}_2 + \tfrac{-12}{6}\boldsymbol{u}_2 + \tfrac{-33}{33/2}\boldsymbol{u}_3 \\ &= \boldsymbol{u}_1 - 2\boldsymbol{u}_2 - 2\boldsymbol{u}_3. \end{aligned}$$

Remark about MyLab Homework

Homework from Section 6.2 will involve questions about orthogonal projection and distance minimization. We discuss these topics in the slides for Section 6.3.

Determine whether the set of vectors is orthogonal.

$$\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$$

MyLab Problems

Assume all vectors are in $\mathbb{R}^n$. Mark each statement True or False. Justify each answer.

a. Not every orthogonal set in $\mathbb{R}^n$ is linearly independent.

- **A.** False. Every orthogonal set of nonzero vectors is linearly independent and zero vectors cannot exist in orthogonal sets.
- **B.** True. Orthogonal sets with fewer than n vectors in $\mathbb{R}^n$ are not linearly independent.
- **C.** False. Orthogonal sets must be linearly independent in order to be orthogonal.
- **D.** True. Every orthogonal set of nonzero vectors is linearly independent, but not every orthogonal set is linearly independent.

Our first main goal in this section is to prove the following theorem:

Theorem (Orthogonal Decomposition Theorem - Theorem 8)

Let W be a subspace of $\mathbb{C}^n$. For every $\boldsymbol{x}$ in $\mathbb{C}^n$, there exist unique $\boldsymbol{y} \in W$ and $\boldsymbol{z} \in W^\perp$ such that $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$.

- Assuming the theorem, we define the **orthogonal projection of $\boldsymbol{x}$ onto W** by

$$\mathrm{proj}_W(\boldsymbol{x}) = \boldsymbol{y}, \quad \text{where} \quad \boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}, \quad y \in W,\ \boldsymbol{z} \in W^\perp.$$

- *Note:* Orthogonal projection is a linear transformation.
- *Note:* If $\boldsymbol{x} \in W$, then $\mathrm{proj}_W(\boldsymbol{x}) = \boldsymbol{x}$.
- Later we will need to figure out how to actually compute these things!

Preliminary Lemmas

Lemma

If S is an independent set in W and T is an independent set in $W^\perp$, then the union of S and T is an independent set.

Proof.

Essential fact: W and $W^\perp$ share only the zero vector. $\square$

Lemma

If $W \subset \mathbb{C}^n$ has dimension p, then $W^\perp$ has dimension $n-p$.

Proof.

Essential facts: Rank-Nullity Theorem and $[\mathrm{Col}\,A]^\perp = \mathrm{Nul}\,(A^*)$. $\square$

Orthogonal Decomposition

Proof of Orthogonal Decomposition.

Let B be a basis for W and C be a basis for $W^\perp$. By the lemmas above, the union of B and C is a basis for $\mathbb{C}^n$. Then every $\boldsymbol{x} \in \mathbb{C}^n$ has a unique representation as $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{y} \in \text{Span}(B)$ and $\boldsymbol{z} \in \text{Span}(C)$. □

As mentioned above, given this decomposition we define

$$\text{proj}_W(\boldsymbol{x}) = \boldsymbol{y}.$$

- Observe that $\text{proj}_W(\boldsymbol{x})$ always belongs to W.
- Observe also that

$$\boldsymbol{x} - \text{proj}_W(\boldsymbol{x}) = \text{proj}_{W^\perp}(\boldsymbol{x}).$$

Simple Example

Example

If W is the xy-plane in $\mathbb{R}^3$, then the orthogonal projection of a vector $\boldsymbol{v} = [v_1 \;\; v_2 \;\; v_3]^T$ is simply $[v_1 \;\; v_2 \;\; 0]^T$.

In general, it is not so obvious how to compute the orthogonal projection onto a subspace...

The Plan

- The orthogonal projection proj_W defines a **linear transformation** from $\mathbb{C}^n$ to $\mathbb{C}^n$ (or $\mathbb{R}^n$ to $\mathbb{R}^n$), but at this moment it is only abstractly defined.
- **Goal 1.** Find a formula for the **matrix representation of proj_W**.
- **Goal 2.** Show that if we have an **orthogonal/orthonormal basis** for W, the formula is very simple.
- **Goal 3.** *Compute some numerical examples!*
- **Goal 4.** Relate orthogonal projection to the problem of **distance minimization**.

Goal 1. Matrix Representation

- **Setup.** $B = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$ is a basis for $W \subset \mathbb{C}^n$.
- We seek a matrix M such that $\text{proj}_W(\boldsymbol{x}) = M\boldsymbol{x}$.
- Let us first find the **B-coordinates of $\text{proj}_W(\boldsymbol{x})$**: write

$$\boldsymbol{x} = \alpha_1 \boldsymbol{w}_1 + \cdots + \alpha_p \boldsymbol{w}_p + \boldsymbol{z}, \quad \boldsymbol{z} \in W^\perp,$$

where

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} = [\text{proj}_W(\boldsymbol{x})]_B = \boldsymbol{\alpha}.$$

Goal 1 (continued)

- Take the inner product of

$$\boldsymbol{x} = \alpha_1 \boldsymbol{w}_1 + \cdots + \alpha_p \boldsymbol{w}_p + \boldsymbol{z}$$

with $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p$. This yields

$$\begin{aligned} \boldsymbol{w}_1^* \boldsymbol{x} &= \alpha_1 \boldsymbol{w}_1^* \boldsymbol{w}_1 + \cdots + \alpha_p \boldsymbol{w}_1^* \boldsymbol{w}_p \\ &\vdots \qquad\qquad \vdots \qquad\qquad\quad \vdots \\ \boldsymbol{w}_p^* \boldsymbol{x} &= \alpha_1 \boldsymbol{w}_p^* \boldsymbol{w}_1 + \cdots + \alpha_p \boldsymbol{w}_p^* \boldsymbol{w}_p. \end{aligned}$$

- This may be written compactly as

$$A^* A \alpha = A^* \boldsymbol{x}, \quad A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p] \in \mathbb{C}^{n \times p}.$$

- This is called the **normal system**. The matrix $A^* A \in \mathbb{C}^{p \times p}$ is called the **Gram matrix.**

Goal 1 (continued)

We will show:

- If $A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p]$ with $B = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$ a basis for W, then the Gram matrix A^*A is **invertible**, and so the normal system has solution
$$\boldsymbol{\alpha} = [\text{proj}_W(\boldsymbol{x})]_B = (A^*A)^{-1}A^*\boldsymbol{x}.$$

Thus (using $\boldsymbol{y} = A[\boldsymbol{y}]_B$) we get:

Theorem (Goal 1. Matrix Representation)

We have
$$proj_W(\boldsymbol{x}) = A(A^*A)^{-1}A^*\boldsymbol{x},$$
where $B = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$ is any basis for W and $A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p]$.

