Notes on Linear Algebra (MA371)

Yu Cao

Spring 2025

Contents

1 Vectors and Matrices

Linear algebra is the study of vectors, matrices, and how they interact.

1.1 Vectors

Vectors in $\mathbf{R}^2$:

Two basic operations on vectors of $\mathbf{R}^2$:

- Multiplying a vector v by a number (scalar multiplication):

- Adding two vectors v and w in $\mathbf{R}^2$ (vector addition):

Linear combination of two vectors:

Two key questions for $\mathbf{R}^2$: Given two vectors v and w in $\mathbf{R}^2$,

- Describe all the combinations

- Find a specific combination

Vectors in $\mathbf{R}^m$:

$$\mathbf{R}^m = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \mid a_1, a_2, \ldots, a_m \in \mathbf{R} \right\}$$

Each element of $\mathbf{R}^m$ is called a vector. The m numbers $a_1, a_2, \ldots, a_m$ are called the components of the vector.

Note: we use the "column vector" notation for vectors. For convenience, we ofter write $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ as $\begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix}^T$, or simply as $(a_1, a_2, \ldots, a_m)$. Which form to use depends on the context.

Scalar multiplication and vector addition in $\mathbf{R}^m$:

$$c \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_m \end{bmatrix}, \qquad \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_m + b_m \end{bmatrix}$$

These two operations can be combined to define **linear combinations** of vectors:

For $v_1, v_2, \ldots, v_n$ in $\mathbf{R}^m$ and $c_1, c_2, \ldots, c_n$ in $\mathbf{R}$, the vector

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

is called a linear combination of $v_1, v_2, \ldots, v_n$.

1.2 Matrices

An $m \times n$ matrix A is a rectangular array of numbers with m rows and n columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

We can think of A as a collection of n column vectors in $\mathbf{R}^m$:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

where $a_1, a_2, \ldots, a_n$ are the columns of A.

Multiplying a matrix by a scalar:

Matrix addition (subtraction):

1.2.1 Matrix-vector multiplication

Dot Product Way:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 7 + 2 \cdot 8 \\ 3 \cdot 7 + 4 \cdot 8 \\ 5 \cdot 7 + 6 \cdot 8 \end{bmatrix} = \begin{bmatrix} 23 \\ 53 \\ 83 \end{bmatrix}$$

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

Column Way:

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

In general, Ax is a linear combination of the columns of A with coefficients given by the components of x.

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

1.2.2 Matrix-matrix multiplication

Dot Product Way:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} \text{row } 1 \cdot \text{col } 1 & \text{row } 1 \cdot \text{col } 2 \\ \text{row } 2 \cdot \text{col } 1 & \text{row } 2 \cdot \text{col } 2 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Column Way: To find

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

one can first calculate

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 19 \\ 43 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 22 \\ 50 \end{bmatrix}$$

Thus, the product is

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Row Way: To find

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

First calculate:

$$\begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix} = \begin{bmatrix}19 & 22\end{bmatrix}, \quad \begin{bmatrix}3 & 4\end{bmatrix}\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix} = \begin{bmatrix}43 & 50\end{bmatrix}$$

So the product is

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix} = \begin{bmatrix}19 & 22\\43 & 50\end{bmatrix}$$

Column-times-Row Way: This way allows us to express AB as a sum of rank 1 matrices.

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

More examples of matrix-matrix multiplication

Find the product (do it!), in four different ways, of AB and BA where $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$

1.2.3 Properties of matrix multiplication

- Matrix multiplication is **not commutative**: AB and BA are not necessarily equal.
- Matrix multiplication is **associative**: $(AB)C = A(BC)$.
- Matrix multiplication is **distributive** over addition: $A(B+C) = AB+AC$ and $(A+B)C = AC+BC$.
- Scalar multiplication is **compatible** with matrix multiplication: $(kA)B = A(kB) = k(AB)$.

1.2.4 Inverse of (Square) Matrix

Definition 1.1. *Let A be an $n \times n$ matrix. If there exists an $n \times n$ matrix B such that*

$$AB = BA = I_n$$

then B is called the ***inverse*** *of A, denoted by A^{-1}.*

Property 1. Suppose A is an $n \times n$ matrix. Then $I_n A = AI_n = A$.

Property 2. Suppose A and B are $n \times n$ matrices and invertible. Then $(AB)^{-1} =$

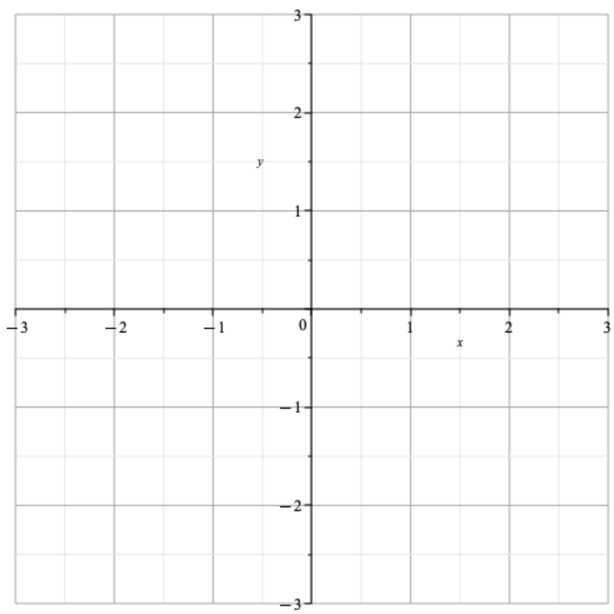
1.3 Geometry of linear equations in $\mathbf{R}^2$

Consider the system of linear equations:

$$\begin{cases} -x + 2y = 3 \\ \quad 2x - y = 0 \end{cases}$$

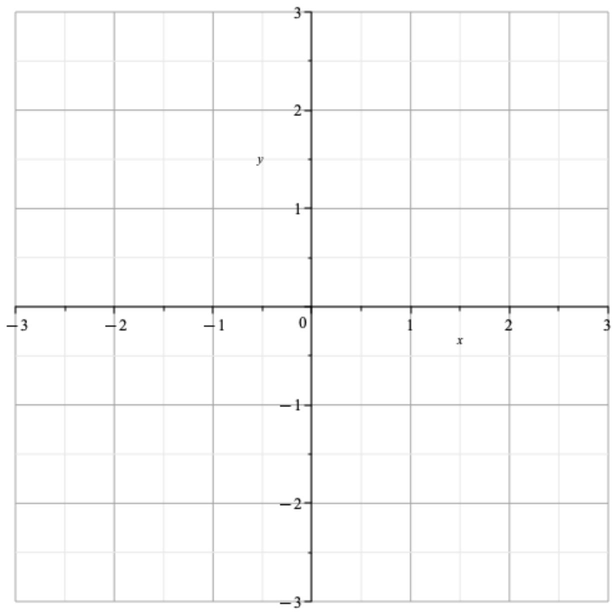
There are three ways in geometry to think about this system of equations.

"Row Picture":



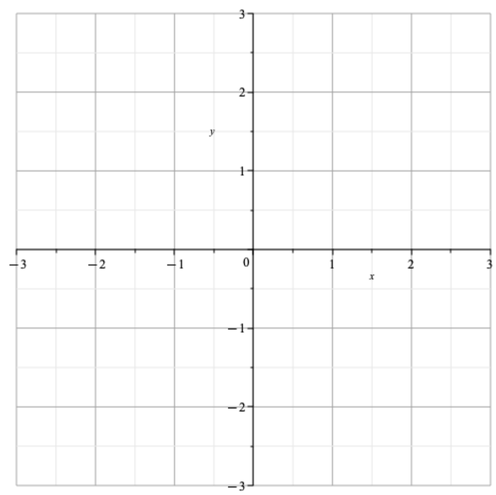
$$\begin{cases} -x + 2y = 3 \\ \ \ 2x - y = 0 \end{cases}$$

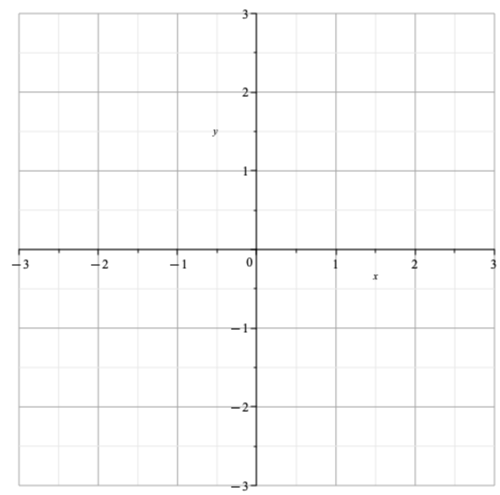
"Column Picture":



$$\begin{cases} -x + 2y = 3 \\ \ \ 2x - y = 0 \end{cases}$$

Linear transformation picture





2 LU Decomposition

2.1 Gaussian Elimination Revisit

Key concepts:

- Row Echelon Form (REF)
- Reduced Row Echelon Form (RREF)
- Elementary Row Operations
- Gaussian Elimination
- Gauss-Jordan Elimination

A matrix is in **Row Echelon Form** if

- All rows having only zero entries, if exist, are at the bottom.
- The leading entry (that is, the left-most nonzero entry) of every nonzero row, called the pivot, is on the right of the leading entry of every row above.

Example 2.1. *The following matrices are in row echelon form (REF):*

$$\begin{bmatrix} 3 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

A matrix is in **Reduced Row Echelon Form (RREF)** if it satisfies the following conditions:

- It is in the row echelon form.
- The leading entry (i.e. pivot) in each nonzero row is 1.
- Each column containing a leading 1 has zeros in all entries above the leading 1.

Example 2.2. *REF vs RREF*

REF but not RREF:

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

RREF:

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

Elementary Row Operations:

There are three elementary row operations that we can perform on a matrix when solving a system of linear equations (equivalently, its augmented matrix). These operations are:

- (Row switching) Exchange any two rows. $R_i \leftrightarrow R_j$
- (Scalar Multiplication) Multiply any row by a **nonzero** scalar. $kR_i \to R_i$ $(k \neq 0)$
- (Row Addition) Add a multiple of one row to another row. $R_i + kR_j \to R_i$

Steps of Gaussian Elimination:

1. **Start with the Full Matrix:** Focus on the leftmost column that hasn't been processed yet.
2. **Find the Pivot:**
   - Locate the first non-zero entry in this column. This entry is called the **pivot**.
   - If the pivot is not already at the top, swap rows to bring it to the top of this column.
3. **Create Zeros Below the Pivot:**
   - Use the pivot row to create zeros in all entries below the pivot.
   - Subtract multiples of the pivot row from the rows beneath it to achieve this.
4. **Move to the Next Column:**
   - Ignore the rows that have already been processed.
   - Focus on the submatrix that starts from the row directly below the current pivot.
   - Repeat the process: find the next pivot, swap if necessary, and create zeros below it.
5. **Continue Until No More Rows or Columns Remain:**
   - Keep repeating the process of finding pivots and making zeros below them until you run out of rows or reach a column that is entirely zero.
   - When this happens, the matrix is in **Row Echelon Form (REF)**.
6. **End:** The process stops when all the remaining rows contain only zeros, or there are no more columns to process.

The process of transforming a matrix into RREF is called **Gauss-Jordan elimination**:

1. Get REF: Use Gaussian elimination to transform the matrix into Row Echelon Form (REF).
2. Make each pivot 1 by row operations: For each pivot (leading entry in a row), divide the row by the pivot value to make it 1.
3. Make all entries above each pivot 1 zero: Start from the last pivot and work your way upwards, using row operations to make all values above each pivot equal to zero.

Note: Steps 2 and 3 are equivalent to the back substitution step in the context of solving systems of equations.

