# Linear Algebra Study Guide

> **ℹ Info**
>
> Hereby released under **Creative Commons BY-SA 4.0**
> └── *provide credit upon repost, all changed versions must be under this license*
>
> jasoncheng.me

> **🔥 Hint**
>
> Your PDF reader may have an **outline** tool! I have embedded heading bookmarks for each chapter, subchapter, and sub-subchapter, so use them to study more efficiently.

---

# 1: Systems of Linear Systems and Matrices

## 1.1: Introduction to Systems of Linear Equations

### Consistency

- A system is **consistent** if it has at least one ordered n-tuple solution.
- **Inconsistent** if it has no solutions.

All linear systems have 0, 1, or infinite solutions.

### Augmented Matrices

An augmented matrix is the matrix form of a linear system.

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & s_{n3} \end{bmatrix}$$

The elementary row operations are:

- ○ Multiply row by a constant
- ○ Add one row to another
- ○ Swap two rows

# Row Echelon Form

- ○ Row echelon form follows these rules:
  - The first nonzero number is 1 (the **leading 1**).
  - All zero rows are at the bottom of the matrix.
  - In any two successive non-zero rows, the leading 1 in the lower one must occur further to the right.
  - Row echelon forms are **not unique**.
- ○ Reduced row echelon form requires one more rule:
  - The column of a leading 1 has zeroes in all other rows
  - Reduced row echelon forms have **pivot positions**.
    - These pivot positions are the leading ones. The nonzero elements that were in those spaces in their non-row echelon forms are **pivots**.

# General Solution

- ○ When there are infinite solutions, the solutions may be expressed as **parametric equations**.
  - Solve the equations for one variable, express the rest as free variables

For example: $x = 2s - t + 3, y = s, z = t$

# 1.2: Gaussian Elimination

## Gaussian-Jordan Elimination

- ○ A way to reduce a matrix to reduced row echelon form
  - Step 1: Move the row with the furthest-left element to the top.
  - Step 2: Reduce the leading element of the top row to 1, by dividing the row by a number.
  - Step 3: Make the elements below the new leading 1 to 0 by multiplying the leading 1's row by a number and adding the row to the rows below
  - Step 4: Repeat Steps 1-3 for the next row down.

**Gaussian Elimination** creates only row echelon form, the extra steps are **Gauss-Jordan Elimination** and creates reduced row echelon form

# Homogenous Linear Systems

○ Homogenous linear systems are of the format:

$$\begin{bmatrix} x_{11} & x_{12} & \ldots & 0 \\ x_{21} & x_{22} & \ldots & 0 \\ x_{31} & x_{32} & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 \end{bmatrix}$$

○ All homogenous linear systems have the trivial solution $x, y, \cdots = 0$. All other solutions are nontrivial. A homogenous linear system can either have no solutions or infinite solutions.

**1.2.1** Free Variable Theorem: If a homogenous linear system has $n$ unknowns and $r$ nonzero rows in its reduced row echelon form, then it has $n - r$ free variables.

**1.2.2** If a homogenous linear system has more unknowns than nonzero lines, it has infinitely many solutions.

# Back-Substitution

○ This technique can solve a system in row echelon form.
  • Step 1. Solve all nonzero lines for their leading variables.
  • Step 2. Successively substitute each line into its preceding line.

# 1.3. Matrices and Matrix Operators

## Row & Column Vectors

○ A matrix with only one row is called a **row vector**.
○ A matrix with only one column is called a **column vector**.

## Square Matrices

○ A matrix with $n \times n$ dimensions is called a square matrix.
○ The diagonal from the top left to bottom right is called the **main diagonal**.
  • The sum of these values is called the **trace**, $tr()$.

## Matrix Operations

○ Matrices are only equal if their elements are equal, and dimensions are equal.
○ The **sum** and **difference** of matrices is each corresponding element being added or subtracted with the other matrix's corresponding element.

- As such, only matrices with the same dimensions can be added or subtracted.
- The **product** of two matrices is found by multiplying each element of a **row** of the first matrix by each element of the **column** of the second matrix, piecewise. The products are added.
    - The **product** of a matrix and a scalar is each element of the matrix being multiplied by the scalar.
    - To determine if two matrices can be multiplied, this rule must be followed:
        - Let two matrices, $A$ and $B$, be multiplied.
        - Write their dimensions like so:
            - $3 \times 4 \qquad 6 \times 3$
        - If the outside values are equal, the multiplication is valid. The inner values are the size of the resultant matrix.
        - Therefore, this multiplication is valid, and the resulting matrix will be of dimensions $4 \times 6$.
        - ^ This is known as the **row-column rule**.

## Partitions and Submatrices

- Matrices can be partitioned into submatrices by inserting lines in between columns or rows.
- This can be used to find the products of only specific columns or lines of two matrices.
    - Let $A$ be a matrix of size $4 \times 8$, and $B$ a matrix of size $5 \times 4$. To find Row 3 of the resultant $AB$, the 3rd row of $A$ can be partitioned and multiplied by $B$.
    - As such:
        - To find the $i$th row of $AB$: ($i$th row of $A$)$B$
        - To find the $j$th column of $AB$: $A$($j$th column of $B$)

**1.3.1** For a square matrix $A$ and column vector $B$ of the same height as $A$, then $A\mathbf{x}$ can be written as each element of $B$ multiplying the corresponding column of $A$, summed.

Example:

$$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

$$2 \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

This can be done with larger matrices to compute columns individually, if the column vector has multiple columns. This example calculated the first column of the resulting matrix.

# Column-Row Expansion

By partitioning two multiply-able matrices $A$ and $B$ into $A$'s component column vectors and $B$'s component row vectors, the product can be calculated easily.

$$AB = c_1 r_1 + c_2 r_2 \ldots c_n r_n$$

As such, there are five total ways to calculate matrix multiplication.

1. Entry by entry (the basic way, Definition 5 in textbook)
2. Row-column method (basically just entry by entry, Formula (5) in textbook)
3. Column by column (1.3.1, Formula (6) in textbook)
4. Row by row (Formula (7) in textbook)
5. Column-row expansion (Formula (11) in textbook)

Note that **matrices cannot be divided, only multiplied by an inverse.**

# Linear System to Matrix

The coefficients of a linear system matrix can be separated into a column vector multiplied by the remaining matrix. That column vector is called the coefficient matrix.

Thus, a linear system can be represented by $A\mathbf{x} = \mathbf{b}$, where $A$ is the coefficient matrix.

# Transposes

The transpose of a matrix $A$ with dimensions $m \times n$ has the rows and columns reversed, with dimensions $n \times m$. Row 1 of $A$ becomes Column 1 of $A'$, Column 1 of $A$ becomes Row 1 of $A'$, and so on. The transpose of square matrices can be obtained by reflecting values across the main diagonal. The main diagonal remains unchanged.

The **trace** of a square matrix is the sum of the values on the main diagonal. There is no trace if the matrix is not a square matrix.

# 1.4: Inverses and Algebraic Properties

## Properties of Matrix Addition

**1.4.1** Properties of matrix arithmetic, where $A, B$ are matrices and $a, b$ are scalars
(a) $A + B = B + A$
(b) $A + (B + C) = (A + B) + C$
(c) $A(BC) = (AB)C$
(d) $A(B + C) = AB + AC$
(e) $(B + C)A = BA + CA$
(f) $A(B - C) = AB - AC$
(g) $(B - C)A = BA - CA$
(h) $a(B + C) = aB + aC$
(i) $a(B - C) = aB - aC$
(j) $(a + b)C = aC + aB$
(k) $(a - b)C = aC - aB$
(l) $a(bC) = (ab)C$
(m) $a(BC) = (aB)C = B(aC)$

## Properties of Matrix Multiplication

The **commutative law** for multiplication DOES NOT APPLY with matrices. Eg, $AB \neq BA$. It can fail for one of these reasons:

- $AB$ is possible but $BA$ isn't due to dimensional reasons (eg. $1 \times 3$ and $3 \times 4$).
- Both $AB$ and $BA$ are possible, but have different sizes (eg. $3 \times 4$ and $4 \times 3$)
- Both $AB$ and $BA$ are possible and have the same size, but different values.