Invertibility of the Gram Matrix

Lemma

If $A \in \mathbb{C}^{n\times p}$ then $A^*A \in \mathbb{C}^{p\times p}$ satisfies

$$Nul(A^*A) = Nul(A) \quad \text{and} \quad Rank(A^*A) = Rank(A).$$

In particular, since $\text{rank}(A) = p$ in our setting, the Gram matrix is invertible.

Proof of Lemma.

Key Fact: $\text{Nul}(A^*) = [\text{Col}(A)]^\perp$, so if $A\boldsymbol{x} \in \text{nul}(A^*)$ then $A\boldsymbol{x} = 0$. $\square$

Goal 2. Orthogonal Basis Case

- **Setup.** Suppose $B = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_p\}$ is an **orthogonal** basis for $W \subset \mathbb{C}^n$.
- Writing $A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p]$, we seek a simple formula for the matrix representation for $\text{proj}_W(\boldsymbol{x})$, namely,

$$A(A^*A)^{-1}A^*.$$

- Since B is an orthogonal basis,

$$(A^*A)^{-1} = \text{diag}\{\tfrac{1}{\|\boldsymbol{w}_1\|^2}, \cdots, \tfrac{1}{\|\boldsymbol{w}_p\|^2}\},$$

and so

$$A(A^*A)^{-1}A^* = \tfrac{1}{\|\boldsymbol{w}_1\|^2}\boldsymbol{w}_1\boldsymbol{w}_1^* + \cdots + \tfrac{1}{\|\boldsymbol{w}_p\|^2}\boldsymbol{w}_p\boldsymbol{w}_p^*.$$

Goal 2. Summary.

Theorem (Goal 2. Orthogonal Basis Case)

Suppose $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ *is an orthogonal basis for* W. *Then* proj_W *has the matrix representation*

$$QQ^* = \frac{1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\mathbf{w}_1^* + \cdots + \frac{1}{\|\mathbf{w}_p\|^2}\mathbf{w}_p\mathbf{w}_p^*,$$

where $Q = [\mathbf{w}_1 \cdots \mathbf{w}_p]$. *That is,*

$$\mathrm{proj}_W(\mathbf{x}) = \frac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \cdots + \frac{\mathbf{w}_p \cdot \mathbf{x}}{\|\mathbf{w}_p\|^2}\mathbf{w}_p.$$

- The projection is written as the sum of p 'rank-one' projections onto the lines spanned by each $\mathbf{w}_j$.
- *Remark:* We saw this formula already when computing coordinates of a vector relative to an orthogonal basis!

Goal 3. Numerical Examples!

Example

Let $W = \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\} \subset \mathbb{R}^3$, where

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

Find the matrix representation for proj_W and compute $\text{proj}_W(\boldsymbol{e}_1)$.

Solution. Write $A = [\boldsymbol{v}_1 \ \boldsymbol{v}_2]$. Compute

$$A^T A = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix} \implies (A^T A)^{-1} = \tfrac{1}{21} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}.$$

Then

$$A(A^T A)^{-1} A^T = \tfrac{1}{21} \begin{bmatrix} 5 & 8 & -4 \\ 8 & 17 & 2 \\ -4 & 2 & 20 \end{bmatrix} \implies \text{proj}_W(\boldsymbol{e}_1) = \tfrac{1}{21} \begin{bmatrix} 5 \\ 8 \\ -4 \end{bmatrix}.$$

Goal 3. Numerical Examples!

Example

For the subspace in the previous example, we may also write $W = \text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} \subset \mathbb{R}^3$, where

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} -2/5 \\ 1/5 \\ 2 \end{bmatrix}.$$

Find the matrix representation for proj_W and compute $\text{proj}_W(\boldsymbol{e}_1)$.

Solution. This is an orthogonal basis, so we get

$$\frac{1}{\|\boldsymbol{w}_1\|^2}\boldsymbol{w}_1\boldsymbol{w}_1^* + \frac{1}{\|\boldsymbol{w}_2\|^2}\boldsymbol{w}_2\boldsymbol{w}_2^* = \ldots$$

Example (Continued)

$$\dots = \tfrac{1}{5}\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \tfrac{5}{21}\begin{bmatrix} 4/25 & -2/25 & -4/5 \\ -2/25 & 1/25 & 2/5 \\ -4/5 & 2/5 & 4 \end{bmatrix}$$
$$= \tfrac{1}{21}\begin{bmatrix} 5 & 8 & -4 \\ 8 & 17 & 2 \\ -4 & 2 & 20 \end{bmatrix}.$$

This the same answer as before, since after all it is the same subspace. To compute $\mathrm{proj}_W(\boldsymbol{e}_1)$, we can write

$$\mathrm{proj}_W(\boldsymbol{e}_1) = \tfrac{\boldsymbol{w}_1\cdot e_1}{\|\boldsymbol{w}_1\|^2}\boldsymbol{w}_1 + \tfrac{\boldsymbol{w}_2\cdot \boldsymbol{e}_1}{\|\boldsymbol{w}_2\|^2}\boldsymbol{w}_2 = \tfrac{1}{21}\begin{bmatrix} 5 \\ 8 \\ -4 \end{bmatrix}.$$

Interlude — why do we care?

- Given any basis for a subspace, we have found an explicit formula for the matrix representation of proj_W.
- This formula is much simpler if we can find an *orthogonal* basis for W.
- We will return to the problem of constructing orthogonal bases in the next section ('the Gram–Schmidt algorithm').
- Before that, we once again ask ourselves... why do we care (about orthogonal projections)?

Goal 4. Distance Minimization

Theorem (Theorem 9 - Best Approximation Theorem)

Let W be a subspace. Then $proj_W(\boldsymbol{y})$ is the closest point in W to $\boldsymbol{y}$, i.e.

$$\|\boldsymbol{y} - proj_W(\boldsymbol{y})\| \leq \|\boldsymbol{y} - \boldsymbol{v}\| \quad \textit{for all} \quad \boldsymbol{v} \in W,$$

with equality if and only if $\boldsymbol{v} = proj_W(\boldsymbol{y})$.

Proof.