Example 2.3. *Find the RREF of the matrix:*

$$\begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix}$$

Example 2.4. *Solve the linear equation* $Ax = b$ *where* 3×3 *matrix* $A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix}$ *and* $b = \begin{bmatrix} 19 \\ 55 \\ 50 \end{bmatrix}$.

2.2 Gaussian Elimination with elementary matrices

How do we solve $Ax = b$ using Gaussian elimination?

Goal: Intepret Gaussian elimination as a sequence of matrix multiplications. (You should review the Row way of matrix-matrix multiplication.)

2.2.1 Elementary matrices

Philosophy: Instead of performing row operations on A, we can multiply A by a sequence of elementary matrices to achieve the same effect.

Elimination matrices: The $n \times n$ elementary matrix $E_{ij}(\ell)$ denotes the matrix obtained by adding $\ell(\neq 0)$ times row j to row i in the identity matrix.

Permutation matrices: The $n \times n$ elementary matrix P_{ij} denotes the matrix obtained by swapping rows i and j in the identity matrix.

Basic properties of elementary matrices:

$E_{ij}(\ell)^{-1} = E_{ij}(-\ell)$

$P_{ij}^{-1} = P_{ij}$

2.2.2 LU factorization of a matrix

Consider the matrix $A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix}$.

2.2.3 Solving linear systems with LU factorization

Example 2.5. *Consider the linear equation $Ax = b$ where 3×3 matrix $A = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix}$ and $b = \begin{bmatrix} 19 \\ 55 \\ 50 \end{bmatrix}$.*

2.2.4 Further remarks on LU factorization

Remark 1: We only do LU factorization for square matrices.

Example 2.6. *Find the LU factorization of the matrix* $A = \begin{bmatrix} 3 & 1 & 0 \\ -3 & 1 & 1 \\ 6 & 8 & 4 \end{bmatrix}$.

Remark 2: The LU factorization is not always possible.

For example: $A := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has no LU factorization.

The factorization $A = LU$ is possible when there is no zeros in the pivots (thus no row exchanges are needed).

3 Vector space $\mathbf{R}^m$ and its subspaces

Modern mathematics is built on the concept of a **set**. A set is a collection of objects. The objects in a set are called the **elements** of the set.

When we consider the set of all the vectors in $\mathbf{R}^m$ together with the two fundamental operations: vector addition and scalar multiplication, we obtain a **vector space**:

3.1 Basic definitions

Definition 3.1. $\mathbf{R}^m$ *together with the operations of vector addition and scalar multiplication is called a* ***vector space****.*

Definition 3.2 (Subspace of $\mathbf{R}^m$)**.** *A subset V of $\mathbf{R}^m$ is called a **subspace** of $\mathbf{R}^m$ if it satisfies the following conditions:*

- *The zero vector $\vec{0}$ is in V.*
- *V is closed under addition: if u and v are in V, then $u + v$ is in V.*
- *V is closed under scalar multiplication: if u is in V and c is a real number, then cu is in V.*

A subspace V of $\mathbf{R}^m$ is also called a vector space.

A particular way to describe a subspace is to use the *span* of a set of vectors:

Definition 3.3 (Span of a set of vectors)**.** *Let* $S = \{v_1, v_2, \ldots, v_n\}$ *be a set of vectors in* $\mathbf{R}^m$*. The* ***span*** *of* S*, denoted by* $\operatorname{span}(S)$*, is the set of all linear combinations of the vectors in* S*:*

$$\operatorname{span}(\{v_1, v_2, \ldots, v_n\}) := \{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : c_1, c_2, \ldots, c_n \in \mathbf{R}\}$$

Theorem 3.1.
The span of a set of vectors in $\mathbf{R}^m$ *is a subspace of* $\mathbf{R}^m$*.*

Definition 3.4. *Let A be an $m \times n$ matrix. The* ***column space*** *of A, denoted by $C(A)$, is the span of the columns of A.*

3.2 Linear independence, basis, and dimension

In the previous section, we have seen that the span of a set of vectors is a subspace. In a spanning set, some vectors may be redundant. Consider the following example.

Example 3.1. *Let* $u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *and* $v = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. *Then* $\operatorname{span}\{u, v\} = \operatorname{span}\{u\}$.

When there is no "redundancy" in a spanning set, we say that the set is linearly independent:

Definition 3.5. *Let V be a vector space and $S = \{v_1, v_2, \ldots, v_n\}$ be a set of vectors in V. The set S is called* ***linearly independent*** *if the only solution to the equation*

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = \vec{0}$$

is $x_1 = x_2 = \cdots = x_n = 0$.

The empty set of vectors, $\emptyset$, is also declared to be linearly independent.

Definition 3.6. *If S is not linearly independent, we say that it is linearly dependent.*

In other words, a set of vectors $S = \{v_1, v_2, \ldots, v_n\}$ is linearly dependent if there exist scalars $c_1, c_2, \ldots, c_n$, not all zero, such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = \vec{0}.$$

Why this condition prevents redundancy?

- If some vector in a set can be written as a linear combination of the others, then it is redundant, i.e., the set is linearly dependent.

Example 3.2. *Let V be a vector space and $w \in V$ a nonzero vector. Prove that the set $\{w\}$ is linearly independent.*

Example 3.3. *Prove that the set of the columns of the matrix* $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix}$ *is linearly independent.*

Note. We are building up knowledge of linear algebra from the ground up. Whenever we <u>prove</u> a statement, we are ONLY allowed to use the definitions and theorems that we have learned so far. In your homework and exams, you should only use the definitions and theorems that have been covered in the course up to that point too.

Example 3.4. *Prove that the set of the columns of the matrix* $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 6 & 0 & 6 \end{bmatrix}$ *is linearly dependent.*

In the previous subsection, we discussed spanning sets and linearly independent sets. We now bring these two concepts together by considering sets that have both properties.

Definition 3.7. *A basis B for a vector space V is a set of vectors in V that*

(1) B spans V, i.e., $\mathrm{span}(B) = V$*;*

(2) B is linearly independent.

Example 3.5. *The set* $B = \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$ *is a basis of* $\mathbf{R}^2$*, called the standard basis of* $\mathbf{R}^2$*.*

Example 3.6. *The set* $B = \{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}\}$ *is a basis of* $\mathbf{R}^2$*.*

Example 3.7. *The set* $B = \{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\}$ *is a basis of the vector space* $V = \{\begin{bmatrix} x \\ x \\ y \end{bmatrix} : x, y \in \mathbf{R}\}$*.*

Example 3.8. *The set* $B = \{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\}$ *is a basis of the vector space* $V = \{\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbf{R}^3 : x + y + z = 0\}$*.*

Vectors in a basis are building blocks for all vectors in the vector space:

Theorem 3.2.
*Let V be a vector space and $B = \{v_1, v_2, \ldots, v_n\}$ be a basis for V. Then every vector $v \in V$ can be written **uniquely** as a linear combination of the vectors in B:*

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

for some scalars $c_1, c_2, \ldots, c_n$.

Theorem 3.3 (Dimension Theorem).
All bases for a vector space V have the same number of vectors.

Definition 3.8. *The **dimension** of a vector space V, denoted by* $\dim(V)$*, is the number of vectors in a basis for V.*

Definition 3.9. *For a matrix A,* $\dim(C(A))$ *is called the **rank** of A, denoted by* $\text{rank}(A)$.

Retrieval practice:

Can you recall the **precise definitions** of the following concepts?

- What is a vector space?
- What is a subspace of a vector space?
- What is the span of a set of vectors?
- What is the column space of a matrix?
- What is a linearly independent set of vectors?
- What is a basis for a vector space?
- What is the dimension of a vector space?
- etc.

What is your *intuition* about these concepts?

Can you provide **examples** of each of these concepts?

How are these concepts **related** to the concepts in the previous chapters? How are they related to each other?

What theorems have we learned so far? What is the precise statement of the theorem? $(*)$

Can you provide examples to illustrate the theorem?

3.3 Four fundamental subspaces associated with a matrix

Given an $m \times n$ matrix A, we can define four subspaces associated with A:

Definition 3.10. *Let A be an $m \times n$ matrix. The **column space** of A, denoted by $C(A)$, is the span of the columns of A.*

Definition 3.11. *Let A be an $m \times n$ matrix. The **null space** of A, denoted by $N(A)$, is the set of all solutions to the homogeneous equation $Ax = \vec{0}$.*

Definition 3.12. *Let A be an $m \times n$ matrix. The **row space** of A, denoted by $C(A^T)$, is the column space of the transpose of A.*

Definition 3.13. *Let A be an $m \times n$ matrix. The **left null space** of A, denoted by $N(A^T)$, is the null space of the transpose of A.*

3.3.1 Finding bases for the four fundamental subspaces of a matrix

Example 3.9. *Find a basis for each of the four fundamental subspaces associated with the matrix*

$$A = \begin{bmatrix} 1 & 7 & 0 & 8 \\ 1 & 7 & 1 & 17 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$A = \begin{bmatrix} 1 & 7 & 0 & 8 \\ 1 & 7 & 1 & 17 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Summary of what we have learned so far for the fundamental subspaces of a matrix A.

To find a basis for $C(A)$: we find $R = \text{ref}(A)$, or $R_0 = \text{rref}(A)$, and the pivot columns of A form a basis for $C(A)$.

To find a basis for $N(A)$: we find $R = \text{ref}(A)$, or $R_0 = \text{rref}(A)$, and the "special solutions" to $Ax = \vec{0}$ form a basis for $N(A)$.

Observations:

- The number of pivot columns in A is the same as the number of pivot columns in R. So the dimension of $C(A)$ is the same as the dimension of $C(R)$, but $C(A) \neq C(R)$ in general.
- $Ax = \vec{0}$ iff $Rx = \vec{0}$. So $N(A) = N(R)$ and in particular, the dimension of $N(A)$ is the same as the dimension of $N(R)$.
- $\dim C(A) + \dim N(A) =$ the number of columns of A.

Remark. So far, we have focused on $C(A)$ and $N(A)$. We can find bases for the row space $C(A^T)$ and left null space $N(A^T)$ by applying the same procedures to the matrix A^T.

However, we can also find bases for $C(A^T)$ and $N(A^T)$ by directly using $R = \text{ref}(A)$ or $R_0 = \text{rref}(A)$. We will come back to this later.

3.3.2 Relation between the column space and the null space of a matrix

Theorem 3.4 (Rank-Nullity Theorem)**.**
Suppose A is an $m \times n$ matrix. Then $\text{rank}(A) + \text{nullity}(A) = n$. *In other words:*

$$\dim(C(A)) + \dim(N(A)) = n.$$

3.4 Structure of solutions to linear equation $Ax = b$

Suppose A is an $m \times n$ matrix. Then $Ax = b$ has a solution if and only if $b \in C(A)$.

Theorem 3.5.
If x_p is a particular solution to $Ax = b$, then any solution to $Ax = b$ can be written as $x = x_p + x_n$ for some $x_n \in N(A)$.

If x_p is a particular solution to $Ax = b$, and $x_n \in N(A)$. Then $x = x_p + x_n$ is also a solution to $Ax = b$.

Because of the structure theorem, we call $x = x_p + x_n$, where x_n may have unspecified parameters, the "complete solution" to $Ax = b$.

Example 3.10. *Find the complete solution to* $Ax = b$ *where* $A = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix}$ *and* $b = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix}$.

Example 3.11. *Find the complete solution to* $Ax = b$ *where* $A = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix}$ *and* $b = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

Observation: $Ax = b$ has a unique solution if and only if $b \in C(A)$ and $N(A) = \{\vec{0}\}$.

3.5 Dimensions of the Four Subspaces

In the previous sections, we studied the column space and null space of a matrix. We now look at the row space and left null space.