In general, matrices **do not commute** unless under special circumstances.

## Zero Matrices

- A matrix with all-zero elements is a zero matrix.
- Usually, the size is irrelevant, so it's denoted by $0$.

**1.4.1** Properties of matrix arithmetic, where $A, B$ are matrices and $a, b$ are scalars
(a) $A + 0 = A$
(b) $A - 0 = A$
(c) $A - A = 0$

(d) $0(A) = 0$
(e) If $c(A) = 0$, $c = 0$ or $A = 0$

Note that the following two arithmetic rules do not apply in matrix arithmetic:

○ If $ab = ac$ and $a \neq 0$, $b = c$.
○ If $ab = 0$ (note $0 \neq 0$), at least one of $ab$ is 0.

# Identity Matrices

A square matrix with only 1s on the main diagonal and 0s everywhere else is the identity matrix. It is denoted by $I$. An identity matrix of a specific size is $I_n$ for the $n \times n$ identity matrix.

Any matrix $A$ of dimension $m \times n$ follows $AI_n = A$ and $I_m A = A$.

○ Thus, the zero matrix fills the position of 0 from arithmetic (cancellation in multiplication) and the identity matrix fills the position of 1 (unchanging in multiplication).

**1.4.3** If $R$ is the reduced row echelon form of a square matrix $A$, then $R$ has either one row of zeroes or $R$ is the identity matrix $I$.

# Inverse Matrices

For a matrix $A$, an inverse matrix $B$ is such that $AB = I$ and $BA = I$. If there is no such matrix $B$, $A$ is not **invertible**, and is **singular**.

An invertible matrix has only one inverse.

**1.4.5** A $2 \times 2$ matrix is only invertible if its **determinant** $\neq 0$. If so, the inverse is given by the formula

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

The determinant of a $2 \times 2$ matrix is denoted by $\det(A) = ad - bc$ or

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

You will learn a formula for a more universal determinant soon, so don't get too attached.

**1.4.6** If $A$ and $B$ are invertible matrices of the same size, $(AB)^{-1} = A^{-1}B^{-1}$. This applies to combinations of invertible matrices of any amount.

# Matrix Powers

$(A)^0 = I$ and $(A)^n = AAA\ldots A_n$ for square matrix $A$

**1.4.7** If $A$ is invertible and $n$ is a nonnegative integer, then:
(a) $A^{-1}$ is invertible and $(A^{-1})^{-1} = A$.
(b) $A^n$ is invertible and $(A^n)^{-1} = A^{-n} = (A^{-1})^n$.
(c) $kA$ is invertible for any nonzero scalar $k$, and $(kA)^{-1} = k^{-1}A^{-1}$.

# Matrix Polynomials

In the form $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n$, the matrix $p(A)$ of size m x m is $p(A) = a_0 I + a_1 A + a_2 A^2 + \ldots a_n A^n$.

# Properties of Transposition

**1.4.8** If $A$ and $B$ are matrices and $k$ is a scalar:
(a) $(A^T)^T = A$
(b) $(A + B)^T = A^T + B^T$
(c) $(A - B)^T = A^T - B^T$
(d) $(kA)^T = kA^T$
(e) $(AB)^T = B^T A^T$

The transpose of a product of any number of matrices is the product of the transposes in the reverse order.

If A is an invertible matrix, then $A^T$ is also invertible, and $(A^T)^{-1} = (A^{-1})^T$.

# 1.5: Elementary Matrices and Inverse Matrices

Matrices A and B are said to be **row equivalent** if either (hence each) can be obtained from the other by a sequence of elementary row operations.

## Elementary Matrices

A matrix E is called an **elementary matrix** if it can be obtained from an identity matrix by performing a single elementary row operation.

**1.5.1** If the elementary matrix E results from performing a certain row operation on $I_m$ and
if A is an m × n matrix, then the product EA is the matrix that results when this same
row operation is performed on A.

Inverse operations are the opposite of the 3 elementary operations.

- ○ Multiply row $i$ by $\frac{1}{c}$
- ○ Interchange rows $i$ and $j$
- ○ Add $-c$ times row $i$ to row $j$

**1.5.2** All elementary matrices are invertible, and the inverted matrix is also an elementary matrix.

**1.5.3** If $A$ is an $n{\times}n$ matrix, then the following statements are equivalent, that is, all true or
all false.
(a) $A$ is invertible.
(b) $Ax = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ is expressible as a product of elementary matrices

## Inverting Matrices

To find the inverse of an invertible matrix A, find a sequence of elementary row operations that reduces A to the identity and then perform that same sequence of operations on $I_n$ to obtain $A^{-1}$ .

How to know if a matrix is uninvertible: While reducing A to $I_n$, a row of 0s will appear.

# 1.6: More on Invertible Matrices

## Properties of Invertible Matrices

**1.6.3** If $A$ is an $n{\times}n$ matrix:
(a) If $B$ is a square matrix satisfying $BA = 1$, then $B = A^{-1}$.
(b) If $B$ is a square matrix satisfying $AB = 1$, then $B = A^{-1}$.

Hence, adding two more statements to **1.5.3**

**1.6.4** If $A$ is an $n{\times}n$ matrix, then the following statements are equivalent, that is, all true or
all false.
(a) $A$ is invertible.
(b) $Ax = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ is expressible as a product of elementary matrices

(e) $Ax = b$ is consistent for every $n \times 1$ matrix $b$.

(f) $Ax = b$ has exactly one set of solutions for every $n \times 1$ matrix $b$.

**1.6.5** If $A$ and $B$ are $n \times n$ matrices of the same size, if $AB$ is invertible, $A$ and $B$ are also invertible.

# Determining Consistency by Elimination

For a matrix A of size $n$ x $n$, append column matrix B with elements $b_1, b_2, \ldots b_n$ and reduce the left side to an identity matrix, manipulating B elements accordingly. Final result is in the form of parametric equations of arbitrary variables $b_1, b_2, \ldots$.

# 1.7: Diagonal, Triangular, and Symmetric Matrices

## Diagonal Matrices

A square matrix where all items off the main diagonal are 0.

Only invertible if all diagonal entries are nonzero

$D^{-1} = 1/d_1, 1/d_2, \ldots$ where all items off the diagonal remain unchanged.

$D^k = d_1^k, d_2^k, \ldots$ where all items off the diagonal remain unchanged. k is any positive integer.

For $DA$, where $D$ is a diagonal matrix and $A$ is any matrix:

- Let $D$ be a diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_n$, and let $A$ be an $m \times n$ matrix with entries $a_{ij}$.
- The product $DA$ is computed by multiplying each row of $A$ by the corresponding diagonal entry of $D$. Specifically, the $i$-th row of $DA$ is given by:

$$(d_1 a_{i1} \quad d_2 a_{i2} \quad \cdots \quad d_n a_{in})$$

For $AD$, where $D$ is a diagonal matrix and $A$ is any matrix:

- The product $AD$ is computed by multiplying each column of $A$ by the corresponding diagonal entry of $D$. Specifically, the $j$-th column of $AD$ is given by:

$$\begin{pmatrix} d_1 a_{1j} \\ d_2 a_{2j} \\ \vdots \\ d_m a_{mj} \end{pmatrix}$$

# Triangular Matrices

(Square matrix)

Upper triangular matrix: All entries below EXCLUDING the main diagonal are zero
Lower triangular matrix: All entries above EXCLUDING the main diagonal are zero
Diagonal matrices are both upper and lower triangular.

**1.7.1** Properties of triangular matrices
(a) The transpose of an upper triangular matrix is lower, and vice versa
(b) Product of two lower triangular matrices is lower, and product of two upper is upper
(c) A triangular matrix is invertible if and only if its diagonal entries are all nonzero
(d) Inverse of a lower is lower, and inverse of upper is upper

# Symmetric Matrices

A square matrix A is symmetric if $A = A^T$.

A simple way to recognize symmetry is to search for an analogue of an element across the main diagonal. Items on the diagonal are arbitrary, but there must be symmetry for everything else. Thus, if A is symmetric, $A_{ij} = A_{ji}$.