By the Pythagorean theorem, for any $\boldsymbol{v}$ in W,

$$\|\boldsymbol{y} - \boldsymbol{v}\|^2 = \|\text{proj}_W(\boldsymbol{y}) - \boldsymbol{v}\|^2 + \|\text{proj}_{W^\perp}(\boldsymbol{y})\|^2 \geq \|\text{proj}_{W^\perp}(\boldsymbol{y})\|^2,$$

with equality if and only if $\boldsymbol{v} = \text{proj}_W(\boldsymbol{y})$. $\square$

- *Remark.* This also shows that the **distance from $\boldsymbol{y}$ to W** equals

$$\|\text{proj}_{W^\perp}(\boldsymbol{y})\| = \|\boldsymbol{y} - \text{proj}_W(\boldsymbol{y})\|.$$

Example

Example

Find the distance from y to $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, where

$$y = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \boldsymbol{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

Solution. We first compute $\text{proj}_W(\boldsymbol{y})$. Since $\boldsymbol{u}_1 \perp \boldsymbol{u}_2$, we can use

$$\text{proj}_W(\boldsymbol{y}) = \tfrac{\boldsymbol{u}_1 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1 + \tfrac{\boldsymbol{u}_2 \cdot \boldsymbol{y}}{\|\boldsymbol{u}_2\|^2}\boldsymbol{u}_2 = \ldots = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}.$$

So the distance from $\boldsymbol{y}$ to W is given by

$$\|\boldsymbol{y} - \text{proj}_W(\boldsymbol{y})\| = \sqrt{45}.$$

- The Gram–Schmidt algorithm takes as input a set of vectors $S_{in} = \{\boldsymbol{w}_1, \ldots \boldsymbol{w}_p\}$ and returns an **orthogonal** set of vectors $S_{out} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ such that $\text{Span}(S_{in}) = \text{Span}(S_{out})$.
- The idea is straightforward: at each stage, one performs an orthogonal projection of $\boldsymbol{w}_j$ away from the span of the preceding vectors.

Theorem (Theorem 11)

Let $S_{in} = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$.

- *Let $\mathbf{v}_1 = \mathbf{w}_1$ and $\Omega_1 = Span\{\mathbf{v}_1\}$.*
- *Let $\mathbf{v}_2 = proj_{\Omega_1^\perp}\mathbf{w}_2$ and $\Omega_2 = Span\{\mathbf{v}_1, \mathbf{v}_2\}$.*
- $\ldots$
- *Let $\mathbf{v}_{j+1} = proj_{\Omega_j^\perp}(\mathbf{w}_{j+1})$ and set $\Omega_{j+1} = Span\{\mathbf{v}_1, \ldots, \mathbf{v}_{j+1}\}$.*

The process ends when $j+1 = p$. It produces the orthogonal set

$$S_{out} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$$

with $Span(S_{out}) = Span(S_{in})$. Finally, observe that

$$\mathbf{v}_{j+1} = \mathbf{0} \iff \mathbf{w}_{j+1} \in \Omega_j,$$

so that if S_{in} is independent then S_{out} contains nonzero vectors.

Example

Example

Find an orthogonal basis for the span of the following vectors:

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{w}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{w}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. Apply Gram-Schmidt. Set $\boldsymbol{v}_1 = \boldsymbol{w}_1$. Then

$$\boldsymbol{v}_2 = \boldsymbol{w}_2 - \frac{\boldsymbol{v}_1 \cdot \boldsymbol{w}_2}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Example (Continued)

Next,

$$\boldsymbol{v}_3 = \boldsymbol{w}_3 - \frac{\boldsymbol{v}_1 \cdot \boldsymbol{w}_3}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 - \frac{\boldsymbol{v}_2 \cdot \boldsymbol{w}_3}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 = \cdots = \mathbf{0}.$$

(This reflects the fact that $\boldsymbol{w}_3 \in \text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.) Finally,

$$\boldsymbol{v}_4 = \boldsymbol{w}_4 - \frac{\boldsymbol{v}_1 \cdot \boldsymbol{w}_4}{\|\boldsymbol{v}_1\|^2} - \frac{\boldsymbol{v}_2 \cdot \boldsymbol{w}_4}{\|\boldsymbol{v}_2\|^2} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4\}$ is an orthogonal basis for $\text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4\}$.

Remark. To make an *orthonormal* basis, divide each basis vector by its length.

QR Factorization of Matrices

- Performing the Gram–Schmidt algorithm for vectors $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$ in $\mathbb{C}^n$ is equivalent to performing a **QR factorization** for the matrix $A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p] \in \mathbb{C}^{n \times p}$.
- QR factorization refers to the following:

Theorem (Theorem 12 - The QR Factorization)

Any matrix $A \in \mathbb{C}^{n \times p}$ can be written as $A = QR$, where the columns of $Q \in \mathbb{C}^{n \times p}$ are orthogonal and $R \in \mathbb{C}^{p \times p}$ is an invertible upper triangular matrix.

To prove this, we need to rewrite the $\boldsymbol{w}$'s in terms of the $\boldsymbol{v}$'s in the Gram–Schmidt algorithm.

QR Factorization (continued)

Since the $\{\boldsymbol{v}_j\}$ are orthogonal, we can write

$$\text{proj}_{\Omega_j}(\boldsymbol{w}_{j+1}) = \tfrac{\boldsymbol{v}_1 \cdot \boldsymbol{w}_{j+1}}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 + \cdots + \tfrac{\boldsymbol{v}_j \cdot \boldsymbol{w}_{j+1}}{\|\boldsymbol{v}_j\|^2}\boldsymbol{v}_j,$$

where we only include nonzero $\boldsymbol{v}$'s in the sum above. So, defining

$$r_{k,j+1} = \begin{cases} \tfrac{\boldsymbol{v}_k \cdot \boldsymbol{w}_{j+1}}{\|\boldsymbol{v}_k\|^2} & \text{if } \boldsymbol{v}_k \neq 0 \\ \text{any number you want!} & \text{if } \boldsymbol{v}_k = 0, \end{cases}$$

we get

$$\boldsymbol{v}_{j+1} = \boldsymbol{w}_{j+1} - \text{proj}_{\Omega_j}(\boldsymbol{w}_{j+1}) = \boldsymbol{w}_{j+1} - r_{1,j+1}\boldsymbol{v}_1 - \cdots - r_{j,j+1}\boldsymbol{v}_j,$$

or equivalently: $\boldsymbol{w}_{j+1} = r_{1,j+1}\boldsymbol{v}_1 + \cdots + r_{j,j+1}\boldsymbol{v}_j + \boldsymbol{v}_{j+1}.$

QR Factorization (conclusion)