3.5.1 The row space and the left null space

Consider the matrix $A = \begin{bmatrix} 1 & 7 & 0 & 8 \\ 1 & 7 & 1 & 17 \\ 1 & 7 & 0 & 8 \end{bmatrix}$. We will find $\dim C(A^T)$ and $\dim N(A^T)$.

Theorem 3.6 (Row-Column Rank Theorem).
Suppose A is an $m \times n$ matrix. Then

$$\dim(C(A)) = \dim(C(A^T)) = \text{rank}(A).$$

Fundamental Theorem of Linear Algebra Part 1

Given an $m \times n$ matrix A, the four fundamental subspaces and their dimensions are summarized as follows:

Suppose $r = \text{rank}(A)$.

3.5.2 Additional exercises

If $m > n$, then m vectors in $\mathbf{R}^n$ are always linearly dependent.

Suppose $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are non-zero vectors in $\mathbf{R}^7$. They span a subspace of $\mathbf{R}^7$. What are the possible dimensions of that vector space?

This question is about an m by n matrix A for which

$$Ax = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ has no solutions and } Ax = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ has exactly one solution.}$$

1. Give all possible information about m and n and the rank r of A.
2. Find all solutions to $Ax = \vec{0}$ and explain your answer.
3. Write down an example of a matrix A that fits the description in part 1.

3.6 Intuition of span and linear independence/dependence

4 Orthogonality

In the previous chapter, we have learned about the four fundamental subspaces associated with a matrix A. The dimensions of these subspaces and their relations are summarized in the Fundamental Theorem of Linear Algebra (Part 1).

We have al seen the structure of solutions to the linear equation $Ax = b$ when it is solvable.

In this chapter, we will study the eqution $Ax = b$ when it is not solvable. (machine learning, AI, etc.)

4.1 Orthogonality of vectors

The **dot product**, also known as the **inner product**, of two vectors $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ in $\mathbf{R}^n$ is defined as

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

If we consider x, y as two column vectors, then $x \cdot y = x^T y$.

The **norm** of a vector $x \in \mathbf{R}^n$ is defined as

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

If $u \in \mathbf{R}^n$ and $\|u\| = 1$, then u is called a **unit vector**.

If $x \cdot y = 0$, then we say that x and y are **orthogonal**.

Example 4.1. *Suppose $u = (1, 2)$, $v = (3, -4)$ and $z = (-2, 1)$. Find $u \cdot v$, $\|v\|$. Are z and v orthogonal?*

The following are some basic properties of the dot product:

For any $x, y \in \mathbf{R}^n$ and $c \in \mathbf{R}$,

- $x \cdot y = y \cdot x$.
- $(x + y) \cdot z = x \cdot z + y \cdot z$.
- $(cx) \cdot y = x \cdot (cy) = c(x \cdot y)$.
- $x \cdot x = \|x\|^2$.

With the basic properties, we can easily prove the following:

Theorem 4.1 (Pythagorean Theorem)**.**
Suppose $x, y \in \mathbf{R}^m$*. If* $x \cdot y = 0$*, then* $\|x+y\|^2 = \|x\|^2 + \|y\|^2$*.*

We also have the following simple observation:

Proposition 4.1.
If $v \neq \vec{0}$*, then* $u = \dfrac{v}{\|v\|}$ *is a unit vector.*

4.1.1 Geometric meaning of the dot product in $\mathbf{R}^2$

For two nonzero vector $v, w \in \mathbf{R}^2$, let $\theta \in [0, \pi]$ be the angle between u and v. Then

$$v \cdot w = \|v\| \cdot \|w\| \cos\theta.$$

Remark. This formula shows that $v \cdot w = 0$ if and only if $\theta = \frac{\pi}{2}$, i.e., v and w are orthogonal.

This formula follows from the Law of Cosines and properties of the dot product. (See also p.14 of the textbook.)

4.2 Orthogonality of subspaces

Definition 4.1. *Two subspaces V and W of $\mathbf{R}^n$ are **orthogonal** if $v^T w = 0$ for every $v \in V$ and $w \in W$.*

Example 4.2. *For any matrix A, the nullspace of A is orthogonal to the row space of A.*

(Two proofs.)

Example 4.3 (**Important**). *There is another pair of fundamental subspaces: the column space $C(A)$ and the left null space $N(A^T)$ are orthogonal. Those are perpendicular inside m-dimensional space $\mathbf{R}^m$.*

Example 4.4. *(p.145)*

There is a very important restriction on the dimensions of any two orthogonal subspaces:

If V and W are orthogonal subspaces in $\mathbf{R}^n$, then $\dim(V) + \dim(W) \leq n$.

(Proof with bases. Omitted.)

Theorem 4.2.
If V and W are orthogonal subspaces in $\mathbf{R}^n$, then $V \cap W = \{\vec{0}\}$.

Two orthogonal subspaces that account for all of $\mathbf{R}^n$ are called **complementary subspaces**.

Definition 4.2. *The orthogonal complement $V^{\perp}$ of V contains all vectors orthogonal to V:*

$$V^{\perp} = \{x \in \mathbf{R}^n : x \cdot v = 0 \textit{ for all } v \in V\}.$$

The two pairs of subspaces in the big picture of linear algebra are actually orthogonal complements:

Fundamental Theorem of Linear Algebra Part 2.

If A is an $m \times n$ matrix, then

- $N(A)$ is the orthogonal complement of the row space $C(A^T)$ in $\mathbf{R}^n$.
- $N(A^T)$ is the orthogonal complement of the column space $C(A)$ in $\mathbf{R}^m$.

Both parts of the Fundamental Theorem of Linear Algebra are summarized in the following picture:

The Matrix is everywhere. It is all around us. Even now, in this very room. You can see it when you look out your window or when you turn on your television. You can feel it when you go to work... when you go to church... when you pay your taxes. It is the world that has been pulled over your eyes to blind you from the truth.

4.3 Projection of vectors

In $\mathbf{R}^2$

Definition 4.3. *Suppose $a, b \in \mathbf{R}^n$ and $a \neq \vec{0}$. The projection of b onto the line spanned by a is defined as*

$$p = proj_a(b) = \left(\frac{a^T b}{a^T a}\right) a.$$

Example 4.5. *Project $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find $proj_a(b)$.*

Now comes the projection matrix. In the formula for p, what matrix is multiplying b?

Example 4.6. *Find the projection matrix* $P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}$ *onto the line through* $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

Remark. Projecting a second time does not change anything, so $P^2 = P$.

4.3.1 Two great inequalities

With projections, we can prove one of the most important inequalities in mathematics:

Theorem 4.3 (Cauchy-Schwarz inequality)**.**
For $u, v \in \mathbf{R}^n$*, we have* $|u \cdot v| \leq \|u\|\|v\|$.

Using the Cauchy-Schwarz inequality, we can also prove the following:

Theorem 4.4 (Triangle inequality)**.**
For $u, v \in \mathbf{R}^n$*, we have* $\|u + v\| \leq \|u\| + \|v\|$.

This is left as a homework exercise.

4.4 Projections onto subspaces

Now we move beyond projection onto a line. Projecting onto an n-dimensional subspace of $\mathbf{R}^m$ is a generalization of projection onto a line.

Start with n vectors $a_1, \cdots, a_n \in \mathbf{R}^m$. Assume that these a's are linearly independent.

Problem: Find the combination $p = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$ closest to a given vector b.

In terms of matrices, we rewrite the problem as follows:

Given $m \times n$ matrix $A = [a_1 \cdots a_n]$ and $b \in \mathbf{R}^m$, find $p = A\hat{x}$ closest to b, i.e., minimize the quantity $\|b - Ax\|^2$.

The key is again in the geometry! We get intuition from $\mathbf{R}^2$ and $\mathbf{R}^3$.

The combination $p = \widehat{x}_1 a_1 + \cdots + \widehat{x}_n a_n = A\widehat{x}$ that is closest to b comes from $\widehat{x}$:

Find $\widehat{x}$ $\quad (n \times 1)$: $\qquad A^{\mathrm{T}}(\mathrm{b} - A\widehat{x}) = \vec{0} \quad$ or $\quad A^{\mathrm{T}} A\widehat{x} = A^{\mathrm{T}}\mathrm{b}$.

This symmetric matrix $A^{\mathrm{T}}A$ is n by n. It is invertible if the a 's are independent. The solution is $\widehat{\boldsymbol{x}} = \left(A^{\mathrm{T}}A\right)^{-1} A^{\mathrm{T}}b$.

The projection of $\boldsymbol{b}$ onto the subspace is p :

Find p $\quad (m \times 1)$: $\qquad p = A\widehat{x} = A\left(A^{\mathrm{T}}A\right)^{-1} A^{\mathrm{T}}\ \mathrm{b}$

The projection matrix P multiplying b:

Find P $\quad (m \times m)$: $\qquad P = A\left(A^{\mathrm{T}}A\right)^{-1} A^{\mathrm{T}}$

Example 4.7. *If* $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ *and* $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ *find* $\widehat{x}$, p *and* P.

4.4.1 Two extreme cases

If $b \in C(A)$,

If $b \perp C(A)$,

4.4.2 General definition of projection matrices

Definition 4.4. *A square matrix P is called a **projection matrix** if $P^2 = P$.*

Remark: If P is a projection matrix, then $I - P$ is also a projection matrix.

4.4.3 Back to $A^T A$

Theorem 4.5.
The square matrix $A^T A$ is invertible if and only if the columns of A are linearly independent.

We will use the following fact, which we will prove later when we focus more on square matrices and wrap things up regarding equivalent descriptions of invertible matrices.

Theorem 4.6.
A square matrix is invertible if and only if its nullspace is trivial, i.e., $\{\vec{0}\}$.

Proof of Theorem 4.5:

4.5 Least Squares Problems

It often happens that $Ax = b$ has no solution: b may not be in the column space of A.

In the previous section, we find a vector p that is closest to b in the column space of A. (This is the projection of b onto the column space of A.)

Note in finding p, we are also able to answer the following question:

> Least squares problem: What is the "best" vector x such that the squared error $\|b - Ax\|^2$ is minimized?

Note: When $Ax = b$ has a solution, denoted as $\widehat{x}$, then the minimized error is

The previous section emphasized the projection p. This section emphasizes $\widehat{x}$, the least squares solution:

> **Definition 4.5.** *The <u>least squares solution</u> to the equation $Ax = b$ is the vector $\widehat{x}$ that minimizes the error $\|b - Ax\|^2$.*

Example 4.8 (Baby Machine Learning)**.** *A crucial application of least squares is fitting a straight line to m points. Consider the problem with three points: Find the "closest" line to the points* $(0, 6), (1, 0)$*, and* $(2, 0)$.

Why $\widehat{x}$ minimizes the error?

By geometry.

By algebra.

By calculus. Most functions are minimized by calculus!

In this example:

$$E = \|b - Ax\|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1 - 0)^2 + (C + D \cdot 2 - 0)^2$$

where the unknowns variables are C and D. At the minimum, the partial derivatives of E with respect to C and D are both zero:

$$\partial E/\partial C = 2(C + D \cdot 0 - 6)(1) + 2(C + D \cdot 1)(1) + 2(C + D \cdot 2)(1) = 0$$
$$\partial E/\partial D = 2(C + D \cdot 0 - 6)(0) + 2(C + D \cdot 1)(1) + 2(C + D \cdot 2)(2) = 0$$

which is simplified to:

$$3C + 3D = 6$$
$$3C + 5D = 0$$

This process can be generalized when there are more than $m > 2$ "data points": $(t_1, b_1), \cdots, (t_m, b_m)$.

The best line $y = C + Dt$ misses the points by vertical distances $e_i = b_i - (C + Dt_i)$.

No line is perfect, and the least squares line minimizes $E = e_1^2 + \cdots + e_m^2$.