**1.7.2** Properties of symmetric matrices. A and B are symmetric matrices of the same size and k is any scalar.
(a) $A^T$ is symmetric.
(b) $A + B$ and $A - B$ are symmetric.
(c) $kA$ is symmetric.
(d) However, $AB$ is NOT always symmetric. $(AB)^T = AB$ if and only if $AB = BA$, $\therefore A$ and $B$ commute.

Not all symmetric matrices are invertible, but if one is invertible, $A^{-1}$ is also symmetric.

Note: $AA^T$ and $A^T A$ for any matrix A are important in many applications of linear algebra. These two products are always guaranteed to be symmetric.

**1.7.5** If $A$ is an invertible **square** matrix, then $AA^T$ and $A^T A$ are also invertible.

# 1.8: Introduction to Linear Transformations

$\mathbb{R}^n$ is the set of all ordered n-tuples of real numbers. Each element of $\mathbb{R}^n$ is called a vector and referred to in bold lowercase. Each ordered n-tuple can be expressed as a column vector.

$(s_1, s_2, \ldots s_n)$ is called comma delimited form.

$\begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{bmatrix}$ is the column vector form.

Let $\mathbf{e}_i$ be the vector in $\mathbb{R}^n$ with a 1 in the $i$ position and 0s everywhere else. Eg. $\mathbf{e}_2 = (0, 1, \ldots 0)$

All vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard basis vectors for $\mathbb{R}^n$, n defined for positive integers.

All other vectors of $\mathbb{R}^n$ can be expressed as a linear combination of the basis vectors.

## Functions and Transformations

For $b = f(a)$: b is the image of f(a), f(a) is the value of f at a, the set A is the domain of f (all values a is defined for), and the set B is the codomain of f (all values f(a) is defined for). The range is a subset of the codomain (all values f(a) can actually be).

Let domain be $\mathbb{R}^n$ and codomain be $\mathbb{R}^m$, sets of vectors. The function f maps from $\mathbb{R}^n$ to $\mathbb{R}^m$: $f : \mathbb{R}^n \to \mathbb{R}^m$

## Matrix Transformations

This is about transforming $\mathbb{R}^n$ to $\mathbb{R}^m$, written $T : \mathbb{R}^n \to \mathbb{R}^m$.

Let the goal be to transform $\mathbf{x}$ into $\mathbf{w}$ (both are column vectors). This can be denoted by:
$\mathbf{w} = A\mathbf{x}$ which becomes $A : \mathbb{R}^n \to \mathbb{R}^m$. This is used when the domain and codomain need to be clear.

When specifying domain and codomain is not important: $\mathbf{w} = T_A(\mathbf{x})$. This can be expressed in schematic form as $\mathbf{x} \xrightarrow{T_A} \mathbf{w}$. The $T_A$ is a stand-in for multiplication by a matrix $A$.

If the $A$ in $T_A$ is a 0 matrix, $T_0(\mathbf{x}) = 0\mathbf{x} = \mathbf{0}$. $T_0$ is called the zero transformation.

If I is the $n \times n$ identity matrix, $T_I(\mathbf{x}) = I\mathbf{x} = \mathbf{x}$. The original column vector is unchanged. $T_I$ is called the identity operator on $\mathbb{R}^n$.

## Properties of Matrix Transformations

**1.8.1** For every matrix A, the transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ has the following properties for all vectors $\mathbf{u}$ and $\mathbf{v}$, and all scalars $k$:
(a) $T_A(0) = 0$
(b) $T_A(k\mathbf{u}) = kT_A(\mathbf{u})$   [Homogeneity property]
(c) $T_A(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v})$   [Additivity property]
(d) $T_A(\mathbf{u} - \mathbf{v}) = T_A(\mathbf{u}) - T_A(\mathbf{v})$

**1.8.2**   $T : \mathbb{R}^n \to \mathbb{R}^m$ is a valid matrix transformation if and only if these relationships are true for all vectors $\mathbf{u}$ and $\mathbf{v}$, and all scalars $k$:
(a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$   [Additivity property]
(b) $T(k\mathbf{u}) = kT(\mathbf{u})$   [Homogeneity property]

These conditions are called linearity conditions. Any transformation that satisfies these conditions is called a linear transformation.

**1.8.3** Every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a matrix transformation, and vice versa. In other words, only for transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, matrix transformations and linear transformations are synonymous.

**1.8.4** If $T_A : \mathbb{R}^n \to \mathbb{R}^m$ and $T_B : \mathbb{R}^n \to \mathbb{R}^m$ are matrix transformations, and if $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$, $A = B$.

# 2: Determinants

## 2.1: Determinants by Cofactor Expansion

### Cofactor Expansion

$2 \times 2$ matrices are invertible only if $ad - bc \neq 0$, where $\det(A) = ad - bc$ or

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

The minor of an element $a_{ij}$ is the determinant of the submatrix that results when row $i$ and column $j$ are deleted. The cofactor of $a_{ij}$ is $(-1)^{i+j} M_{ij}$.

Pattern that decides if the cofactor is positive or negative times the minor:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**2.1.1** If $A$ is an $n$ x $n$ matrix, then regardless of which row or column of $A$ is chosen, the number obtained by multiplying the entries in that row or column by the corresponding cofactors and adding the resulting products is always the same.

By any row or column of $A$, multiplying by corresponding cofactors and adding them together is called the determinant, and the sums of each column and row themselves are called cofactor expansions.

**2.1.2** If $A$ is an $n$ x $n$ triangular matrix, $\det(A)$ is the product of the diagonal entries.

# 2.2: Evaluating Determinants by Row Reduction

**2.2.1** If $A$ is an $n$ x $n$ matrix, $\det(A) = 0$ if there is a column or row of 0s.

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

Most theorems work for columns and rows.

**2.2.3** For square matrix $A$:
(a) If $B$ is the matrix that results when a single row or column of $A$ is multiplied by a scalar $k$, then $\det(B) = k \det(A)$.
(b) If $B$ is the matrix that results when two rows or columns of $A$ are interchanged, then $\det(B) = -det(A)$.
(c) If $B$ is the matrix that results when a multiple of one row/column of $A$ is added to another row/column of $A$, then $\det(B) = \det(A)$

## Elementary Matrices

**2.2.4** For square elementary matrix $E$:
(a) If $E$ is the matrix that results when a single row of $I_n$ is multiplied by a nonzero number $k$, then $\det(E) = k$.
(b) If $E$ is the matrix that results from interchanging two rows of $I_n$, then $\det(E) = -1$.

(c) If $E$ is the matrix that results when a multiple of one row of $I_n$ is added to another row of $I_n$, then $\det(E) = 1$.

**2.2.5** If $A$ is an $n$ x $n$ matrix with two proportional columns or rows, $\det(A) = 0$.

# 2.3: Properties of Determinants; Cramer's Rule

**2.3.1** If $A$ is an $n$ x $n$ matrix and $k$ is a scalar, $\det(kA) = k^n\det(A)$.

$\det(A+B) \neq \det(A) + \det(B)$

**2.3.2** Let $A, B,$ and $C$ be $n \times n$ matrices that differ only in a single row, say the $r$th, and assume that the $r$th row of $C$ can be obtained by adding corresponding entries in the $r$th rows of $A$ and $B$. Then

$\det(C) = \det(A) + \det(B)$

The same result holds for columns.

**2.3.3** A square matrix $A$ is invertible if and only if $\det(A) \neq 0$.

**2.3.4** For two matrices $AB$ of the same size, $\det(AB) = \det(A) \cdot \det(B)$.

**2.3.5** If $A$ is invertible, $\det(A^{-1}) = \frac{1}{\det(A)}$.

The transpose of the matrix of cofactors of $A$ is the adjoint, $\mathrm{adj}(A)$.