Rewrite $\boldsymbol{w}_{j+1} = r_{1,j+1}\boldsymbol{v}_1 + \cdots + r_{j,j+1}\boldsymbol{v}_j + \boldsymbol{v}_{j+1}$ in vector form:

$$\boldsymbol{w}_{j+1} = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_{j+1}] \begin{bmatrix} r_{1,j+1} \\ \vdots \\ r_{j,j+1} \\ 1 \end{bmatrix},$$

So $A = QR$, where $A = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_p]$, $Q = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p]$ and

$$R = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & r_{p-1,p} \\ & & & 1 \end{bmatrix}, \quad r_{k,j+1} = \begin{cases} \frac{\boldsymbol{v}_k \cdot \boldsymbol{w}_{j+1}}{\|\boldsymbol{v}_k\|^2} & \text{if } \boldsymbol{v}_k \neq 0 \\ \text{anything} & \text{if } \boldsymbol{v}_k = 0. \end{cases}$$

Example

Example

Return to the previous example. Then the QR factorization for $A = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_4]$ is given by $A = QR$, with $Q = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{0} \ \mathbf{v}_4]$ and

$$R = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 0 & 1 & -1 & 1/2 \\ 0 & 0 & 0 & c \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where c is arbitrary. The coefficients are determined by

$$r_{k,j+1} = \frac{\mathbf{v}_k \cdot \mathbf{w}_{j+1}}{\|\mathbf{v}_k\|^2} \quad \text{for} \quad \mathbf{v}_k \neq 0.$$

Practice Problem

Example

Let

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{w}_1 \perp \boldsymbol{w}_2.$$

Extend $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ to an orthogonal basis for $\mathbb{R}^4$.

Solution Sketch:

- Write $A = [\boldsymbol{w}_1 \;\; \boldsymbol{w}_2]$ and $W = \mathrm{Col}(A)$.

(i) Find a basis for $W^{\perp} = \mathrm{Nul}(A^T)$ (row reduction).

(ii) Apply Gram-Schmidt to the basis obtained in (i).

- $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and the basis obtained in (ii) give you an orthogonal basis for $\mathbb{R}^4$.

Section 6.5 - Least-Squares Problems

- One of the first topics we discussed was to determine consistency and find solutions to systems of the form $A\boldsymbol{x} = \boldsymbol{b}$.
- When $A\boldsymbol{x} = \boldsymbol{b}$ is **not** consistent, we would like to find an $\boldsymbol{x}$ that comes 'as close as possible' to solving the system.

Definition

If A is an $m \times n$ matrix and $\boldsymbol{b} \in \mathbb{C}^m$, a **least squares solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}} \in \mathbb{C}^n$ such that

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$$

for all $\boldsymbol{x} \in \mathbb{C}^n$.

Note: If $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then any solution is automatically a least squares solution.

Least Squares Solutions; Projections; Normal System

- The system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent if $\boldsymbol{b}$ does not belong to $\mathrm{Col}(A)$.
- To remedy this, we instead consider the system

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \quad \hat{\boldsymbol{b}} := \mathrm{proj}_{\mathrm{Col}(A)}(\boldsymbol{b}),$$

  which always has a solution (and is equivalent to $A\boldsymbol{x} = \boldsymbol{b}$ if $\boldsymbol{b} \in \mathrm{Col}(A)$).
- Furthermore, these are guaranteed to be least squares solutions by the best approximation theorem, cf.

$$\|\boldsymbol{b} - \hat{\boldsymbol{b}}\| \leq \|\boldsymbol{b} - \boldsymbol{v}\| \quad \text{for any} \quad \boldsymbol{v} \in \mathrm{Col}(A).$$

- To compute the matrix representation for the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$, we need to solve the **normal system** $A^*A\alpha = A^*\boldsymbol{b}$. The projection is then given by $A\alpha$.

Least Squares Solutions

Theorem (Theorem 13)

The normal system $A^*A\boldsymbol{x} = A^*\boldsymbol{b}$ is always consistent. Solutions to this system are precisely the least squares solutions to $A\boldsymbol{x} = \boldsymbol{b}$.

Proof.

We showed $\mathrm{rank}(A^*A) = \mathrm{rank}(A^*)$, which implies $\mathrm{col}(A^*A) = \mathrm{col}(A^*)$.

Next, if $A^*A\hat{\boldsymbol{x}} = A^*\boldsymbol{b}$, then

$$\boldsymbol{b} - A\hat{\boldsymbol{x}} \in \mathrm{Nul}(A^*) = [\mathrm{Col}(A)]^\perp.$$

Since $A\hat{\boldsymbol{x}} \in \mathrm{Col}(A)$, it follows that $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}} = \mathrm{proj}_{\mathrm{Col}(A)}(\boldsymbol{b})$.
Similarly, if $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$ then $A^*A\hat{\boldsymbol{x}} = A^*\hat{\boldsymbol{b}} = A^*\boldsymbol{b}$, since

$$\boldsymbol{b} - \hat{\boldsymbol{b}} \in [\mathrm{Col}(A)]^\perp = \mathrm{Nul}(A^*).$$

□

Example

Show that $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, where

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Then find the least squares solution(s) to $A\boldsymbol{x} = \boldsymbol{b}$.

Solution. First, the system is inconsistent since

$$[A \mid \boldsymbol{b}] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Example (Continued)

Example (Continued)

Now compute $A^T A$ and $A^T \boldsymbol{b}$ and perform row reduction to find

$$[A^T A \mid A^T \boldsymbol{b}] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

giving the least squares solutions

$$\hat{\boldsymbol{x}} = \begin{bmatrix} 1/3 \\ -1/3 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

- In the previous example, the free variable appeared due to the fact that A has a nontrivial null space.
- The least square solution is unique if and only if the columns of A are independent, which holds if and only if A^*A is invertible. In this case, the unique solution is

$$\hat{\boldsymbol{x}} = (A^*A)^{-1}A^*\boldsymbol{b}.$$

Example

Example

Find the least squares solution(s) for the inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution. We find

$$A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

is invertible. So the unique solution is

$$\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b} = \cdots = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Least Squares Error

- The **least squares error** for the system $A\boldsymbol{x} = \boldsymbol{b}$ is defined by the smallest possible value of

$$\|\boldsymbol{b} - A\boldsymbol{x}\|$$

  over all choices of $\boldsymbol{x}$. It is achieved by choosing any least squares solution $\hat{\boldsymbol{x}}$ (cf. the best approximation theorem).
- The least squares error computes the distance between $\boldsymbol{b}$ and $\mathrm{Col}(A)$.