We want to find the "least squares solution" to $Ax = b$:

$$Ax = b \quad \text{is} \quad \begin{array}{c} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \vdots \\ C + Dt_m = b_m \end{array} \quad \text{with} \quad A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \qquad x = \begin{bmatrix} C \\ D \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Remark: the column space of A is so thin that almost certainly b is outside of it. When b happens to lie in the column space, all the "data points" happen to lie on the line; in this case, $p = b$. Then $Ax = b$ is solvable and the errors are $e = \vec{0}$.

In general:

Solve $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$ for $x = \widehat{x} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$.

Example 4.9. *Suppose we take measurements at the* 21 *equally spaced times* $t = -10, -9, \ldots, 9, 10$. *All measurements are* $b_i = 0$ *except that* $b_{11} = 1$ *at the middle time* $t = 0$.

(a) Using least squares, what are the best $\widehat{C}$ and $\widehat{D}$ to fit those 21 points by a straight line $y = C + Dt$?

(b) You are projecting the vector b onto what subspace? (Give a basis.) Find a nonzero vector perpendicular to that subspace.

(c) What is the minimized squared error $E = \sum e_i^2$ where $e_i = b_i - (\widehat{C} + \widehat{D}t_i)$?

Remark: A has orthogonal columns when the measurement times t_i add to zero. In this case, $A^T A$ is diagonal and the least squares solution is easy to find.

Theorem: If $v_1, \cdots, v_n \in \mathbf{R}^m$ are orthogonal nonzero vectors, then they are linearly independent.

4.6 Orthonormal vectors

Orthogonal vectors are very nice to work with. Suppose A is an $m \times n$ matrix with orthogonal columns $a_1, \cdots, a_n$. Then the matrix $A^T A$ is very simple!

If the orthogonal vectors are unit vectors, it is even better.

Definition 4.6. *The set of vectors* $\{q_1, \cdots, q_n\}$ *is called* ***orthonormal*** *if*

$$q_i^{\mathrm{T}} q_j = \begin{cases} 0 & \textit{when } i \neq j \quad \textit{(orthogonal vectors)} \\ 1 & \textit{when } i = j \quad \textit{(unit vectors: } \|q_i\| = 1) \end{cases}$$

Remark. Let $Q = [q_1 \cdots q_n]$. The columns of Q are orthonormal if and only if

Note: Q is not required to be square; QQ^T is not necessarily I.

Theorem 4.7.
When Q is square, $Q^T Q = I$ implies that Q is invertible, i.e., $Q^T Q = QQ^T = I$ and $Q^{-1} = Q^T$.

Vocabulary: A SQUARE matrix Q is called an **orthogonal matrix** if $Q^T Q = I$.

Example 4.10 (Rotation in $\mathbf{R}^2$)**.** *Q rotates every vector in the plane by the angle θ:*

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad Q^{-1} = Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Example 4.11 (Permutation)**.** *Permutation matrices have a 1 in every row and a 1 in every column.*

These matrices change the order to (y, z, x) and (y, x):

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

The inverse of a permutation matrix is its transpose: $Q^{-1} = Q^T$. The inverse puts the components back into their original order:

"Inverse = transpose":

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Theorem 4.8.
If Q has orthonormal columns $\left(Q^{\mathrm{T}}Q = I\right)$, it leaves lengths unchanged :

Same length for Qx: $\quad \|Qx\| = \|x\|$ *for every vector x.*

Remark. If Q has orthonormal columns, then the least squares solution to $Qx = b$ is $\widehat{x} = Q^T b$. The projection of b onto the column space of Q is $p = QQ^T b = Pb$.

4.7 The Gram-Schmidt process

Start with three linearly independent vectors a, b, c. We want to find an *orthonormal basis* for the subspace spanned by a, b, c.

Big picture:

We first construct three orthogonal vectors A, B, C from a, b, c

and then we divide A, B, C by their lengths to get three orthonormal vectors $q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}, q_3 = \frac{C}{\|C\|}$.

Gram-Schmidt:

(i) $A = a$.

(ii) $B = b - \dfrac{A^T b}{A^T A} A$.

(iii) $C = c - \dfrac{A^T c}{A^T A} A - \dfrac{B^T c}{B^T B} B$.

Finally, let

$$q_1 = \frac{A}{\|A\|}, \quad q_2 = \frac{B}{\|B\|}, \quad q_3 = \frac{C}{\|C\|}.$$

Example 4.12. *Apply Gram-Schmidt to the following vectors:*

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}, \quad c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}.$$

Remark. The Gram-Schmidt process can be used for any number of (independent) vectors, not just three.

Remark. The orthonormal vectors q_1, q_2, q_3 can be used in each step of the process.

Modified Gram-Schmidt process: Given a, b, c:

$$
\begin{aligned}
A &= a & q_1 &= \frac{A}{\|A\|} \\
B &= b - (q_1^T b)q_1 & q_2 &= \frac{B}{\|B\|} \\
C &= c - (q_1^T c)q_1 - (q_2^T c)q_2 & q_3 &= \frac{C}{\|C\|}
\end{aligned}
$$

4.8 Projections using orthonormal bases

Recap: Suppose $a_1, \cdots, a_n$ are linearly independent vectors. Given a vector b, we can find the projection of b onto the subspace spanned by $a_1, \cdots, a_n$.

We can also find the projection of b onto the subspace spanned by $q_1, \cdots, q_n$, which are orthonormal vectors obtained from $a_1, \cdots, a_n$ using the Gram-Schmidt.

Remark: Decomposition of the projection p

Example 4.13 (An old example). *Find the projection of* $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ *onto the column space of the matrix*

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Use the Gram-Schmidt process to find an orthonormal basis for $C(A)$ first. And then use the orthonormal basis to find the projection of b onto $C(A)$.

4.9 QR factorization

We started with a matrix A with linearly independent columns: $a_1, \cdots, a_n$. We ended with a matrix Q with orthonormal columns: $q_1, \cdots, q_n$. How are A and Q related?

Back to Gram-Schmidt

Observation. At every step $a_1, \cdots, a_k$ are combinations of $q_1, \cdots q_k$. Later q's are not involved:

Given a, b, c:

$$A = a \qquad\qquad q_1 = \frac{A}{\|A\|}$$

$$B = b - (q_1^T b)q_1 \qquad\qquad q_2 = \frac{B}{\|B\|}$$

$$C = c - (q_1^T c)q_1 - (q_2^T c)q_2 \qquad\qquad q_3 = \frac{C}{\|C\|}$$

(**QR factorization**) From independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, Gram-Schmidt constructs orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. The matrices with these columns satisfy $\boldsymbol{A} = \boldsymbol{QR}$. Then $\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}$ is upper triangular because later $\boldsymbol{q}$'s are orthogonal to earlier $\boldsymbol{a}$'s.

Example 4.14. *Find th QR factorization of the matrix:* $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix}$

Follow-up question: What is the QR factorization of the following matrix A?

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

Remark: Least squares with QR

Suppose A is an $m \times n$ matrix with linearly independent columns. The least squares solution to $Ax = b$ is given by the normal equation $A^T Ax = A^T b$.

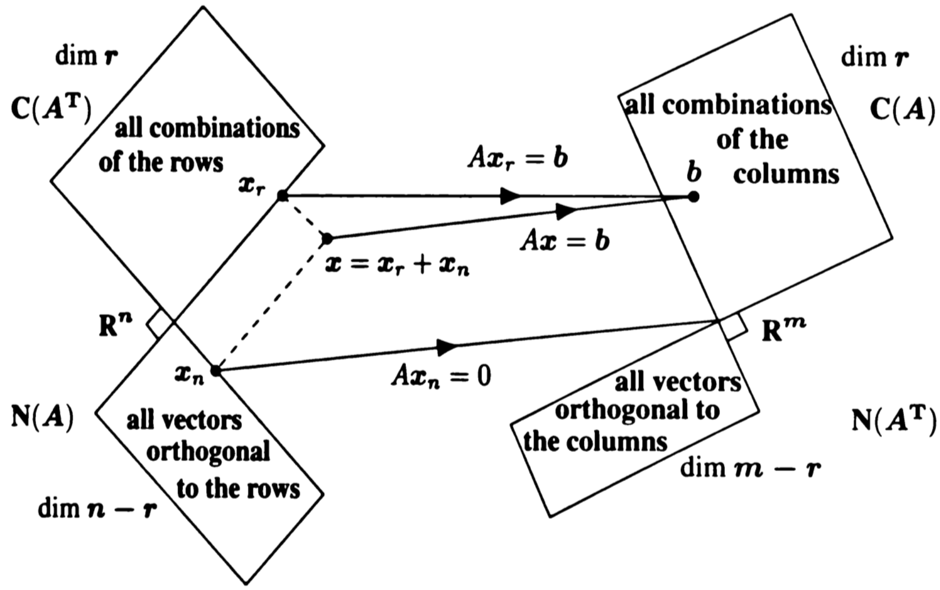
If A has orthonormal columns,

If A has orthogonal columns,

If A does not have orthogonal columns, we can still use the QR factorization to solve the least squares problem.

If $A = QR$ is the QR factorization of A, then the normal equation becomes

4.10 Big pictures



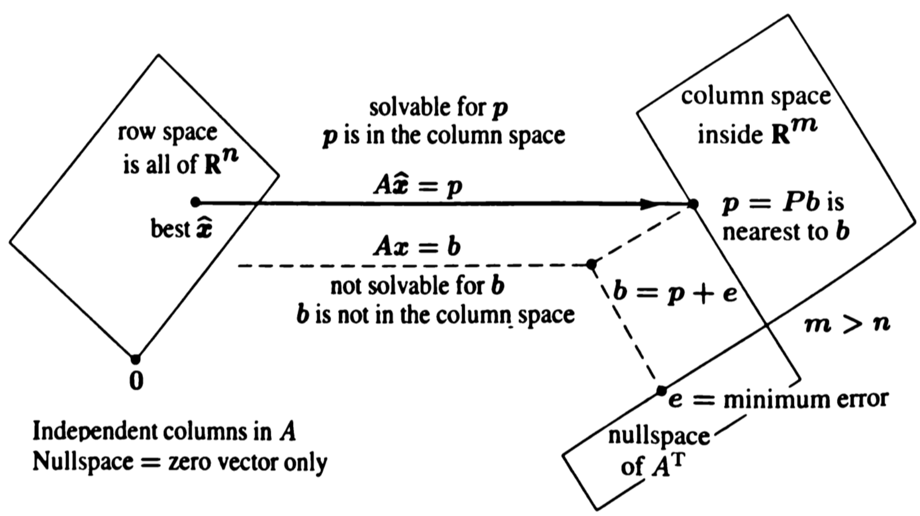


Figure 4.7: The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$.

5 Determinants

Roughly speaking, the determinant is a number associated with a *square* matrix. It is denoted by $\det(A)$ or $|A|$. This number contains a lot of information about the matrix A. For example, it can tell us whether A is invertible. More importantly, it can tell us how the matrix transforms volumes in space.

For the moment, **forget** everything you think you "know" about determinants. We will define it from scratch.

5.1 Axioms of determinants

We will define the determinant of a square matrix A by three axioms. These axioms are not derived from anything else; they are just *assumed* to be true. Together, they define what a determinant is.

"It's not who you are underneath, it's what you ***do*** *that defines you."*

— Batman

Three axioms of determinants.

The determinant det is a function that takes a square matrix A and returns a number $\det(A)$. The function det satisfies the following three axioms:

Axiom 1. $\det(I) = 1$.

Axiom 2. Swapping two rows multiplies det by -1.

Axiom 3. The determinant function det is linear in each row separately:

$$\det(\ldots, r_i + r_i', \ldots) = \det(\ldots, r_i, \ldots) + \det(\ldots, r_i', \ldots),$$

and similarly for scalar multiples:

$$\det(\ldots, \lambda r_i, \ldots) = \lambda \det(\ldots, r_i, \ldots).$$

As a combination, this axiom says that

$$\det(\ldots, \lambda_1 r_1 + \lambda_2 r_2, \ldots) = \lambda_1 \det(\ldots, r_1, \ldots) + \lambda_2 \det(\ldots, r_2, \ldots).$$

Example 5.1. *Using the three axioms, compute the determinant of the* 2×2 *matrix:*

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

Question: How are these axioms related to the 2×2 determinant formula I learned in high school?