**2.3.6** $A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$.

**2.3.7 Cramer's Rule:** If $A\mathbf{x} = \mathbf{b}$ is a system of linear equations where $\det(A) \neq 0$, the system has a unique solution. The solution is $x^j = \frac{\det(A)}{\det(A_j)}$ for every column j, where $A_j$ is the matrix obtained by replacing the $j$th column of $A$ with the column vector $\mathbf{b}$. (It is usually more efficient, for $n > 3$, to solve systems with $n$ unknowns and $n$ equations, to use Gauss-Jordan elimination)

**2.3.8** For square matrix $A$, all of these statements are equivalent:
(a) $A$ is invertible.
(b) $A\mathbf{x} = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ can be expressed as a product of elementary matrices.
(e) $A\mathbf{x} = b$ is consistent for every $n$ x 1 matrix $\mathbf{b}$.
(f) $A\mathbf{x} = b$ has exactly one solution for every $n$ x 1 matrix $\mathbf{b}$.
(g) $\det(A) \neq 0$.

# 3: Euclidean Vector Spaces

## 3.1: Vectors in 2-Space, 3-Space, and n-Space

### Vector Notation/Formatting

- Direction of arrow signifies **direction**
- Length of arrow signifies **magnitude**
- Tail of arrow is the **initial point**
- Tip of arrow is the **terminal point**

Vectors are in bold, $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \ldots$, scalars are lowercase italics, $a, b, c, d, \ldots$

For a vector $\mathbf{v}$ with initial point $A$ and terminal point $B$, $\mathbf{v} = \overrightarrow{AB}$

Vectors are **equivalent** if and only if they have the same direction and magnitude

A vector with its initial and terminal points at the same point is called the **zero vector**, $\mathbf{0}$. You can make up its direction to whatever is convenient

### Vector Addition

Parallelogram Rule
For two vectors $\mathbf{v}$ and $\mathbf{w}$ with coinciding initial points, take the vectors as sides of a parallelogram, and the sum $\mathbf{v} + \mathbf{w}$ is equal to the vector that shares an initial point at $\mathbf{v}$ and $\mathbf{w}$ and through the parallelogram's diagonal to the opposing vertex.

Triangle Rule
For two vectors $\mathbf{v}$ and $\mathbf{w}$ with $\mathbf{w}$'s initial point at $\mathbf{v}$'s terminal point, the sum $\mathbf{v} + \mathbf{w}$ is equal to the vector from the tip of $\mathbf{v}$ (initial point) to the tail of $\mathbf{w}$ (terminal point).

$\mathbf{v} + \mathbf{w}$ is commutative, $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$. This means you can slap together $\mathbf{v}$ and $\mathbf{w}$ in any order or format you want to take advantage of the parallelogram or triangle rules.

$\mathbf{v} + \mathbf{w} + \mathbf{x}$ is associative, $\mathbf{v} + (\mathbf{w} + \mathbf{x}) = (\mathbf{v} + \mathbf{w}) + \mathbf{x}$. Use tip to tail when dealing with >2 vectors.

### Vector Subtraction

The same as addition but you flip the negative vector around.

# Scalar Multiplication

Works as expected, multiplies the vector's length by that scalar. $k\mathbf{v} = \mathbf{0}$ for $k = 0$ or $\mathbf{v} = \mathbf{0}$.

# Collinearity

Since translating a vector around a vector space technically isn't changing it, being collinear = being parallel. The zero vector is also parallel to all vectors.

# Components

When working with vectors that do not start at the origin, we can define a vector with specific initial and terminal points.

Let $\overrightarrow{P_1P_2}$ be the vector defined by initial and terminal point $P_1(x_1, y_1), P_2(x_2, y_2)$

The components of the vector $\overrightarrow{P_1P_2}$ can be calculated by $\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1)$. Can be extended to 3-vectors and so on.

# n-Space

An **ordered n-tuple** is a sequence of $n$ real numbers. The set of all ordered n-tuples is the **real n-space**, $\mathbb{R}^n$.

$(s_1, s_2, \ldots s_n)$ is called comma delimited form.

$\begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{bmatrix}$ is the column vector form.

**3.1.1** If $\mathbf{u}, \mathbf{v},$ and $\mathbf{w}$ are factors in $\mathbb{R}^n, \mathbf{0}$ is a zero vector, and if $k$ and $m$ are scalars:
(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(b) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(c) $\mathbf{u} + 0 = 0 + \mathbf{u} = \mathbf{u}$
(d) $\mathbf{u} + (-\mathbf{u}) = 0$
(e) $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
(f) $(k + m)\mathbf{u} = k\mathbf{u} + m\mathbf{u}$
(g) $k(m\mathbf{u}) = (km)\mathbf{u}$
(h) $1\mathbf{u} = \mathbf{u}$

**3.1.2** If $\mathbf{v}$ is a factor in $\mathbb{R}^n, \mathbf{0}$ is a zero vector, and $k$ is a scalar:
(a) $0\mathbf{v} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{v} = -\mathbf{v}$

The reason these properties are nice is because calculations on vectors can be done without doing them component-wise.

## Linear Combinations

If $\mathbf{w}$ is a vector in $R^n$, then $\mathbf{w}$ is said to be a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ in $R^n$ if it can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r$$

where $k_1, k_2, \ldots, k_r$ are scalars. These scalars are called the **coefficients** of the linear combination. In the case where $n = 1$, the formula becomes $\mathbf{w} = k_1\mathbf{v}_1$. As such, a linear combination of a single vector becomes just a scalar multiple of that vector.

# 3.2: Norm, Dot Product, and Distance

The **norm** of a vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|$. It is also called the **length** or **magnitude**, and is computed through $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^3 + \cdots + v_n^2}$ for all components $v$ of $\mathbf{v}$.

**3.2.1** If $\mathbf{v}$ is a factor in $R^n$ and $k$ is a scalar:

(a) $\|\mathbf{v}\| \geq 0$

(b) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

(c) $\mathbf{v} = -\mathbf{v}$

## Unit Vectors

A vector of norm 1 is a **unit vector**. Make a unit vector by multiplying a vector by the reciprocal of its norm. As such, a unit vector is defined as $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ share the same direction. This process is called **normalization**.

The **standard unit vectors** are vectors in the positive cardinal directions of the coordinate axes. For example, standard unit vectors of $\mathbb{R}^2$ are $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$. **All vectors in $\mathbb{R}^n$ can be expressed as a linear combination of standard unit vectors.**

## Distance

For two points $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$, the **distance** between $\mathbf{u}, \mathbf{v}$ is

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_1)^2 + \cdots + (u_n - v_n)^2}$$

# Dot Product

Takes two vectors and finds their scalar product. $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$, or you can multiply each component together and add them up, like $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots u_n v_n$ for $\mathbf{u} = (u_1, u_2, \ldots u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots v_n)$.

Useful information about $\theta$ if you know the dot product:
$\theta$ is acute if $\mathbf{u} \cdot \mathbf{v} > 0$.
$\theta$ is obtuse if $\mathbf{u} \cdot \mathbf{v} < 0$.
$\theta$ is 90 deg if $\mathbf{u} \cdot \mathbf{v} = 0$.

If $\mathbf{u} = \mathbf{v}$, there is a special case $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$, and $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

**3.2.2** Algebraic properties of dot products. If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in $R^n$, and $k$ is a scalar, then:
(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$  [Symmetry property]
(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$  [Distributive property]
(c) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$  [Homogeneity property]
(d) $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = 0$  [Positivity property]

**3.2.3** Additional properties of dot products. If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in $R^n$, and $k$ is a scalar, then:
(a) $\mathbf{0} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{0} = 0$
(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
(c) $\mathbf{u} \cdot (\mathbf{v} - \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{w}$
(d) $k(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot k(\mathbf{v})$

# Cauchy-Schwarz Inequality

Solving the dot product for $\theta$:
$\theta = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)$

This only holds true for $-1 \leq \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$ because arccos is only defined for that range. Conveniently, the **Cauchy-Schwarz Inequality** proves the inequality always holds for nonzero $\mathbb{R}^n$ vectors.

**3.2.4** Cauchy–Schwarz Inequality: If $\mathbf{u} = (u_1, u_2, \ldots u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots v_n)$ are vectors in $\mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

or in component form,

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leq (u_1^2 + u_2^2 \cdots + u_n^2)^{\frac{1}{2}}(v_1^2 + v_2^2 \cdots + v_n^2)^{\frac{1}{2}}$$

Don't get too attached, there's a more general form later.

## Geometry in R^n

You know that: (a) the sum of two side lengths of a triangle is at least as large as the third side, and (b) the shortest distance between two points is a straight line. Here are those for $\mathbb{R}^n$.