Examples

Example

- In the first example, the least squares error is

$$\|A\hat{\boldsymbol{x}} - \boldsymbol{b}\| = \tfrac{2}{3}\sqrt{3}.$$

- In the second example, the least squares error is

$$\|A\hat{\boldsymbol{x}} - \boldsymbol{b}\| = 2\sqrt{21}.$$

Other Approaches...

- If the columns of A are orthogonal, then we can compute $\hat{\boldsymbol{b}}$ simply and then solve $A\boldsymbol{x} = \hat{\boldsymbol{b}}$.
- If A has linearly independent columns and $A = QR$ is the QR factorization of A, then the least squares solution is given by

$$\hat{\boldsymbol{x}} = R^{-1}Q^*\boldsymbol{b},$$

since then

$$A\hat{\boldsymbol{x}} = QQ^*\boldsymbol{b},$$

and QQ^* is the orthogonal projection onto $\mathrm{Col}(A)$.

Let $A = \begin{bmatrix} 6 & 1 \\ -4 & -2 \\ 4 & 9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 31 \\ -18 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

$A\mathbf{u} = \begin{bmatrix} 46 \\ -28 \\ 14 \end{bmatrix}$ (Simplify your answer.)

$A\mathbf{v} = \begin{bmatrix} 16 \\ -8 \\ -6 \end{bmatrix}$ (Simplify your answer.)

Compare $A\mathbf{u}$ and $A\mathbf{v}$ with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$?

- A. $A\mathbf{u}$ is closer to $\mathbf{b}$ than $A\mathbf{v}$ is. Thus, $\mathbf{v}$ cannot be a least-squares solution of $A\mathbf{x} = \mathbf{b}$, but $\mathbf{u}$ can be.
- B. $A\mathbf{v}$ is closer to $\mathbf{b}$ than $A\mathbf{u}$ is. Thus, $\mathbf{u}$ cannot be a least-squares solution of $A\mathbf{x} = \mathbf{b}$, but $\mathbf{v}$ can be.
- C. $A\mathbf{u}$ and $A\mathbf{v}$ are equally close to $\mathbf{b}$. Thus, both can be the least-squares solution of $A\mathbf{x} = \mathbf{b}$.
- D. $A\mathbf{u}$ and $A\mathbf{v}$ are equally close to $\mathbf{b}$. Thus, neither can be the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

MyLab Probems

True or False Questions:

a. The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

b. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that $\|\mathbf{b} - A\mathbf{x}\| \le \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

d. Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

e. If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

- This section covers several applications, including (i) least squares lines and linear models, (ii) more general least squares curves, (iii) multiple regression.
- We focus on the case of least squares lines.

Least-Squares Lines

- Suppose we want to fit data points $(x_1, y_1), \dots, (x_n, y_n)$ to a line $y = \beta_0 + \beta_1 x$. This corresponds to trying to solve the linear system

$$X\boldsymbol{\beta} = \mathbf{y}, \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

- Typically this system will not be consistent, so instead we find the **least squares solution** β.
- This yields the least-squares line for the data.
- This is equivalent to minimizing the length of the **residual vector** $\varepsilon = \mathbf{y} - X\boldsymbol{\beta}$ over all choices of $\boldsymbol{\beta}$.
- This extends naturally to higher order polynomial approximations.

Multiple Regression

- This technique also extends to the case when the data depends on multiple variables. For example, if one assumes a relationship of the form $y = \beta_0 + \beta_1 u + \beta_2 v$ (a plane instead of a line), then we should find the least squares solution to $X\boldsymbol{\beta} = \boldsymbol{y}$, where

$$X = \begin{bmatrix} 1 & u_1 & v_1 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

The following is the definition of a (real) inner product space:

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0 \quad$ and $\quad \langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

General Properties

Whenever we have an inner product on a vector space, we get the following:

- Length, distance, angle
- Cauchy–Schwarz inequality
- Triangle inequality
- Orthogonality, Pythagorean theorem
- Orthogonal bases
- Orthogonal projections
- Gram–Schmidt algorithm...

Examples

- Let $V = \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ be a positive definite real symmetric matrix ($A = A^T$). Then
$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = x^T Ay$$
is a real inner product.
- Let $V = \mathbb{C}^n$ and $A \in \mathbb{C}^{n\times n}$ be a positive definite hermitian matrix ($A = A^*$). Then
$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^* A\boldsymbol{y}$$
is a (complex) inner product.
- Let $V = C([0, 2\pi])$. Then
$$\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\, dt$$
is a real inner product on V.

Section 6.8 - Applications of Inner Product Spaces

- In the book, several applications are discussed, including weighted least squares, trend analysis, and Fourier series.
- We will focus on a short discussion of Fourier series.

- The aim of Fourier series is to represent an arbitrary continuous function f on $[0, 2\pi]$ as a linear combination of waves of fixed frequencies.
- In particular, for each $n = 1, 2, \dots,$ we want to find the best approximation to f using the functions from

$$S_n = \{1, \cos t, \cos 2t, \dots, \cos nt, \sin t, \sin 2t, \dots, \sin nt\}.$$

- We already know what to do: we should use the orthogonal projection

$$f_n = \mathrm{proj}_{\mathrm{Span}(S_n)} f.$$

- *Note:* This notion of orthogonality and projection is given in terms of the inner product on $C([0, 2\pi])$!

Orthogonality

- The orthogonal projection onto $\mathrm{Span}(S_n)$ is straightforward to compute because S_n is an orthogonal set!

Example

For $m \neq n$,

$$\begin{aligned}\langle \cos mt, \cos nt \rangle &= \int_0^{2\pi} \cos mt \cos nt \, dt \\ &= \tfrac{1}{2} \int_0^{2\pi} [\cos((m+n)t) + \cos((m-n)t) \, dt = \cdots = 0.\end{aligned}$$

Orthogonal Projection

- Since S_n is orthogonal, we may write

$$f_n = \tfrac{1}{2}a_0 + a_1 \cos t + \cdots + a_n \cos nt + b_1 \sin t + \cdots + b_n \sin nt,$$

where

$$a_k = \frac{\langle \cos kt, f \rangle}{\| \cos kt \|^2} = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos kt \, dt,$$

$$b_k = \frac{\langle \sin kt, f \rangle}{\| \sin kt \|^2} = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin kt \, dt$$

for $k \geq 1$, and $a_0 = \frac{1}{\pi} \int_0^{2\pi} f(t)\, dt$.

Example

EXAMPLE 4 Find the nth-order Fourier approximation to the function $f(t) = t$ on the interval $[0, 2\pi]$.