Question: It feels like you just flipped the game. I could have "proved" these axioms starting from the 2×2 formula. So why do we need to start from axioms?

Answer: What you learned in high school is called the Leibniz formula; more on this later. It works well for 2×2 or 3×3 matrices, but it becomes "messy" and impractical for larger sizes. The axioms, on the other hand, are much easier to work with—and they let us prove all the key properties cleanly and systematically.

5.2 Leibniz formula

Similar to the 2×2 determinant:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

we have a formula for the 3×3 determinant. These formulas are called the **Leibniz formula**.

Repeatedly applying axioms A1-A3, we can show that

$$\begin{aligned}
\begin{vmatrix} a & b & c \\ p & q & r \\ x & y & z \end{vmatrix} &= \begin{vmatrix} a & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & z \end{vmatrix} + \begin{vmatrix} a & 0 & 0 \\ 0 & 0 & r \\ 0 & y & 0 \end{vmatrix} + \begin{vmatrix} 0 & b & 0 \\ p & 0 & 0 \\ 0 & 0 & z \end{vmatrix} \\
&+ \begin{vmatrix} 0 & b & 0 \\ 0 & 0 & r \\ x & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ p & 0 & 0 \\ 0 & y & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ 0 & q & 0 \\ x & 0 & 0 \end{vmatrix} \\
&= aqz \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} + ary \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} + bpz \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} \\
&+ brx \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} + cpy \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} + cqx \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} \\
&= aqz + brx + cpy - ary - bpz - cqx\,.
\end{aligned}$$

There are $3! = 6$ terms in the formula since there are $3!$ permutation matrices. The signs of the terms are determined by the determinant of the permutation matrix.

The formula is summarized as the Sarrus Rule:

$$\left|\begin{array}{ccc|cc} a & b & c & a & b \\ p & q & r & p & q \\ x & y & z & x & y \end{array}\right.$$

For $n \times n$ matrices in general, the Leibniz formula is more complicated:

$$\det(A) = \text{Sum of } n! \text{ terms}\,.$$

Each term is the determinant of a permutation matrix times the product of the entries of A corresponding to that permutation.

Remark. The Leibniz formula is not practical for large matrices.

5.3 Cofactor expansions

In this section, we give a recursive formula for the determinant of a matrix. This formula is called the **cofactor expansion**.

The starting point of the cofactor expansion is the determinant of a 1×1 matrix:

$$\det[a] = a.$$

To describe the cofactor expansions, wee need the following definition:

Let A be an $n \times n$ matrix.

1. The (i, j) **minor**, denoted A_{ij}, is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i th row and the j th column.

2. The (i, j) **cofactor** C_{ij} is defined in terms of the minor by

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

Example 5.2. *For the matrix*

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

compute A_{23} and C_{23}.

Theorem 5.1 (Cofactor expansion)**.**

Let A be an $n \times n$ matrix with entries a_{ij}.

1. For any $i = 1, 2, \ldots, n$, we have

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} = a_{i1} C_{i1} + a_{i2} C_{i2} + \cdots + a_{in} C_{in}$$

This is called cofactor expansion along the ith row.

2. For any $j = 1, 2, \ldots, n$, we have

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} = a_{1j} C_{1j} + a_{2j} C_{2j} + \cdots + a_{nj} C_{nj}$$

This is called cofactor expansion along the jth column.

Example 5.3. *Compute the determinant of the following matrix using cofactor expansion along the first row:*

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

5.4 Properties of determinants

The Leibniz formula and the cofactor expansion are two ways to compute the determinant of a matrix. Overall, the formulas are useful for small matrices. However, they are not the only ways.

There are various properties of determinants that can be used to compute them. These properties are all consequences of the three axioms of determinants.

1. If two rows of a matrix are equal, then its determinant is zero.
2. Adding or subtracting a multiple of one row to another row does not change the determinant.
3. If a matrix has a row of zeros, then its determinant is zero.
4. The determinant of a (upper or lower) triangular matrix is the product of its diagonal entries.
5. The determinant of a product of matrices is the product of their determinants:

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

6. A square matrix A is invertible if and only if its determinant is nonzero.
7. The determinant of the inverse of a matrix is equal to the reciprocal of the determinant of the original matrix:

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

8. The determinant of the transpose of a matrix is equal to the determinant of the original matrix:

$$\det(A^T) = \det(A).$$

Proposition 5.1.
If two rows of a matrix are equal, then its determinant is zero.

Proof. Suppose A has two equal rows. Swap those two rows to get the same matrix A. By Axiom 2,

$$\det A \;=\; -\det A,$$

hence $\det A = 0$. $\square$

Example 5.4. *Compute* $\det(A)$ *where* $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$.

Proposition 5.2.
Adding or subtracting a multiple of one row to another row does not change the determinant.

Proof. Let A' be obtained from A by replacing row i with $\mathrm{row}_i + \lambda\,\mathrm{row}_j$, $j \neq i$. By multilinearity (Axiom 3) in row i,

$$\det A' = \det\big(\ldots, \mathrm{row}_i + \lambda\,\mathrm{row}_j, \ldots\big) = \det\big(\ldots, \mathrm{row}_i, \ldots\big) + \lambda\,\det\big(\ldots, \mathrm{row}_j, \ldots\big).$$

But in the second term the matrix has two identical rows, so that determinant is zero by the previous proposition. Thus $\det A' = \det A$. $\square$

Example 5.5. *The following two matrices have the same determinant:*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 19 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 19 \end{bmatrix}$$

Proposition 5.3.
If a matrix has a row of zeros, then its determinant is zero.

Proof. If row i of A is the zero vector, then by linearity in that row (Axiom 3),

$$\det\big[\dots,\vec{0},\dots\big] = \det\big[\dots,0\cdot v,\dots\big] = 0\cdot\det\big[\dots,v,\dots\big] = 0.$$

$\square$

Example 5.6. *Compute* $\det(A)$ *where* $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 0 & 0 \end{bmatrix}$.

Proposition 5.4.
The determinant of a (upper or lower) triangular matrix is the product of its diagonal entries.

Idea: Use cofactor expansion along the first row or column.

Example 5.7. *Compute the determinant of the following matrices:*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{bmatrix}$$

Proposition 5.5.
For any two $n \times n$ matrices A and B,

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

Proposition 5.6.
A square matrix A is invertible iff $\det(A) \neq 0$.

Proposition 5.7.
If A is a square matrix and A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.

Proposition 5.8.
If A is a square matrix, then $\det(A^T) = \det(A)$.

Proof. Suppose A has LU factorization $A = LU$, where L is lower triangular and U is upper triangular. Then

$$A^T = (LU)^T = U^T L^T.$$

Thus

$$\det(A^T) = \det(U^T L^T) = \det(L^T)\ \det(U^T) = \det(L)\ \det(U) = \det(A).$$

In general, there exists a permutation matrix P such that $PA = LU$. (This is called LU factorization with pivoting.) Then

$$\det(P)\det(A) = \det(L)\det(U) = \det(U).$$

On the other hand,

$$A^T P^T = U^T L^T$$

and thus

$$\det(A^T)\det(P^T) = \det(U^T)\det(L^T) = \det(U)\det(L) = \det(U).$$

But $\det(P^T) = \det(P)$ (why?), so we have

$$\det(A^T)\det(P) = \det(P)\det(A).$$

□

5.5 Cramer's rule and inverse matrix

Cramer's idea for solving $Ax = b$ when $\det(A) \neq 0$:

Replace the first column of I by x. This triangular M_1 has determinant x_1. When you multiply by A, the first column becomes Ax which is b. The other columns of B_1 are copied from A.

$$\begin{array}{c}\text{Key idea}\\ AM_1 = B_1\end{array} \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1$$

Take determinants of the three matrices to find x_1 :

$$\text{Product rule} \quad (\det A)\,(x_1) = \det B_1 \quad \text{or} \quad x_1 = \det B_1 / \det A.$$

This is the first component of x in Cramer's Rule. Changing a column of A gave B_1. To find x_2 and B_2, put the vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ into the second columns of I and A :

$$\begin{array}{c}\text{Same idea}\\ AM_2 = B_2\end{array} \begin{bmatrix} & & \\ a_1 & a_2 & a_3 \\ & & \end{bmatrix} \begin{bmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{bmatrix} = \begin{bmatrix} & & \\ a_1 & b & a_3 \\ & & \end{bmatrix} = B_2.$$

Take determinants to find $(\det A)\,(x_2) = \det B_2$. This gives $\boldsymbol{x_2} = (\det \boldsymbol{B_2})\,/(\det \boldsymbol{A})$.

Example 5.8. *Solve the system of equations with Cramer's idea:*

$$\begin{aligned} 3x_1 + 4x_2 &= 2 \\ 5x_1 + 6x_2 &= 4 \end{aligned}$$

Theorem 5.2 (Cramer's Rule)**.**
If $\det A$ *is not zero,* $A\boldsymbol{x} = \boldsymbol{b}$ *is solved by determinants:*

$$x_1 = \frac{\det B_1}{\det A} \quad x_2 = \frac{\det B_2}{\det A} \quad \dots \quad x_n = \frac{\det B_n}{\det A}$$

The matrix B_j *has the* j *th column of* A *replaced by the vector* b*.*

Consider the linear system

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Assume $a_1b_2 - b_1a_2$ is nonzero. Then, with the help of determinants, x and y can be found with Cramer's rule as

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} = \frac{c_1b_2 - b_1c_2}{a_1b_2 - b_1a_2}, \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} = \frac{a_1c_2 - c_1a_2}{a_1b_2 - b_1a_2}.$$

The rules for 3×3 matrices are similar. Given

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}.$$

Then the values of x, y and z can be found as follows:

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}, \text{ and } z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

Remark: Cramer's rule is ____________ practical for large matrices. It is mainly of theoretical interest.

Theorem 5.3.
Let A be an $n \times n$ matrix. If A is invertible, then the inverse of A can be computed by

$$A^{-1} = \frac{C^T}{\det(A)},$$

where C^T is the transpose of the cofactor matrix C of A.

*The matrix C^T is called the **adjugate** of A, denoted* $\operatorname{adj}(A)$.

Idea. We describe the idea of the proof in the 3×3 case. The proof is similar for larger matrices.

Consider the matrix equation $AX = I$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

6 Eigenvalues and Eigenvectors

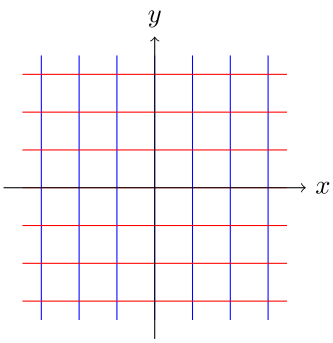
6.1 Introduction

Matrices as Linear Transformations

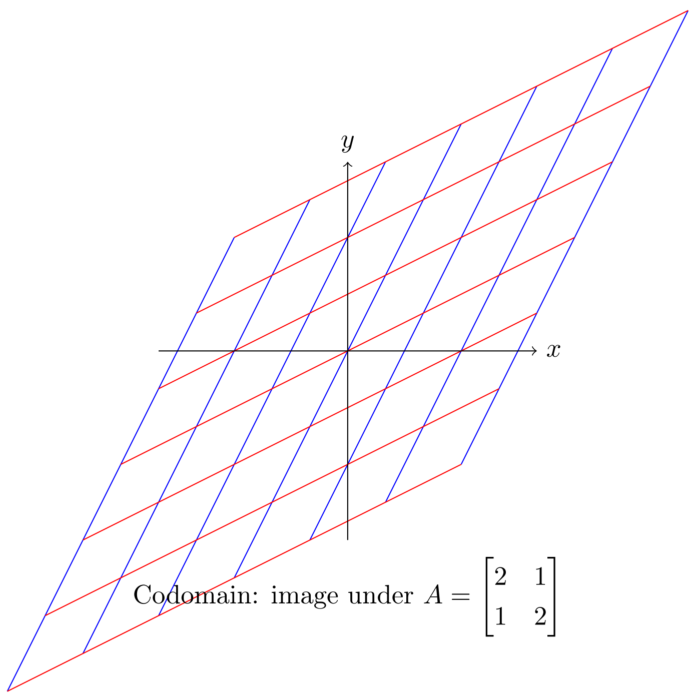
An $m \times n$ matrix A defines a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$: $L_A(x) = Ax$. The function L_A has the following properties:

- $L_A(x + y) = L_A(x) + L_A(y)$ for all $x, y \in \mathbf{R}^n$.
- $L_A(cx) = cL_A(x)$ for all $x \in \mathbf{R}^n$ and all scalars $c \in \mathbf{R}$.