**3.2.5** If $\mathbf{u}, \mathbf{v},$ and $\mathbf{w}$ are vectors in $R^n$,
(a) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ [Triangle inequality for vectors]
(b)
distance$(\mathbf{u}, \mathbf{v}) \leq$ distance$(\mathbf{u}, \mathbf{w}) +$ distance$(\mathbf{w}, \mathbf{v})$ [Triangle inequality for distances]

This is a version of the typical fact that for parallelograms, the sum of the squares of the diagonals is equal to the sum of the squares of the four sides.

**3.2.6** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $R^n$, then $\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$.

So basically, even though $\mathbb{R}^{4+}$ is a murky dark soup that destroys your brain if you think too hard about it, it's not that different from $\mathbb{R}^2$ and $\mathbb{R}^3$, and many properties generalize.

Important relationship between the norm and dot product:
**3.2.7** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $R^n$, then $\mathbf{u} \cdot \mathbf{v} = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$.

## Dot Products & Matrix Multiplication

Dot products can be seen as matrix multiplication. How $\mathbf{u}$ and $\mathbf{v}$ are arranged depends on if they're expressed as column or row vectors. Arrange vectors or take their transposes accordingly to make sure matrix multiplication is valid.

Essentially this:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^T\mathbf{v}$$

and

$$\mathbf{u} \cdot A\mathbf{v} = A^T\mathbf{u} \cdot \mathbf{v}$$

# 3.3: Orthogonality

## Orthogonality

Two nonzero vectors $\mathbf{u}, \mathbf{v}$ in $R^n$ are orthogonal (perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$. The zero vector in $R^n$ is orthogonal to every vector in $R^n$.

The normal is a way of signifying the slope and inclination of a vector. It is a nonzero vector $\mathbf{n}$ orthogonal to the line/plane.

**3.3.1** If $a$ and $b$ are constants, and both are not zero, the equations for $\mathbf{n}$ in $\mathbb{R}^2$ and $\mathbb{R}^3$ are:
(a) $ax + by + c = 0$
(b) $ax + by + cz + d = 0$
where (a, b) and (a, b, c) is the line/plane of the normal.

**3.3.2** Projection Theorem: If $\mathbf{u}$ and $\mathbf{a}$ are vectors in $\mathbb{R}^n$, and if $\mathbf{a} \neq 0$, then $\mathbf{u}$ can be expressed in exactly one way in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is a scalar multiple of a and $\mathbf{w}_2$ is orthogonal to $\mathbf{a}$.

You can call $\mathbf{w}_1$ the **orthogonal projection of u along a** or **vector component of u along a**, and $\mathbf{w}_2$ **vector component of u orthogonal to a**.

Basically,

$$\text{proj}_a \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2} \mathbf{a} \quad \text{(vector component of } \mathbf{u} \text{ along } \mathbf{a})$$

$$\mathbf{u} - \text{proj}_a \mathbf{u} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2} \mathbf{a} \quad \text{(vector component of } \mathbf{u} \text{ orthogonal to } \mathbf{a})$$

# Norm of a Projection

The norm of a projection can be derived from the equations above, and is

$$\|\text{proj}_\mathbf{a} \mathbf{u}\| = \frac{|\mathbf{u} \cdot \mathbf{a}|}{\|\mathbf{a}\|}$$

Using the conventional definition of a dot product, you can also define it as

$$\|\text{proj}_\mathbf{a} \mathbf{u}\| = \|\mathbf{u}\| |\cos \theta|$$

**3.3.3** Pythagorean Theorem for $\mathbb{R}^n$: For two **orthogonal** vectors in $\mathbb{R}^n$,
$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

There's also some extra stuff but I have no idea what it's talking about so seek it out if you wish

# 3.4: Geometry of Linear Systems

# Vector and Parametric Equations

**3.4.1** For a line with a point $\mathbf{x}_0$ that is parallel to the vector $\mathbf{v}$, the equation of that line is $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$ for a scalar $t$.

**3.4.2** For a plane with a point $\mathbf{x}_0$ that is parallel to the noncollinear vectors $\mathbf{v}_1, \mathbf{v}_2$, the equation of the plane is $\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2$

These equations are called **vector forms**. When points are substituted in to solve for $t$, it is called the **parametric form**.

## Lines through Two Points

The **two-point vector equations** are $\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)$ or $\mathbf{x} = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1$. They define the line segment from $\mathbf{x}_0$ to $\mathbf{x}_1$.

**3.4.3** For a $m \times n$ matrix $A$, the set of solutions of the homogenous linear system $A\mathbf{x} = 0$ consists of all vectors in $\mathbb{R}^n$ that are orthogonal to every row vector in $A$.

# 3.5: Cross Product

Find cross product $a \times b$ in 3-space by making a matrix such that

$$\begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

and take the determinant.

**3.5.1** For 3-space vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$:
(a) $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$  [$\mathbf{u} \times \mathbf{v}$ is orthogonal to u]
(b) $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$  [$\mathbf{u} \times \mathbf{v}$ is orthogonal to v]
(c) $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$  [Lagrange's identity]
(d) $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$  [vector triple product]
(e) $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}$  [vector triple product]

**3.5.2** If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are any vectors in 3-space and $k$ is any scalar, then:
(a) $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$
(b) $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w})$
(c) $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{w})$
(d) $k(\mathbf{u} \times \mathbf{v}) = (k\mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (k\mathbf{v})$
(e) $\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$
(f) $\mathbf{u} \times \mathbf{u} = \mathbf{0}$

## Geometric Interpretation of Cross Products

**3.5.3** If $\mathbf{u}$ and $\mathbf{v}$ are any vectors in 3-space, $\|\mathbf{u} \times \mathbf{v}\|$ is equal to the area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$. As such, the area of the triangle determined by $\mathbf{u}$ and $\mathbf{v}$ is $\frac{\|\mathbf{u} \times \mathbf{v}\|}{2}$.

The special product $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ is called the **scalar triple product**. To find the value, arrange the values of the three vectors and find

$$
\begin{matrix}
u_1 & u_2 & u_3 \\
v_1 & v_2 & v_3 \\
w_1 & w_2 & w_3
\end{matrix}
$$

**3.5.4** The absolute value of $\det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$ is equal to the area of the 2-space parallelogram determined by $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. The volume of the parallelepiped determined by three vectors in 3-space is the same way. As such the volume of such a parallelepiped is equal to $|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|$

**3.5.5** If three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ all have the same initial point, they lie in the same plane if and only if $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 0$.

# 4: General Vector Spaces

## 4.1: Real Vector Spaces

## Vector Space Axioms

Let $V$ be an arbitrary nonempty set of objects. If the following axioms are satisfied by all objects $\mathbf{u}, \mathbf{v}, \mathbf{w}$ that are in $V$ and all scalars $k$ and $m$, then we call $V$ a **vector space** and we call the objects in $V$ **vectors**.

1. If $\mathbf{u}$ and $\mathbf{v}$ are objects in $V$, then $\mathbf{u} + \mathbf{v}$ 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 exists an object in $V$, called the **zero vector**, that is denoted by $\mathbf{0}$ and has the property that $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u}$ in $V$.
5. For each $\mathbf{u}$ in $V$, there is an object $-\mathbf{u}$ in $V$, called a **negative** of $\mathbf{u}$, such that $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$.
6. If $k$ is any scalar and $\mathbf{u}$ is any object in $V$, then $k\mathbf{u}$ is in $V$.
7. $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
8. $(k + m)\mathbf{u} = k\mathbf{u} + m\mathbf{u}$

9. $k(m\mathbf{u}) = (km)(\mathbf{u})$

10. $1\mathbf{u} = \mathbf{u}$

Axioms 1, 4, and 6 are very important, since most of the other axioms can be derived from them. 1 is called **closure under addition** and 6 is called **closure under scalar multiplication**.