SOLUTION Compute

$$\frac{a_0}{2} = \frac{1}{2}\cdot\frac{1}{\pi}\int_0^{2\pi} t\,dt = \frac{1}{2\pi}\left[\frac{1}{2}t^2\Big|_0^{2\pi}\right] = \pi$$

and for $k > 0$, using integration by parts,

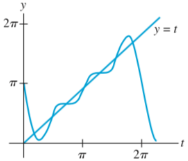
$$a_k = \frac{1}{\pi}\int_0^{2\pi} t\cos kt\,dt = \frac{1}{\pi}\left[\frac{1}{k^2}\cos kt + \frac{t}{k}\sin kt\right]_0^{2\pi} = 0$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} t\sin kt\,dt = \frac{1}{\pi}\left[\frac{1}{k^2}\sin kt - \frac{t}{k}\cos kt\right]_0^{2\pi} = -\frac{2}{k}$$

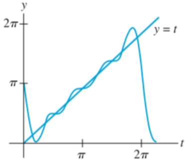
Thus the nth-order Fourier approximation of $f(t) = t$ is

$$\pi - 2\sin t - \sin 2t - \frac{2}{3}\sin 3t - \cdots - \frac{2}{n}\sin nt$$

Figure 3 shows the third- and fourth-order Fourier approximations of f. ■



(a) Third order



(b) Fourth order

- For $f \in C([0,2\pi])$, the **Fourier series expansion** of f is given by

$$f(t) = \tfrac{1}{2}a_0 + \sum_{m=1}^{\infty}(a_m \cos mt + b_m \sin mt),$$

  where a_m, b_m are defined as above.
- This series converges to f in the sense of norm convergence, namely,

$$\lim_{n\to\infty} \|f - \mathrm{proj}_{\mathrm{Span}(S_n)} f\| = 0.$$

Math 3108 - Fall 2019

Chapter 7: Symmetric Matrices and Quadratic Forms

- A real matrix is **symmetric** if $A = A^T$.
- In this section, we will show:

Theorem (Spectral Theorem for Symmetric Matrices)

Every symmetric matrix is diagonalizable. In fact, we can find an **orthonogonal basis** *of eigenvectors. This means*

$$A = PDP^T$$

for a real diagonal matrix D and an **orthogonal** *matrix P.*

- In fact, we will prove a **spectral theorem for normal matrices**, which means $AA^* = A^*A$. This includes symmetric matrices as a special case.

Motivation: Orthogonality of Eigenvectors

- Why might we expect the spectral theorem should be true?

Theorem (Theorem 1)

If A is a symmetric matrix, then any two eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof.

Let $\lambda_1, \boldsymbol{v}_1$ and $\lambda_2, \boldsymbol{v}_2$ be eigenvalue/eigenvector pairs with $\lambda_1 \neq \lambda_2$. Then

$$\boldsymbol{v}_1 \cdot A\boldsymbol{v}_2 = \lambda_2 \, \boldsymbol{v}_1 \cdot \boldsymbol{v}_2$$

while at the same time $\boldsymbol{v}_1 \cdot A\boldsymbol{v}_2 = A^T \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = \lambda_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2$. □

- This also holds for normal matrices (using Theorem 2 below).
- This does not solve the entire problem... it says nothing about diagonalizability in the first place.

Definitions

Recall the following:

- A real matrix is **symmetric** if $A = A^T$.
- A complex matrix is **hermitian** or **self-adjoint** if $A = A^*$.
- A real matrix is **orthogonal** if $P^T P = I$.
- A complex matrix is **unitary** if $P^* P = I$.

We introduce some new terminology, as well:

- A complex matrix is **normal** if $A^* A = AA^*$.

Example

Symmetric, hermitian, orthogonal, and unitary matrices are all normal. So are **skew-adjoint** matrices, which satisfy $A^* = -A$.

Main Result: Spectral Theorem

Our goal is the following:

Theorem (Spectral Theorem for Normal Matrices)

A matrix A is normal if and only if it is unitarily similar to a diagonal matrix, i.e.

$$A = PDP^*$$

for some diagonal matrix $D = diag\{\lambda_1, \ldots, \lambda_n\}$ and some unitary matrix $P = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_n]$.

In particular, we may write A as a sum of rank one orthogonal projections:

$$A = \lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^* + \cdots + \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^*.$$

- The final expression is called a **spectral decomposition** of A.
- This result will imply the spectral theorem for symmetric matrices.

Main Ingredients

- We need two main ingredients to prove the spectral theorem:

Theorem (Schur Factorization)

Any $A \in \mathbb{C}^{n\times n}$ can be written in the form $A = PUP^*$ for some unitary matrix P and some upper triangular matrix U.

Theorem (Theorem 2)

If A is normal and $\lambda, \boldsymbol{v}$ is an eigenvalue/eigenvector pair for A, then $\bar{\lambda}, \boldsymbol{v}$ is an eigenvector/eigenvalue pair for A^*.

- In particular, after we apply the Schur factorization to a normal matrix to write $A = PUP^*$, the second theorem will imply that U is actually diagonal.

Proof of Schur Factorization

Proof.

Suppose it holds for $(n-1)\times(n-1)$ matrices. Let A be $n\times n$.

Let $\lambda_1, \mathbf{v}_1$ be an eigenvalue/eigenvector pair for A with $\|\mathbf{v}_1\| = 1$.

Extend to an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and let $P_1 = [\mathbf{v}_1 \cdots \mathbf{v}_n]$.

Note P_1 is unitary and we can write $AP_1 = P_1 \begin{bmatrix} \lambda_1 & \mathbf{w} \\ \mathbf{0} & M \end{bmatrix}$.

Now write $M = QU_0Q^*$ with Q unitary, U upper triangular.
Define

$$P_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix}, \quad P = P_1P_2.$$

Then P is unitary and

$$P^*AP = P_2^* \begin{bmatrix} \lambda_1 & \mathbf{w} \\ \mathbf{0} & M \end{bmatrix} P_2 = \begin{bmatrix} \lambda & \mathbf{w}Q \\ \mathbf{0} & U_0 \end{bmatrix}.$$

□

Proof of Theorem 2

Theorem

If A is normal and $\lambda, \boldsymbol{v}$ is an eigenvalue/eigenvector pair for A, then $\bar{\lambda}, \boldsymbol{v}$ is an eigenvector/eigenvalue pair for A^*.

Proof.