Particularly, an $n \times n$ square matrix A defines a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$.

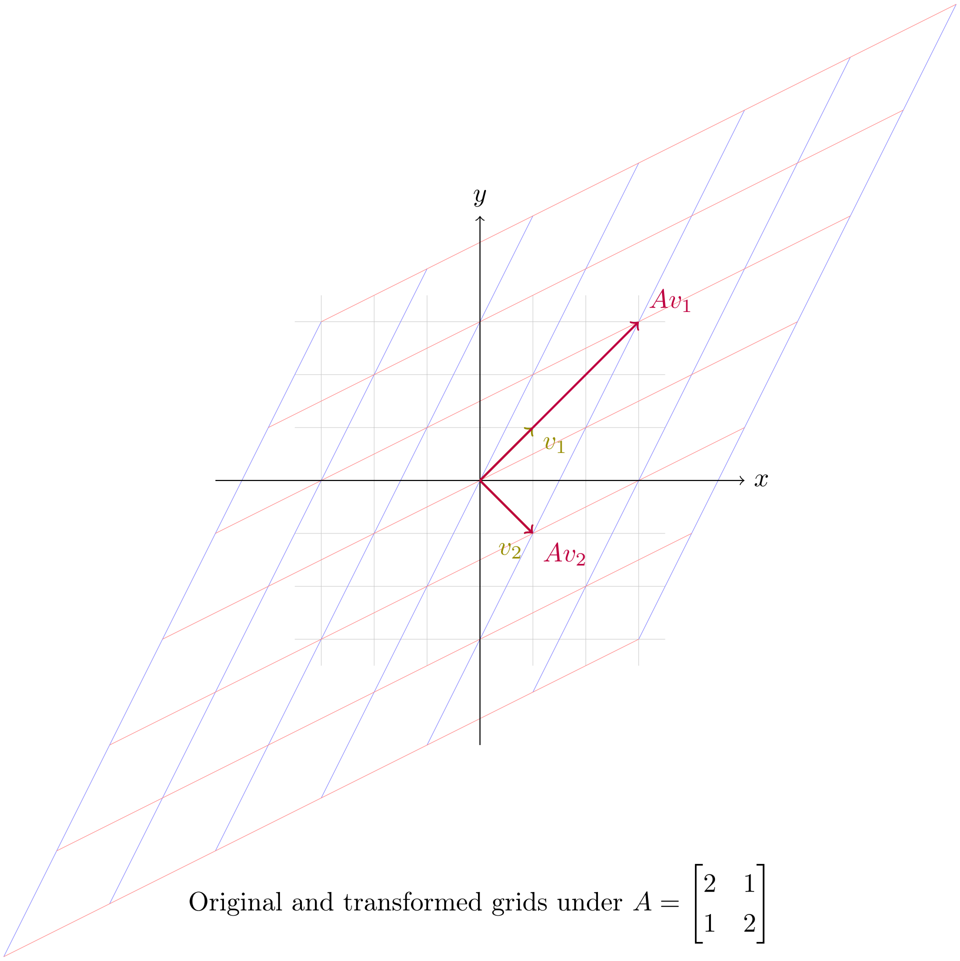


Domain: original grid



Codomain: image under $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Given a square matrix A, the transformation $x \mapsto Ax$ moves vectors in a variety of directions. It often happens that there are special vectors on which the action of A is quite simple.



Original and transformed grids under $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Eigenvalues and eigenvectors are a new way to see into the heart of a matrix.

Definition 6.1. *Let A be an $n \times n$ matrix. If $Ax = \lambda x$ for some scalar λ and some nonzero vector x in $\mathbf{R}^n$, then*

- *λ is called an* eigenvalue *of A, and*
- *x is called an* eigenvector *of A corresponding to λ.*

(λ, x) is called an eigenpair *of A.*

The term eigenvalue comes from the German word eigen, meaning "own" or "proper," reflecting values intrinsic to a transformation; it was introduced by Hilbert's school in the early 20th century.

Remark: if x is an eigenvector of A corresponding to λ, then cx is also an eigenvector of A corresponding to λ for any nonzero scalar c.

Example 6.1. *A square matrix A is singular if and only if 0 is an eigenvalue of A.*

Example 6.2. *For a projection matrix P, if $x \in C(P)$, then $Px = x$. If $x \in N(P)$, then $Px = 0x$. So the eigenvalues of a projection matrix are 0 and 1.*

Example 6.3. *The exchange matrix* $E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *has eigenvalues* 1 *and* -1*.*

Example 6.4 (Markov matrix)**.** *If each columns of a matrix* A *adds to* 1*. Then* $\lambda = 1$ *is an eigenvalue of* A*. For example,* $\lambda = 1$ *is an eigenvalue of the matrix* $A = \begin{bmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{bmatrix}$.

The Equation for the Eigenvalues

Eigenvalues: The number λ is an eigenvalue of A if and only if $A - \lambda I$ is singular.

The eigenvectors corresponding to λ make up $N(A - \lambda I) \setminus \{\vec{0}\}$.

Summary. To find eigenvalues and eigenvectors of a square matrix A:

1. Compute $\det(A - \lambda I)$
2. Find the roots of this polynomial by solving $\det(A - \lambda I) = 0$.
3. For each eigenvalue λ, solve the system $(A - \lambda I)x = \vec{0}$.

Example 6.5. *Find the eigenvalues and eigenvectors of the matrix* $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

6.2 Diagonalizing a matrix

The diagonal matrices are the "simplest" matrices.

Example 6.6. *Consider the diagonal matrix*

$$D = \operatorname{diag}(2, -1, 3) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Find the eigenvalues and eigenvectors of D.

Observation. If $D = \operatorname{diag}(d_1, d_2, \cdots, d_n)$,

then the eigenvalues of D are ________________

and the eigenvectors of D are

Example 6.7. *Compute the following expressions:*

$$\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}^{100}$$

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}^{100}$$

Observation. If $D = \operatorname{diag}(d_1, d_2, \cdots, d_n)$, then $D^k = \operatorname{diag}(d_1^k, d_2^k, \cdots, d_n^k)$ for any positive integer k.

In general, we can **diagonalize** a matrix A if A has n linearly independent eigenvectors.

Diagonalization. Suppose the $n \times n$ matrix A has n linearly independent eigenvectors $x_1, \cdots, x_n$. Put those x_i into the columns of an eigenvector matrix X and put the corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$ into a diagonal matrix Λ, the eigenvalue matrix. Then we have $AX = X\Lambda$.

Since the matrix X is invertible (why?),

$AX = X\Lambda \quad \text{is} \quad \boldsymbol{X^{-1}AX = \Lambda} \quad \text{or} \quad \boldsymbol{A = X\Lambda X^{-1}}.$

Definition 6.2. *Let A be an $n \times n$ matrix. When A has n linearly independent eigenvectors, we say that A is diagonalizable, i.e., there exists an invertible matrix X and a diagonal matrix Λ such that $A = X\Lambda X^{-1}$. We say that A is diagonalized by the matrix X.*

When A is diagonalizable, we can compute the kth power A^k easily:

Example 6.8. *Consider the matrix $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. Compute A^{100}.*

Remark 1. A and Λ have the same eigenvalues. The eigenvectors are different.

Remark 2. The eigenvalues in Λ come in the same order as the eigenvectors in X.

Example 6.9. *Consider the matrix* $A = \begin{bmatrix} 2 & 4 \\ 0 & 6 \end{bmatrix}$. *It has eigenvalues* $\lambda_1 = 2$ *and* $\lambda_2 = 6$ *and corresponding eigenvectors* $x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Remark 3. Some matrices have too few linearly independent eigenvectors. Those matrices cannot be diagonalized.

Example 6.10. *Consider the following matrix:*

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

Remark 4. There is no connection between invertibility and diagonalizability:

Invertibility is concerned with the eigenvalues (all $\lambda \neq 0$).

Diagonalizability is concerned with the eigenvectors (n independent eigenvectors).

Example 6.11. *Consider the following four matrices:*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Theorem 6.1 (Independent $\boldsymbol{x}$ from different $\boldsymbol{\lambda}$).
Eigenvectors that correspond to distinct eigenvalues are linearly independent. Consequently, an $n \times n$ matrix that has n different eigenvalues (no repeated $\boldsymbol{\lambda}$'s) must be diagonalizable.

Proof. Suppose A has n distinct eigenvalues $\lambda_1, \cdots, \lambda_n$ and corresponding eigenvectors $x_1, \cdots, x_n$. We want to show that $x_1, \cdots, x_n$ are linearly independent.

To describe the idea of the proof, we first work with the case $n = 2$.

Suppose $c_1 x_1 + c_2 x_2 = \vec{0}$ for some scalars c_1 and c_2.

Multiplying by A gives:

Multiplying by λ_2 gives:

Subtraction leaves:

Similarly, multiplying by A, multiplying by λ_1, and subtracting gives:

Only the combination with $c_1 = c_2 = 0$ gives $c_1 x_1 + c_2 x_2 = \vec{0}$. So the eigenvectors x_1 and x_2 are linearly independent.

This proof extends directly to n eigenvectors.

Suppose $c_1 x_1 + c_2 x_2 + \cdots + c_n x_n = \vec{0}$.

Multiplying by A gives:

Multiplying by λ_n gives:

Subtraction leaves:

x_n is gone.

Now multiplying by A and by λ_{n-1} and subtracting will remove x_{n-1}.

Eventually only x_1 will be left:

$$(\lambda_1 - \lambda_n)(\lambda_1 - \lambda_{n-1}) \cdots (\lambda_1 - \lambda_2) c_1 x_1 = \vec{0}$$

Thus $c_1 = 0$. Similarly every $c_i = 0$.

6.3 Symmetric matrices

In the previous section, we have seen that an $n \times n$ matrix A is diagonalizable **if** it has n different eigenvalues. For example, the following upper triangular matrix is diagonalizable (recall thed definition!):

$$A = \begin{bmatrix} \pi & 2 & 3 & 4 \\ 0 & 6 & 7 & 8 \\ 0 & 0 & 10 & 11 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

An eigenvalue λ is repeated if it appears more than once as a root of the characteristic equation $\det(A - \lambda I) = 0$. The number of times it occurs is called its *algebraic multiplicity*.

For example, if $\det(A - \lambda I) = (\lambda - 2)^2(\lambda - 3)$, then $\lambda = 2$ is a repeated eigenvalue with algebraic multiplicity 2, while $\lambda = 3$ has algebraic multiplicity 1.

When A has repeated eigenvalues, it may or may not be diagonalizable. For example,

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

both have (repeated) eigenvalue $\lambda = 1$. But A is ________________, while B is ________________

In this section, we will focus on another type of matrices: symmetric matrices. These matrices have the property that they are diagonalizable even if they have repeated eigenvalues.