Some important vector spaces are:

- $V = \{\mathbf{0}\}$, the zero vector space.
- $V = \{\mathbb{R}^n\}$
- $V = \{\mathbb{R}^\infty\}$, an infinite set of real numbers
- $V = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R} \right\}$, all 2x2 matrices with real entries
  - Vector spaces of matrices of a specific size are denoted as $M_{mn}$
- The vector space of functions defined on all $x$ from $(a, b)$ is $F(a, b)$ with square brackets for closed intervals

## Vector Properties

**4.1.1** Let $V$ be a vector space, $\mathbf{u}$ a vector in $V$, and $k$ a scalar.

(a) $0\mathbf{u} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{u} = -\mathbf{u}$

(d) If $k\mathbf{u} = \mathbf{0}$, either $k = 0$ or $\mathbf{u} = \mathbf{0}$

Sets can define their addition and scalar multiplication rules differently than algebraic operations. For example, $k\mathbf{u}$ can be defined to equal $\mathbf{u}^k$. If they pass all the axioms under these defined conditions, they are still a vector space.

# 4.2: Subspaces

A subset $W$ of a vector space $V$ is a **subspace** if $W$ is a vector space under the addition and scalar multiplication defined with $V$. Since a subspace $W$ must be a vector space, it must obey all ten axioms. $W$ inherits many axioms from $V$, meaning they do not need to be proven for $W$, just $V$.

$W$ does not inherit these axioms:

- Axiom 1: Closure of $W$ under addition
- Axiom 4: Existence of $\mathbf{0}$ in $W$
- Axiom 5: Existence of a negative of every vector in $W$, inside $W$

  ○  Axiom 6: Closure of $W$ under scalar multiplication

**4.2.1** $W$ is a subspace of $V$ if and only if:
(a)  It is closed under addition
(b)  It is closed under scalar multiplication

The conditions in **4.2.1** must be the same addition and scalar multiplication as defined in $V$.

All vector spaces have, at minimum, **two** subspaces: itself and $W = \{\mathbf{0}\}$, the zero subspace.

Special case: In $\mathbb{R}^2$ and $\mathbb{R}^3$, lines and planes through $(0, 0)$ are subspaces.

  ○  The set of all continuous functions on $(-\infty, \infty)$ is a subspace of $F(-\infty, \infty)$ called $C(-\infty, \infty)$.
  - The set of functions with a derivative **that is continuous** is denoted by $C^1(-\infty, \infty)$, where 1 can be up to $\infty$ for which order derivative is referred to. It is also a subspace of $F$.
  ○  The set of all polynomials is also a subspace of $F$ denoted by $P_\infty$.
  - However, the set of all polynomials with degree $n$ is **not** a subspace of $F$. But, the set of all polynomials with degree **including and under** a non-negative integer $n$ **is** a subspace of $F$. That will be denoted by $P_n$.

# The Hierarchy of Function Spaces

Subspaces can be within subspaces. Eg. $C^m$ is inside $C^1$ which is inside $C$ which is inside $F$.

# Building Subspaces

**4.2.2** If $W_1, W_2, \ldots, W_r$ are subspaces of a vector space $V$, then the intersection ($\cap$) of these subspaces is also a subspace.

**4.2.3** The set of solutions of a homogenous equation $A\mathbf{x} = \mathbf{0}$ of $m$ equations and $n$ unknowns is a subspace of $\mathbb{R}^n$.

# 4.3: Spanning Sets

## Spanning Sets

A vector $\mathbf{w}$ in $V$ is said to be a **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ if it can be expressed as $\mathbf{w} = k_1\mathbf{v}_1, k_2\mathbf{v_2}, \ldots, k_r\mathbf{v}_r$ for scalars $k$ (**coefficients**).

**4.3.1** For a non-empty set of vectors in $V$ called $S$:

(a) A set containing all possible linear combinations of vectors in $S$ is a subspace of $V$

(b) That set is the 'smallest' subspace of $V$ that contains all the vectors in $S$.

If that set is called $W$, then we can say that $W = \text{span}(S)$ or $W = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ for the vectors $\mathbf{w}$ of $S$. Essentially, if a set $W$ is said to span a space $V$, every vector in $V$ can be created through linear combinations of $W$

Note: Spanning sets are not unique.

**4.3.2** If $S$ $(\mathbf{v}_1, \ldots)$ and $T$ $(\mathbf{w}_1, \ldots)$ are nonempty sets of vectors in a vector space $V$,

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$$

is true if and only if each vector in $S$ is a linear combination of those in $T$, and vice versa.

# 4.4: Linear Independence

If $S$ is a set of two or more vectors in a vector space $V$, it is **a linearly independent set** if no vector in $S$ can be created via linear combination of the others. If there is only one vector in $S$, it is only linearly independent if that vector is nonzero.

**4.4.1** A nonempty set $S$ in a vector space $V$ is linearly independent if and only if the only coefficients satisfying

$$k_1 \mathbf{v}_1 + k_2 \mathbf{v}_2 + \cdots + k_r \mathbf{v}_r = \mathbf{0}$$

are $k_1 = 0, k_2 = 0, \ldots, k_r = 0$.

The set $1, x, x^2, x^3, \ldots, x^n$ forms a linearly independent set in $P_n$.

**4.4.2** A set with finitely many vectors that contains $\mathbf{0}$ is linearly dependent. A set with two vectors is linearly independent if and only if the two vectors are not scalar multiples of one another.

**4.4.3** For a set $S$ of vectors in $\mathbb{R}^n$ with $r$ vectors, if $r > n$, $S$ is linearly dependent.

## Linear Independence of Functions

To find linear independence, try to use identities and algebra to solve for dependence. If not possible, use the **Wronskian**.

The Wronskian is the determinant of the matrix formed by the derivatives of the functions in a set $W$, up to the $n - 1$ times derivative.

$$W(y_1, y_2, \ldots, y_n)(x) = \begin{vmatrix} y_1(x) & y_2(x) & \cdots & y_n(x) \\ y_1'(x) & y_2'(x) & \cdots & y_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(x) & y_2^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{vmatrix}$$

**4.4.4** If the Wronskian of a set of functions $W$ is not zero on $(-\infty, \infty)$, then the set forms a linearly independent set of vectors in $C^{(n-1)}(-\infty, \infty)$. Note that you must choose the value of $n$ to create a square Wronskian.

# 4.5: Coordinates and Basis

## Basis of a Vector Space

A vector space $V$ is **finite-dimensional** if there is a finite set of vectors in $V$ that spans $V$. It is **infinite-dimensional** if no such set exists.

If $S$ is a set of vectors in a finite-dimensional $V$, then $S$ is called a **basis** for $V$ if $S$ spans $V$ and $S$ is linearly independent.

Remember that if $\det(A) \neq 0$ for an augmented matrix $A$, it will have only the trivial solution in a homogenous system, and it will be consistent for all values of a nonhomogeneous system- i.e it will be a basis for $\mathbb{R}^n$.

## Coordinates Relative to a Basis

**4.5.1** If $S$ is a basis for a vector space $V$, then every vector $\mathbf{v}$ in $V$ can be expressed as a linear combination of the vectors in $S$ in only one way.

If $S$ is an ordered basis (a basis where the order of the vectors in $S$ matters) for a vector space $V$ and

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

is the expression of $\mathbf{v}$ in terms of the vectors of $S$, then the scalars $c_1, c_2, \ldots, c_n$ are the **coordinates of v relative to the basis** $S$. A vector made of the scalars is called the **coordinate vector of v relative to** $S$, denoted by $(\mathbf{v})_s = (c_1, c_2, \ldots, c_n)$.

# 4.6: Dimension

## Dimensionality

**4.6.1** All bases of a finite-dimensional (spanning space has finite vectors) vector space have the same number of vectors.

**4.6.2** Let $V$ be a finite-dimensional vector space and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be any basis for $V$.

(a) If a set in $V$ has more than $n$ vectors, then that set is linearly dependent.

(b) If a set in $V$ has fewer than $n$ vectors, then that set cannot span $V$.

The **dimension** or **degrees of freedom** of a vector space $V$ is denoted by $\dim(V)$ and is the number of vectors in a basis for $V$. The zero vector space has $\dim(\mathbf{0}) = 0$.