For any $\lambda, \boldsymbol{v}$ and a normal matrix A,

$$\begin{aligned} \|(A-\lambda I)\boldsymbol{v}\|^2 &= [(A-\lambda I)\boldsymbol{v}]^*(A-\lambda I)\boldsymbol{v} \\ &= \boldsymbol{v}^*(A^*-\bar{\lambda} I)(A-\lambda I) \\ &= \boldsymbol{v}^*(A-\lambda I)(A^*-\bar{\lambda} I)\boldsymbol{v} \\ &= \|(A^*-\bar{\lambda} I)\boldsymbol{v}\|^2. \end{aligned}$$

$\square$

Proof of the Spectral Theorem

Proof of the Spectral Theorem.

We focus on showing normal implies unitarily diagonalizable.

Apply Schur factorization: $A = PUP^*$, $P = [\boldsymbol{v}_1, \ldots \boldsymbol{v}_n]$.

Write $U = [c_{ij}]$ and observe $A\boldsymbol{v}_1 = c_{11}\boldsymbol{v}_1$, and so $A^*\boldsymbol{v}_1 = \bar{c}_{11}\boldsymbol{v}_1$.

But $A^*P = PU^*$, so

$$\bar{c}_{11}\boldsymbol{v}_1 = \bar{c}_{11}\boldsymbol{v}_1 + \cdots + \bar{c}_{1n}\boldsymbol{v}_n \implies c_{1j} = 0, \quad j = 2, \ldots, n.$$

This shows

$$U = \begin{bmatrix} c_{11} & \mathbf{0} \\ \mathbf{0} & \tilde{U} \end{bmatrix}.$$

Now repeat the argument with $A\boldsymbol{v}_2 = c_{22}\boldsymbol{v}_2$...

It follows that U is diagonal. □

Spectral Theorem for Symmetric Matrices

- If A is a hermitian matrix (i.e. $A = A^*$), then it is normal and hence we can write $A = PDP^*$ with D diagonal and P unitary. But then

$$PDP^* = A = A^* = PD^*P^* \implies D = D^* \implies D \text{ is real},$$

so that hermitian matrices have **real** eigenvalues.

- Similarly, if A is real and symmetric (i.e. $A = A^T$), then we can write

$$A = PDP^T$$

where D is a real diagonal matrix and P is a real orthogonal matrix.

Spectral Theorem for Hermitian/Symmetric Matrices

Theorem (Spectral Theorem for Hermitian Matrices)

A matrix A is hermitian if and only if it can be factored as $A = PDP^*$ for a unitary matrix P and a real diagonal matrix D.

Theorem (Spectral Theorem for Symmetric Matrices)

A matrix A is symmetric if and only if it can be factored as $A = PDP^T$ for an orthogonal matrix P and a real diagonal matrix D.

Example

Orthogonally diagonalize the matrix

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

which has characteristic polynomial $-(\lambda - 7)^2(\lambda + 2)$.

Solution. Using the techniques of Chapter 5, we compute bases for the eigenspaces:

$$\lambda = 7 \implies \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}$$

and

$$\lambda = -2 \implies \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}.$$

Example (Continued)

Example (Continued)

Now apply Gram-Schmidt to find an orthogonal basis for Span$\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$: this yields

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}.$$

Finally, normalize each matrix to form an orthogonal matrix:

$$P = [\frac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|} \ \frac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|} \ \frac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|}].$$

Then $A = PDP^T$, with $D = \text{diag}\{7, 7, -2\}$.

- Many matrices that occur in applications are not square. For such matrices, there is an important notion related to eigenvalues and diagonalization, namely the **singular value decomposition**.

Definition (Singular Values)

Let A be an $n \times p$ matrix. The **singular values** of A are given by

$$\sigma_j := \sqrt{\lambda_j}, \quad j = 1, \ldots, p,$$

where $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues of the $p \times p$ matrix A^*A.

- As A^*A is hermitian, it has real eigenvalues. If $\lambda, \boldsymbol{v}$ is an eigenvalue/eigenvector pair,

$$\lambda = \tfrac{1}{\|\boldsymbol{v}\|^2}\boldsymbol{v}^*(A^*A)\boldsymbol{v} = \tfrac{\|A\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2} \geq 0.$$

Singular Values, Singular Vectors

- In the following, we fix an $n \times p$ matrix A. Then A^*A has an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ corresponding to eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$.
- The singular values are given by $\sigma_j = \sqrt{\lambda_j}$.
- If $\mathrm{Rank}(A) = r$, then $\sigma_{r+1} = \cdots = \sigma_p = 0$.
- $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_p\}$ is an orthonormal basis for $\mathrm{Nul}(A^*A) = \mathrm{Nul}(A)$.
- $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for $\mathrm{Col}(A^*) = [\mathrm{Nul}(A)]^\perp$.

Singular Values, Singular Vectors

Lemma

The vectors

$$\boldsymbol{u}_j = \frac{1}{\sigma_j} A\boldsymbol{v}_j, \quad j = 1, \ldots, r$$

form an orthonormal basis for $Col(A)$.

Proof.

By the basis theorem, it is sufficient to check orthonormality:

$$\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \frac{1}{\sigma_i \sigma_j} \boldsymbol{v}_i^* (A^* A \boldsymbol{v}_j) = \frac{\lambda_j}{\sigma_i \sigma_j} \boldsymbol{v}_i^* \boldsymbol{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

□

- Now let $\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n\}$ be any orthonormal basis for $[\text{Col}(A)]^\perp$.
- Define $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ (the **left singular vectors**) and $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p]$ (the **right singular vectors**), both of which are unitary.
- By construction:

$$AV = U\Sigma, \quad \Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad D = \text{diag}\{\sigma_1, \ldots, \sigma_r\}.$$

- For the first r columns we use $A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$.
- For the remaining columns, use $\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_p\}$ belong to $\text{Nul}(A)$.

Singular Value Decomposition

Theorem (Singular Value Decomposition)

For any $n \times p$ matrix A with rank r, there exists a decomposition

$$A = U\Sigma V^*,$$

where

- *U is an $n \times n$ unitary matrix,*
- *V is a $p \times p$ unitary matrix,*
- *Σ is an $n \times p$ with the form*

$$\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad D = diag\{\sigma_1, \ldots, \sigma_r\},$$

where D is an $r \times r$ block diagonal matrix in the upper left corner containing the nonzero singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ of A.

Example

Example

Let us describe the process of finding a singular value decomposition of a real matrix $A \in \mathbb{R}^{n \times p}$.

- Diagonalize $A^T A$ with an *orthonormal* basis of eigenvectors.
- Build the matrices V and Σ.
- Construct the first r columns of U (where $r = \text{Rank}(A)$).
- If $r < p$, build the remaining columns of U by finding an orthonormal basis for $[\text{Col}(A)]^\perp = \text{Nul}(A^T)$.
    - This requires finding a basis for the null space of A^T and then possibly applying the Gram–Schmidt algorithm and normalization.