Definition 6.3. *A square matrix S is called **symmetric** if $S^T = S$, i.e., $S_{ij} = S_{ji}$ for all i, j.*

Example 6.12.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Theorem 6.2 (Spectral Theorem)**.**
If S is a real symmetric matrix, then it has the following properties:

1. *All the eigenvalues of S are real.*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
3. *There are n orthogonal eigenvectors of S. (We could further normalize them to be orthonormal.)*

The word "spectral" comes from the Latin *spectrum*, meaning "appearance" or "image." In linear algebra, the spectrum of a matrix refers to its set of eigenvalues. The spectral theorem describes how a symmetric matrix can be diagonalized using its eigenvalues and orthonormal eigenvectors.

Remark. A consequence of the Spectral theorem is that every real symmetric matrix S is diagonalizable. Additionally, S can be diagonalized by an orthogonal matrix (!) Q such that

$$S = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

Example 6.13. *The matrix* $S = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ *has eigenvalues* $\lambda_1 = 4$ *and* $\lambda_2 = 2$ *and corresponding eigenvectors* $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

6.4 Complex numbers, vectors and matrices

A complex number is a number of the form $z = a + bi$, where a and b are real numbers and i is the imaginary unit, defined by its property $i^2 = -1$.

The set of complex numbers is denoted by $\mathbf{C}$.

Fundamental Theorem of Algebra. Every polynomial of degree n has n roots in $\mathbf{C}$.

For a complex number $z = a + bi$,

- $\mathrm{Re}(z) = a$ is the real part of z; $\mathrm{Im}(z) = b$ is the imaginary part of z.
- $\overline{z} = a - bi$ is the conjugate of z.
- $|z| := \sqrt{\overline{z}z} = \sqrt{a^2 + b^2}$ is the modulus of z.

Complex arithmetics.

For complex number $z = a + bi$ and $w = c + di$, we have:

- $z + w = (a + c) + (b + d)i$.
- $z - w = (a - c) + (b - d)i$.
- $zw = (ac - bd) + (ad + bc)i$.
- $\dfrac{z}{w} = \dfrac{z\bar{w}}{w\bar{w}} = \dfrac{(a + bi)(c - di)}{c^2 + d^2} = \dfrac{(ac + bd) + (bc - ad)i}{c^2 + d^2}$.

Complex vectors: $\mathbf{C}^n := \{(z_1, \cdots, z_n) : z_i \in \mathbf{C}\}$.

Complex dot product.

$v = (v_1, \cdots, v_n), w = (w_1, \cdots, w_n)$, then

$$v \cdot w = \sum_{i=1}^{n} \overline{v_i} w_i = \overline{v_1} w_1 + \cdots + \overline{v_n} w_n$$

In terms of matrix multiplication, we have $v \cdot w = \overline{v}^T w$.

Norm of a complex vector: $\|v\| = \sqrt{v \cdot v} = \sqrt{\overline{v}^T v}$.

Example 6.14. *Let* $u = (1+i, 2-3i, 4)$ *and* $v = (2, 3+4i, 5)$*. Compute* $u \cdot v$ *and* $\|u\|$*.*

Complex transpose.

Example 6.15. *Compute the complex transpose of the following matrices:*

$$A = \begin{bmatrix} 1 & 2+i \\ 3-4i & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1+i & 2i & 0 \\ 3-4i & 5 & -i \end{bmatrix}$$

Definition 6.4. *If* $A^* = A$*, i.e.,* $a_{ij} = \overline{a_{ji}}$*, then* A *is called a* ***Hermitian matrix****.*

Example 6.16. *The following matrix is Hermitian:*

$$A = \begin{bmatrix} 1 & 2+i \\ 2-i & 3 \end{bmatrix}$$

Remark. The diagonal entries of a Hermitian matrix are real.

Theorem 6.3 ((Complex) Spectral Theorem)**.**
If A is a Hermitian matrix, then it has the following properties:

1. *All the eigenvalues of A are real.*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
3. *There are n orthogonal eigenvectors of A. (We could further normalize them to be orthonormal.)*

Remark. A consequence of the Spectral theorem is that every Hermitian matrix A is diagonalizable. Additionally, A can be diagonalized by a unitary matrix U such that

$$A = U\Lambda U^{-1} = U\Lambda U^*$$

Note. A matrix U is called **unitary** if $U^*U = I$. The columns of a unitary matrix are orthonormal vectors.

Example 6.17. *Let $S = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$. S is Hermitian.*

Eigenvalues of S are $\lambda = 8$ and $\lambda = -1$. Its eigenvectors are $x = \begin{bmatrix} 1 \\ 1+i \end{bmatrix}$ and $y = \begin{bmatrix} 1-i \\ -1 \end{bmatrix}$.

Partial proof of the Spectral Theorem.

If A is Hermitian, i.e., $A^* = A$, then all the eigenvalues of A are real.

Proof. Suppose $Ax = \lambda x$ with $x \neq \vec{0}$.

$$x^*Ax = \lambda x^*x$$

The number x^*x

The number x^*Ax is real because:

So the ratio $\lambda = \dfrac{x^*Ax}{x^*x}$ is real.

If A is Hermitian, then the eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof. Assume

$$Av = \lambda v, \quad Aw = \mu w, \quad \text{with } \lambda \neq \mu.$$

Since A is Hermitian, we have $A = A^*$. We compute $(Av)^*w$ in two ways.

First, using $Av = \lambda v$:

$$(Av)^*w = (\lambda v)^*w = \overline{\lambda}\, v^*w.$$

But $\lambda \in \mathbf{R}$, since eigenvalues of Hermitian matrices are real, so:

$$(Av)^*w = \lambda v^*w.$$

Second, using the Hermitian property:

$$(Av)^*w = v^*Aw = v^*(\mu w) = \mu v^*w.$$

Equating the two expressions:

$$\lambda v^*w = \mu v^*w \quad \Rightarrow \quad (\lambda - \mu)v^*w = 0.$$

Since $\lambda \neq \mu$, it follows that $v^*w = 0$. $\square$

Summary. Real vs Complex matrices

From real to complex matrices, we are replacing the transpose A^T with the complex transpose $A^* := \overline{A}^T$.

A^* share similar properties with A^T:

Real orthogonality:

Complex orthogonality:

Real symmetric matrices:

Complex Hermitian matrices:

Real orthogonal matrices:

Complex unitary matrices:

Positive definite matrices.

For real symmetric matrices $S = S^T$, we know that all the eigenvalues of S are real. If all the eigenvalues of S are positive, then S is called a **positive definite matrix**.

> **Theorem 6.4.**
> *A real symmetric matrix S is positive definite if and only if the energy $x^T Sx$ is positive for all nonzero vectors x.*

Proof.

Part I.

Suppose all the eigenvalues of S are positive. Then the real spectral theorem tells us that

$$S = Q\Lambda Q^T$$

where Q is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix with positive entries.

If x is a nonzero vector, then we can write $x = Qy$ for some nonzero vector y.

Then

$$x^T Sx = x^T Q\Lambda Q^T x = y^T \Lambda y = y_1^2 + \cdots + y_n^2 > 0.$$

Part II.

Suppose $x^T Sx > 0$ for all nonzero vectors x.

Let (λ, y) be an arbitrary eigenpair of S. Then we have

$$y^T Sy = \lambda y^T y.$$

Thus $\lambda = \dfrac{y^T Sy}{y^T y} > 0$.

Example 6.18. *The energy function* $E(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 5y^2 > 0$ *for any* $(x,y) \neq (0,0)$.

6.5 Properties of eigenvalues and eigenvectors

Theorem 6.5.
If (λ, x) is an eigenpair of A, then we have the following properties:

1. (λ^k, x) *is an eigenpair of* A^k *for any positive integer* k.
2. (λ^{-1}, x) *is an eigenpair of* A^{-1} *if* $\lambda \neq 0$.
3. $(\lambda + \mu, x)$ *is an eigenpair of* $A + \mu I$ *for any scalar* μ.

Example 6.19. *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 0 & 1 & 4 \end{bmatrix}$. *Find the eigenvalues and eigenvectors of* A^2, $A^{-1} - I$.

For A, found eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$. Eigenvectors: $x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $x_2 = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$, $x_3 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$.

Theorem 6.6.
Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$. Then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n, \quad \operatorname{trace}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Similar matrices

Suppose the eigenvalue matrix Λ is fixed. As we change the eigenvector matrix X, we get a whole family of different matrices $A = X\Lambda X^{-1}$. These matrices all have the eigenvalues in Λ.

All those matrices A (with the same Λ) are called **similar**.

This idea extends to matrices that can't be diagonalized.

Definition 6.5. *Two matrices A and B are called* ***similar*** *if there exists an invertible matrix C such that $A = CBC^{-1}$.*

Note that in this definition, B is not necessarily diagonal and thus the columns of C are not necessarily eigenvectors of A. We only require that B is invertible.

Theorem 6.7.
If A and B are similar matrices, then they have the same eigenvalues.

6.6 Equivalent conditions for invertible matrices

There have been many equivalent conditions we have seen in this course for square matrices.

From the most recent section to the beginning of the course, we have **proved** the following three groups of equivalent conditions for an $n \times n$ matrix A:

- All the eigenvalues of A are nonzero if and only if $N(A) = \{\vec{0}\}$.

- A is invertible if and only if $\det(A) \neq 0$.

- $N(A) = \{\vec{0}\}$ if and only if $\operatorname{rank}(A) = n$.
- $N(A) = \{\vec{0}\}$ if and only if the columns of A are linearly independent.
- $N(A) = \{\vec{0}\}$ if and only if the rows of A are linearly independent.
- $N(A) = \{\vec{0}\}$ if and only if $Ax = \vec{0}$ has only the trivial solution $x = \vec{0}$.

What we have not proved is Theorem 4.6, which states that the following conditions are equivalent:

A is invertible if and only if $N(A) = \{\vec{0}\}$.

This theorem will connect all the equivalent conditions above, particularly, it connects all those conditions with the invertibility of A.

First, recall the <u>definition</u> of A being invertible.

Proof of Theorem 4.6.

Proof of Theorem 4.6.

Part I. Suppose A is invertible. We want to show that $N(A) = \{\vec{0}\}$.

Let $x \in N(A)$. Then $Ax = \vec{0}$. Multiplying both sides by A^{-1} gives $x = A^{-1}Ax = A^{-1}\vec{0} = \vec{0}$. Thus $N(A) \subseteq \{\vec{0}\}$. But we know that $\{\vec{0}\} \subseteq N(A)$, so $N(A) = \{\vec{0}\}$.

Part II. Suppose $N(A) = \vec{0}$. Then the rank-nullity theorem implies that $\text{rank}(A) = n$ and thus $C(A) = \mathbf{R}^n$.

So there exists a matrix B such that $AB = I$. (Why?)

We want to show that $BA = I$. It suffices to show that $BA - I$ is the zero matrix.

First we observe that $N(B) = \{\vec{0}\}$. Since otherwise, $By = \vec{0}$ for some $y \neq \vec{0}$. Then

$$y = Iy = (AB)y = A(By) = A(\vec{0}) = \vec{0}$$

which is a contradiction. So B has independent rows.

Now we let $C := BA - I$. Then

$$CB = (BA - I)B = B(AB) - B = BI - B = O.$$

Since B has independent rows, C must be the zero matrix and thus $BA = I$.

7 Vector Spaces and Linear Transformations

In the previous chapters, we studied

- the vector space $\mathbf{R}^n$ and its subspaces,
- the $m \times n$ real matrices, which multiply vectors in $\mathbf{R}^n$ to give vectors in $\mathbf{R}^m$

The goals of this brief chapter:

- introduce the abstract concept of vector spaces and linear transformations
- see how these abstract concepts relate to what we have learned in the previous chapters.