○ $\dim(\mathbb{R}^n) = n$

○ $\dim(P_n) = n + 1$

○ $\dim(M_{mn}) = m \cdot n$

**4.6.3 Plus-Minus Theorem**: Let $S$ be a non-empty set of vectors in a vector space $V$.

(a) If $S$ is a linearly independent set, and if $\mathbf{v}$ is a vector in $V$ that is outside span($S$), then the set $S \cup \{\mathbf{v}\}$ that results by inserting $\mathbf{v}$ into $S$ is still linearly independent.

(b) If $v$ is a vector in $S$ that is expressible as a linear combination of other vectors in $S$, and if $S - \{v\}$ denotes the set obtained by removing $\mathbf{v}$ from $S$, then $S$ and $S - \{\mathbf{v}\}$ span the same space; that is,

$$\text{span}(S) = \text{span}(S - \{\mathbf{v}\})$$

Basically, if you add a vector to a linearly independent set that can't create that vector, it stays linearly independent, and if you remove a dependent vector from a dependent set, the span does not change.

**4.6.4** Let $V$ be an $n$-dimensional vector space, and $S$ be a set in $V$ with $n$ vectors. $S$ is a basis for $V$ if it spans $V$ OR $S$ is linearly independent.

**4.6.5** Let $V$ be a finite-dimensional vector space, and $S$ be a set in $V$ with a finite number of vectors.

(a) If $S$ spans $V$ but is not a basis for it, $S$ can become a basis by removing appropriate vectors from $S$.

(b) If $S$ is linearly independent but is not a basis for $V$, it can become a basis by adding appropriate vectors to $S$.

**4.6.6** If $W$ is a subspace of a finite-dimensional vector space $V$,

(a) $W$ is also finite-dimensional.

(b) The dimension of $W$ is smaller than the dimension of $V$.

(c) $W = V$ if and only if $\dim(W) = \dim(V)$.

# 4.7: Change of Basis

## Coordinate Maps

If $S$ is a basis for a finite-dimensional vector space $V$, and the coordinate vector of a vector $\mathbf{v}$ is $(\mathbf{v})_s = (c_1, c_2, \ldots, c_n)$, we can map $\mathbf{v} \to (\mathbf{v})_s$. This creates a connection between the vectors in the vector space and the vectors in $\mathbb{R}^n$. This map is called the **coordinate map relative to** $S$ from $V$ to $\mathbb{R}^n$. Basically, the basis of a vector is useful to find the unique corresponding coordinate vector in $\mathbb{R}^n$ to a vector in $V$.

## Change of Basis

To change basis from $B_1 \to B_2$, do this:

○ For each vector in $B_1$ called $\mathbf{v}_n$, find a linear combination such that $\mathbf{v}_n = k_1 \mathbf{u}_1 \ldots$ for every vector $\mathbf{u}_n$ in $B_2$. The scalars $k_n$ should be arranged into a column.
○ Put together all the columns from all the combinations formed with every vector of $\mathbf{v}_n$ to create a square **transition matrix** $P$ such that the $j$-th column corresponds to the coefficients of the $j$-th vectors in $B_2$.
○ Then, $[\mathbf{v}]_{B_2} = P[\mathbf{v}]_{B_1}$ for every vector $\mathbf{v}$ in $V$. Plug in the found values.

**4.7.1** If $P$ is the transition matrix from a basis $B$ to $B'$, then $P$ is invertible and $P^{-1}$ is the transition matrix from $B'$ to $B$.

A more efficient way to change basis is as follows:

○ Form a matrix [new basis | old basis] in which the basis vectors are in columns.
○ Use elementary row operations to reduce the matrix on the left, the new basis, to $I$.
○ The transition matrix is the remaining matrix on the right.

**4.7.2** Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be any basis for $\mathbb{R}^n$ and $S$ be the standard basis for $\mathbb{R}^n$. If the bases are in column form, then the basis from $B$ to $S$ can be described as

$$P_{B \to S} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$$

# 4.8: Row Space, Column Space, and Null Space

If $A$ is a $n \times m$ matrix:

○ The subspace spanned by the row vectors of $A$ is denoted by $\text{row}(A)$ and is called the **row space**.

○ The subspace spanned by the column vectors of $A$ is denoted by $\text{col}(A)$ and is called the **column space**.

○ The solution space of the homogenous system $A\mathbf{x} = \mathbf{0}$, a subspace of $\mathbb{R}^n$, is denoted by $\text{null}(A)$ and is called the **null space**.

## Relationship between $Ax = 0$ and $Ax = b$

$A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{b}$ are **corresponding equations** if they share a matrix $A$.

**4.8.1** A system of equations $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the column space of $A$.

**4.8.2** If $\mathbf{x}_0$ is any solution of a consistent linear system $A\mathbf{x} = \mathbf{b}$, and $S = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for the null space of $A$, then every solution of $A\mathbf{x} = \mathbf{b}$ can be expressed in the form

$$\mathbf{x} = \mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

For all choices of scalars $c_1, c_2, \ldots$, the vector $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

$\mathbf{x}_0$ is called the **particular solution** of $A\mathbf{x} = \mathbf{b}$, and the remaining part is called the **general solution** of $A\mathbf{x} = \mathbf{0}$. Basically, we can say that **the general solution of a consistent linear system can be expressed as the sum of one solution of the system and the general solution of the corresponding homogenous system.**

**4.8.3**
(a) Row equivalent matrices have the same row space.
(b) Row equivalent matrices have the same null space.

**4.8.4** If a matrix $A$ is reduced to a row echelon form $A_r$ (NOT RREF), then the row vectors with leading ones (nonzero row vectors) form a basis for the row space of $A$, and the column vectors with leading ones form a basis for the column space of $A_r$.

However, since the order of columns can be changed by elementary row operations, this method cannot find the column space of $A$, only of $A_r$.

## Bases Formed From Row and Column Vectors

**4.8.5** If $A$ and $B$ are row equivalent matrices, then
(a) A given set of column vectors of $A$ is linearly independent if and only if the corresponding column vectors of $B$ are also linearly independent.

**(b)** A given set of column vectors of $A$ forms a basis for the column space of $A$ if and only if the corresponding column vectors of $B$ form a basis for the column space of $B$.

Note that you can transpose a matrix to turn row vectors into column vectors and vice versa to take advantage of these theorems. (remember to transpose them again when you're done.)

How to find the basis for the space spanned by a set of vectors:

1. Form a matrix whose columns are the vectors from the set
2. Reduce to row echelon
3. The matrix columns corresponding to the columns with leading ones form a basis for the matrix.
4. Obtain dependency equations by forming the columns without leading ones through linear combinations of the leading one columns (in the reduced row echelon form)
5. These equations work for the original matrix as well

# 4.9: Rank, Nullity, and the Fundamental Matrix Spaces

**4.9.1** The row space and column space of a matrix $A$ have the same dimension.

## Rank and Nullity

- The dimension of the row/column space of a matrix $A$ is called the **rank**, denoted by $\mathrm{rank}(A)$.
- The dimension of the null space of a matrix $A$ is called the **nullity**, denoted by $\mathrm{nullity}(A)$.

The rank of a $m \times n$ matrix is $\min(m, n)$.

**4.9.2 Dimension Theorem for Matrices:** If $A$ is a matrix with $n$ number of columns,

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n$$

**4.9.3** If $A$ is an $m \times n$ matrix, then
(a) $\mathrm{rank}(A) =$ the number of leading variables in the general solution of $A\mathbf{x} = \mathbf{0}$.
(b) $\mathrm{nullity}(A) =$ the leading number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$.

**4.9.4** If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system of $m$ equations and $n$ unknowns, and $A$ has rank $r$, then the general solution has $n - r$ parameters.

# The Fundamental Spaces of a Matrix

The fundamental spaces of a matrix $A$ and its transpose $A^T$ are:

- Column space of $A$
- Row space of $A$
- Null space of $A$
- Column space of $A^T$
- Row space of $A^T$
- Null space of $A^T$

But only 4 are distinct:

- Column space of $A$
- Row space of $A$
- Null space of $A$
- Null space of $A^T$ (or **left null space**, $\dim\left(\text{null}(A^T)\right)$ is called **left nullity**)

**4.9.4** For any matrix $A$, $\text{rank}(A) = \text{rank}(A^T)$.

For a $m \times n$ matrix $A$:

- $\dim\left(\text{row}(A)\right) = r$
- $\dim\left(\text{col}(A)\right) = r$
- $\dim\left(\text{null}(A)\right) = n - r$
- $\dim\left(\text{null}(A^T)\right) = m - r$

# A Geometric Link Between the Fundamental Spaces

If $W$ is a subset of $\mathbb{R}^n$, then the set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in $W$ is called the **orthogonal complement** to $W$ and is denoted by $W^\perp$.