A few applications...

- If A is an invertible $n \times n$ matrix, the ratio σ_1/σ_n is called the **condition number** of A, which is related to the sensitivity of the solution to $A\boldsymbol{x} = \boldsymbol{b}$ to changes/errors in the entries of A.
- Since orthogonal matrices in $\mathbb{R}^{2\times 2}$ represent rotations/reflections of the plane, applying the singular value decomposition to a matrix transformation of the plane reveals that every such transformation is the composition of three transformations: rotation/reflection, scaling, and rotation/reflection.
- In terms of numerical analysis, singular value decomposition is generally faster and more accurate than eigenvalue decomposition. In particular, SVD is prevalent in many modern applications.

Section 7.5 - Applications to Image Processing and Statistics

- Suppose we have a $p \times n$ matrix of data, say $A = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_n]$.
- The **sample mean** is defined to be $\frac{1}{N}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$, and for simplicity we assume we have normalized the data to have mean zero.
- The **covariance matrix** of A is defined by the $p \times p$ matrix

$$S = \frac{1}{n-1} AA^T.$$

- The diagonal entries of S represent the variance of the coordinates x_i of data vectors $\boldsymbol{X}$; the total variance is the sum of the diagonal entries (called the **trace of S**).
- The off-diagonal entries s_{ij} of S represent the **covariance** of x_i and x_j. We call x_i and x_j **uncorrelated** if $s_{ij} = 0$.

Principal Component Analysis

- **Goal.** Find an **orthogonal** $p \times p$ matrix P that determines a change of variables $\boldsymbol{X} = P\boldsymbol{Y}$ such that the variables y_j are uncorrelated and arranged in order of decreasing variance.
- Our data matrix is transformed to $B = P^T A$, which has covariance matrix

$$\tfrac{1}{n-1}BB^T = \tfrac{1}{n-1}(P^T A)(P^T A)^T = \tfrac{1}{n-1}P^T AA^T P.$$

- In particular, our problem is equivalent to orthogonal diagonalization of AA^T (which is connected to the singular value decomposition of the transpose A^T of the data matrix).
- Arranging the eigenvalues in decreasing order, the corresponding unit eigenvectors are called the **principal components** of the data.
- The new variables represent the directions of maximal variance (after projecting away from the previous directions).

Dimensionality Reduction

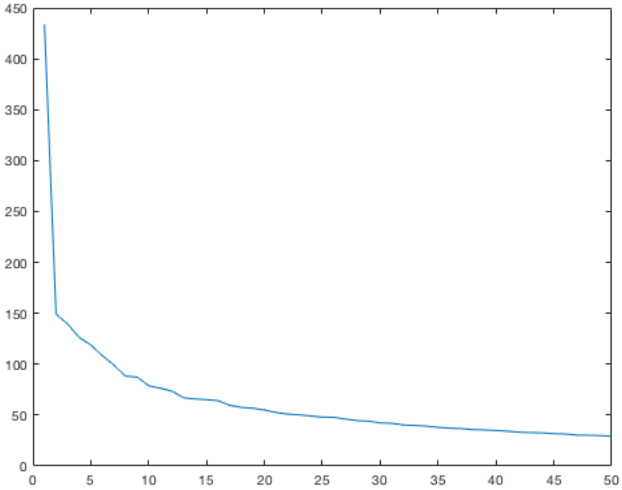
- Orthogonal changes of variables do not change the total variance of the data.
- In many cases, one finds that nearly all of the variance is captured in eigenvalues corresponding to the first few principal components.
- This allows us to find low-dimensional approximations to high-dimensional data! Extremely useful for data analysis, data interpretation, data compression... and on and on.

Example

Example

- Download $n = 5000$ images of handwritten digits from the MNIST database.
- Each image is represented by a vector in $\mathbb{R}^p$, where $p = 28 \times 28 = 784$.
- This gives the $p \times n$ data matrix A.
- Perform SVD on A^T, giving
$$A^T = U\Sigma V^T.$$
- The singular values drop off very quickly.

Dropoff of Singular Values



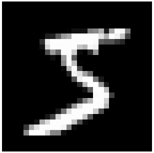
Example (Continued)

Example (Continued)

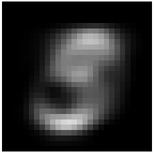
- Let's make a 5-dimensional approximation to this data.
- Given a vector $\boldsymbol{X}$ in our data set, we have the new representation $\boldsymbol{X} = V\boldsymbol{Y}$, i.e. $\boldsymbol{Y} = V^T\boldsymbol{X}$.
- We keep only the first 5 entries of $\boldsymbol{Y}$ and set the rest to zero; call this $\boldsymbol{Y}_{app}$.
- Then we apply V to get the approximation $\boldsymbol{X} \sim V\boldsymbol{Y}_{app}$.

Examples

True Image:

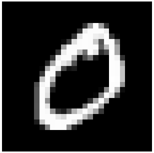


Approximation using first 5 principal components:

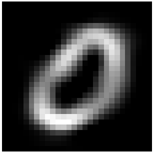


Examples

True Image:

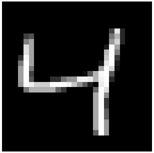


Approximation using first 5 principal components:

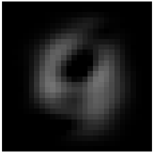


Examples

True Image:

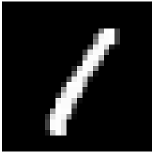


Approximation using first 5 principal components:

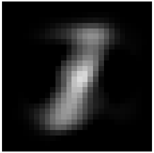


Examples

True Image:

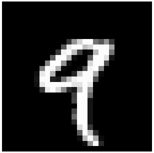


Approximation using first 5 principal components:

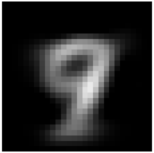


Examples

True Image:

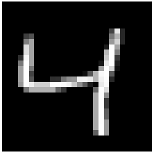


Approximation using first 5 principal components:

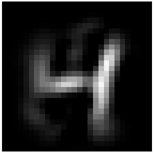


The third image looked pretty bad...

True Image:



Approximation using first 20 principal components:



Some Final Remarks

- We used 5000 samples of all different digits. This would have been much more accurate if they had all been the same digit.
- This provides a very crude method for image compression.
- This type of analysis forms the basis for many modern techniques in machine learning, data analysis, compression, etc.

Thanks!

Thanks for a great semester!