7.1 What is a vector space, really?

Vectors are more than arrows or lists of numbers.

Roughly speaking, a vector space is a set of objects (called vectors) that we can "add" and "multiply" by scalars.

Depending on what scalars we are allowed to use, we may have a vector space over the real numbers **R**, or the complex numbers **C**. And we will call it as **F**-vector space where **F** may be **R** or **C**.

The following are all examples of vector spaces:

Example 7.1. $\mathbf{R}^n$ *and its subspaces are all vector spaces over* $\mathbf{R}$*.* $\mathbf{C}^n$ *and its subspaces are all vector spaces over* $\mathbf{C}$*.*

Example 7.2. *The set* $\mathbf{R}^{m\times n}$ *of all* $m \times n$ *matrices with real entries is a vector space over* $\mathbf{R}$*.*

Example 7.3. *The set* $P_n(\mathbf{R})$ *of all polynomials with real coefficients of degree at most* n *is a vector space over* $\mathbf{R}$*.*

Example 7.4. *The set* $C^\infty(\mathbf{R})$ *of all smooth functions* $f : \mathbf{R} \to \mathbf{R}$ *is a vector space over* $\mathbf{R}$*. "Smooth" means that* f *is infinitely differentiable.*

Example 7.5. *The set of all convergent sequences of real numbers is a vector space over* $\mathbf{R}$*.*

Example 7.6. *The set of all the solutions to the ODE:* $y'' + y = 0$ *is a vector space over* $\mathbf{R}$*.*

Definition 7.1 (Axioms of vector spaces)**.** *A vector space V over a field $\mathbf{F}$ is a set of objects, called vectors, together with two operations, called (vector) addition and scalar multiplication, that satisfy the following properties:*

A1. *if $x, y \in V$, then $x + y \in V$.*

A2. *$x + y = y + x$ for all $x, y \in V$.*

A3. *$(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.*

A4. *There exists a unique zero vector $\vec{0}_V \in V$ such that $x + \vec{0}_V = x$ for all $x \in V$.*

A5. *For each $x \in V$, there exists a unique vector $-x \in V$ such that $x + (-x) = \vec{0}_V$.*

A6. *if $\alpha \in \mathbf{F}$ and $x \in V$, then $\alpha x \in V$.*

A7. *$1x = x$ for all $x \in V$.*

A8. *$\alpha(\beta x) = (\alpha\beta)x$ for all $\alpha, \beta \in \mathbf{F}$ and all $x \in V$.*

A9. *$\alpha(x + y) = \alpha x + \alpha y$ for all $\alpha \in \mathbf{F}$ and all $x, y \in V$.*

A10. *$(\alpha + \beta)x = \alpha x + \beta x$ for all $\alpha, \beta \in \mathbf{F}$ and all $x \in V$.*

- A **vector space over F** is also called an **F-vector space**.
- An **R**-vector space is called a **real vector space**.
- An **C**-vector space is called a **complex vector space**.

Theorem 7.1.
Let V be a $\mathbf{F}$-vector space. Then $0x = \vec{0}_V$ for all $x \in V$.

Span, linear independence, basis and dimension

Like $\mathbf{R}^n$, the following are fundamental concepts for vector spaces:

- subspaces
- span
- linear independence/dependence
- basis
- dimension

Definition 7.2. *Let V be an $\mathbf{F}$-vector space. A subset $W \subset V$ is called a* ***subspace*** *of V if the following conditions hold:*

1. *The zero vector $\vec{0}_V$ is in W.*
2. *If $x, y \in W$, then $x + y \in W$.*
3. *If $\alpha \in \mathbf{F}$ and $x \in W$, then $\alpha x \in W$.*

Definition 7.3. *Let V be an $\mathbf{F}$-vector space and let $v_1, \cdots, v_k \in V$. A* ***linear combination*** *of $v_1, \cdots, v_k$ is an expression of the form*

$$c_1 v_1 + \cdots + c_k v_k$$

where $c_1, \cdots, c_k \in \mathbf{F}$ are scalars.

The set of all linear combinations of $v_1, \cdots, v_k$ is denoted by $\operatorname{span}(v_1, \cdots, v_k)$.

Definition 7.4. *Let V be an $\mathbf{F}$-vector space and $S = \{v_1, \cdots, v_k\} \subset V$. The set S is said to be* ***linearly independent*** *if the only linear combination of $v_1, \cdots, v_k$ that equals the zero vector is the trivial combination, i.e., $c_1v_1 + \cdots + c_kv_k = \vec{0}_V$ implies that $c_1 = c_2 = \cdots = c_k = 0$.*

Otherwise, S is said to be ***linearly dependent****.*

Definition 7.5. *Let V be an $\mathbf{F}$-vector space. A set of vectors $B = \{v_1, \cdots, v_k\} \subset V$ is called a* ***basis*** *of V if the following two conditions hold:*

1. *The set B is linearly independent.*
2. *The set B spans the vector space V, i.e., $\mathrm{span}(B) = V$.*

The number k of vectors in a basis B is called the ***dimension*** *of the vector space V, denoted by $\dim(V)$.*

Remark. *When a finite basis does not exist, we say that the dimension of V is infinite.*

Example 7.7. *Consider the vector space* $V = \mathbf{R}^{2\times 2}$*, the set of all* 2×2 *matrices with real entries.*

Example 7.8. *Consider the vector space* $V = P_3(\mathbf{R})$*, the set of all polynomials with real coefficients of degree at most* 3*.*

Example 7.9. *Consider the vector space* $V = P(\mathbf{R})$*, the set of all polynomials with real coefficients.*

7.2 Linear transformations

Consider the vector spaces $\mathbf{R}^n$ and $\mathbf{R}^m$. An $m \times n$ matrix A can be thought of as a transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$: a vector $x \in \mathbf{R}^n$ is transformed into the vector $T(x) = Ax \in \mathbf{R}^m$.

More importantly, such a transformation is linear:

- $A(x + y) = A(x) + A(y)$ for all $x, y \in \mathbf{R}^n$.
- A(αx)=$\alpha A(x)$ for all $\alpha \in \mathbf{R}$ and all $x \in \mathbf{R}^n$.

Definition 7.6. *Let V and W be two $\mathbf{F}$-vector spaces. A function $T : V \to W$ is called a* ***linear transformation*** *if the following conditions hold:*

1. *$T(x + y) = T(x) + T(y)$ for all $x, y \in V$.*
2. *$T(\alpha x) = \alpha T(x)$ for all $\alpha \in \mathbf{F}$ and all $x \in V$.*

Remark. Linear transformations preserve linear combinations:

$u = c_1 v_1 + \cdots + c_n v_n$ must transform to $T(u) = c_1 T(v_1) + \cdots + c_n T(v_n)$.

Remark. Linear transformations send the zero vector in the domain to the zero vector in the codomain.

Theorem 7.2.
Let V and W be two $\mathbf{F}$-vector spaces and $T : V \to W$ be a linear transformation. Then T transforms the zero vector in V to the zero vector in W, i.e., $T(\vec{0}_V) = \vec{0}_W$.

Example 7.10. *Let* $V = \mathbf{R}^3$ *and* $W = \mathbf{R}^1$. *Choose a fixed vector* $a = (1, 3, 4)$ *and define* $T(v) := a \cdot v$. *Then* $T : V \to W$ *is a linear transformation.*

Example 7.11 (Rotation)**.** *Let* $V = \mathbf{R}^2$ *and* $W = \mathbf{R}^2$. *Define* T *as the transformation that rotates every vector in* $\mathbf{R}^2$ *by an angle* $\frac{\pi}{6}$ *counterclockwise. Is* T *a linear transformation?*

Example 7.12. *Let* $V = \mathbf{R}^2$ *and* $W = \mathbf{R}^2$. *Choose a fixed vector* $a = (1, 3)$ *and define* $T(v) := v + a$. *Is* T *a linear transformation?*

Example 7.13. *Let* $V = P_3(\mathbf{R})$ *and* $W = P_3(\mathbf{R})$. *Define* $T(p(x)) = p'(x)$*, the derivative of the polynomial* $p(x)$. *Is* T *a linear transformation?*

Composition of linear transformations

Let V, W and U be three $\mathbf{F}$-vector spaces. Let $T : V \to W$ and $S : W \to U$ be two linear transformations. Then the composition of S and T, denoted by $S \circ T$, is defined as

$$S \circ T : V \to U, \quad S \circ T(v) := S(T(v))$$

for all $v \in V$.

Remark. The composition of two linear transformations is a linear transformation.

Theorem 7.3.
Let V, W and U be three $\mathbf{F}$-vector spaces. Let $T : V \to W$ and $S : W \to U$ be two linear transformations. Then the composition $S \circ T : V \to U$ is a linear transformation.

Remark. Recall that an $m \times n$ matrix times an $n \times q$ matrix is an $m \times q$ matrix.

Basis tells all.

Let V and W be two vector spaces and $T : V \to W$ is a linear transformation. Suppose we know $T(v)$ for all vectors $v_1, \cdots, v_n$ in a basis B of V. Then we know $T(u)$ for any vector $u \in V$. (!)

7.3 Matrix representations of linear transformations

For finite dimensional vector spaces, we can represent a linear transformation by a matrix.

Before we can do that, we need to introduce the concept of **coordinates** of vectors with respect to an ordered basis.

Definition 7.7. *Let V be an n dimensional $\mathbf{F}$-vector space and $A = (\alpha_1, \cdots, \alpha_n)$ be an ordered basis of V. Then every vector $x \in V$ can be expressed as a unique linear combination of the basis vectors:*

$$x = x_1\alpha_1 + \cdots + \cdots + x_n\alpha_n$$

where $x_1, \cdots, x_n \in \mathbf{F}$ are scalars.

We write

$$[x]_A := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

*and call it the **coordinate vector** of x with respect to the ordered basis A.*

Remark. Through coordinate vectors, we have built the connection between an n-dimensional $\mathbf{F}$-vector space V and $\mathbf{F}^n$. (We say that any n-dimensional $\mathbf{F}$-vector space is *isomorphic* to $\mathbf{F}^n$.)

Example 7.14. *Consider* $V = P_3(\mathbf{R})$, *and an ordered basis* $A = (1, x, x^2, x^3)$. *Find the coordinate vector of the polynomial* $p(x) = 2 + 3x + 4x^2$ *with respect to the ordered basis* A.

Example 7.15. *Consider* $V = P_3(\mathbf{R})$, *and an ordered basis* $B = (1 + x, x + x^2, x^2 + x^3, x^3)$. *Find the coordinate vector of the polynomial* $p(x) = 2 + 3x + 4x^2$ *with respect to the ordered basis* B.

We can now encode the linear transformation $T : V \to W$ by a matrix.

Definition 7.8. *Let V and W be two finite-dimensional $\mathbf{F}$-vector spaces. Let $\alpha = (\alpha_1, \cdots, \alpha_n)$ be an ordered basis of V and $\beta = (\beta_1, \cdots, \beta_m)$ be an ordered basis of W.*

*Let $T : V \to W$ be a linear transformation. Then the **matrix representation** of T with respect to the bases α and β is the matrix $[T]_\alpha^\beta$:*

$$[T]_\alpha^\beta = \begin{bmatrix} \\ [T(\alpha_1)]_\beta & \cdots & [T(\alpha_n)]_\beta \\ \\ \end{bmatrix}$$

Example 7.16. *Consider $V = P_3(\mathbf{R})$ and $W = P_3(\mathbf{R})$. Let $\alpha = (1, x, x^2, x^3)$ be an ordered basis of V and $\beta = (1, x, x^2, x^3)$ be an ordered basis of W. Let $T(p(x)) = p'(x)$, the derivative of the polynomial $p(x)$. Find the matrix representation of T with respect to the bases α and β.*

.

The story of Linear Algebra is not over yet...