**4.9.6** If $W$ is a subspace of $\mathbb{R}^n$, then:
(a) $W^\perp$ is a subspace of $\mathbb{R}^n$.
(b) The only vector common to $W$ and $W^\perp$ is $\mathbf{0}$.
(c) The orthogonal complement to $W^\perp$ is $W$, or $(W^\perp)^\perp = W$

**4.9.4** If $A$ is an $m \times n$ matrix, then:
(a) The null space of $A$ and row space of $A$ are orthogonal complements in $\mathbb{R}^n$.

(b) The null space of $A^T$ and the column space of $A$ are orthogonal complements in $\mathbb{R}^n$.

**4.9.8** For square matrix $A$ with NO DUPLICATE ROWS/COLUMNS, all of these statements are equivalent:
(a) $A$ is invertible.
(b) $A\mathbf{x} = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ can be expressed as a product of elementary matrices.
(e) $A\mathbf{x} = b$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(f) $A\mathbf{x} = b$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(g) $\det(A) \neq 0$.
(h) The column vectors of $A$ are linearly independent.
(i) The row vectors of $A$ are linearly independent.
(j) The column vectors of $A$ span $\mathbb{R}^n$.
(k) The row vectors of $A$ span $\mathbb{R}^n$.
(l) The column vectors of $A$ form a basis for $\mathbb{R}^n$.
(m) The row vectors of $A$ form a basis for $\mathbb{R}^n$.
(n) $A$ has rank $n$.
(o) $A$ has nullity 0.
(p) The orthogonal complement of the null space of $A$ is $\mathbb{R}^n$.
(q) The orthogonal complement of the row space of $A$ is $\{0\}$.

# 5: Eigenvalues and Eigenvectors

## 5.1: Eigenvalues and Eigenvectors

### Eigenvalues and Eigenvectors

If $A$ is $n \times n$, a vector $\mathbf{x}$ is called an eigenvector of $A$ if $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar $\lambda$. $\lambda$ is called an **eigenvalue** while $\mathbf{x}$ is called the **eigenvector** corresponding to $\lambda$.

**5.1.1** If $\lambda$ is an eigenvalue of $A$, then it is only an eigenvalue if and only if it satisfies $\det(\lambda I - A) = 0$. This equation is the **characteristic equation** of A.

**5.1.2** The eigenvalues of a triangular matrix are the entries on the main diagonal.

**5.1.3** The following statements are equivalent for an $n \times n$ matrix $A$:
(a) $\lambda$ is an eigenvalue of $A$
(b) $\det(\lambda I - A) = 0$
(c) $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.
(d) There is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$

# Finding Eigenvalues and Bases for Eigenspaces

Finding eigenspaces:

- Find all eigenvalues
- Solve $(\lambda I - A)\mathbf{x} = \mathbf{0}$ for $\mathbf{x}$
- Use parametric equations to find column vectors

**5.1.4** A square matrix $A$ is invertible if and only if $\lambda = 0$ is not an eigenvalue.

**5.1.5** For square matrix $A$ with NO DUPLICATE ROWS/COLUMNS, all of these statements are equivalent:
(a) $A$ is invertible.
(b) $A\mathbf{x} = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ can be expressed as a product of elementary matrices.
(e) $A\mathbf{x} = b$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(f) $A\mathbf{x} = b$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(g) $\det(A) \neq 0$.
(h) The column vectors of $A$ are linearly independent.
(i) The row vectors of $A$ are linearly independent.
(j) The column vectors of $A$ span $\mathbb{R}^n$.
(k) The row vectors of $A$ span $\mathbb{R}^n$.
(l) The column vectors of $A$ form a basis for $\mathbb{R}^n$.
(m) The row vectors of $A$ form a basis for $\mathbb{R}^n$.
(n) $A$ has rank $n$.
(o) $A$ has nullity 0.
(p) The orthogonal complement of the null space of $A$ is $\mathbb{R}^n$.
(q) The orthogonal complement of the row space of $A$ is $\{0\}$.
(r) $\lambda = 0$ is not an eigenvalue of $A$.

# 5.2: Diagonalization

## Diagonalization

A **similarity transformation** is when $A$ is mapped to $P^{-1}AP$ for some invertible matrix $P$. Doing this preserves some properties of $A$. These properties are called **invariant under similarity**.

The following are invariant under similarity:

1. **Determinant**: $A$ and $P^{-1}AP$ have the same determinant.

2. **Invertibility**: $A$ is invertible if and only if $P^{-1}AP$ is invertible.
3. **Rank**: $A$ and $P^{-1}AP$ have the same rank.
4. **Nullity**: $A$ and $P^{-1}AP$ have the same nullity.
5. **Trace**: $A$ and $P^{-1}AP$ have the same trace.
6. **Characteristic polynomial**: $A$ and $P^{-1}AP$ have the same characteristic polynomial.
7. **Eigenvalues**: $A$ and $P^{-1}AP$ have the same eigenvalues.
8. **Eigenspace dimension**: If $\lambda$ is an eigenvalue of $A$ (and hence of $P^{-1}AP$), then the eigenspace of $A$ corresponding to $\lambda$ and the eigenspace of $P^{-1}AP$ corresponding to $\lambda$ have the same dimension.

If $A$ and $B$ are square matrices, $B$ is similar to $A$ if there is an invertible matrix $P$ such that $B = P^{-1}AP$. $B$ is a diagonal matrix such that the eigenvalues of $A$ are on the main diagonal.

If $A$ is a square matrix, it is diagonalizable if it is similar to some diagonal matrix. Or, there is some matrix $P$ such that $P^{-1}AP$ is diagonal.

**5.2.1** The following statements are equivalent for an $n \times n$ matrix $A$:
(a) $A$ is diagonalizable.
(b) $A$ has $n$ linearly independent eigenvectors.

**5.2.2** The following statements are equivalent for an $n \times n$ matrix $A$:
(a) If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix $A$, and if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are the corresponding eigenvectors, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set.
(b) An $n \times n$ matrix with $n$ distinct eigenvalues is diagonalizable.

# Procedure to Diagonalize a Matrix

Procedure for diagonalizing a matrix:

- Find basis vectors of a matrix.
- Form them into a matrix such that each basis vector is a column.
- Use $P^{-1}AP$ to find $A$.

# Computing Large Powers of Matrices

**5.2.3** If $\lambda$ is an eigenvalue of $A$ and $\mathbf{x}$ is its corresponding eigenvector, $\lambda^k$ for a nonnegative integer $k$ is an eigenvalue of $A^k$, and $\mathbf{x}$ is its corresponding eigenvector.

When taking large powers of matrices, use

$$B = P^{-1}AP$$
$$A = P^{-1}BP$$
$$A^k = P^{-1}B^kP$$

where $k$ can be set to an arbitrarily large power.

While an $n \times n$ matrix with $n$ distinct eigenvalues is diagonalizable, the inverse is **false**. If a matrix is diagonalizable, it does not guarantee it has $n$ eigenvalues.

**5.2.4** For an $n \times n$ matrix $A$:
(a) For every eigenvalue of $A$, the geometric multiplicity is less than or equal to the algebraic multiplicity.
(b) $A$ is diagonalizable if and only if its characteristic polynomial can be expressed as a product of linear factors, and the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity.

# 5.3: Complex Vector Spaces

## Complex Numbers

If $z = a + bi$ is a complex number, then

- $\text{Re}(z) = a$ and $\text{Im}(z) = b$ are called the **real part** of $z$ and the **imaginary part** of $z$ respectively.
- $|z| = \sqrt{a^2 + b^2}$ is called the **modulus** (or **absolute value**) of $z$.
- $\bar{z} = a - bi$ is called the **complex conjugate** of $z$. I like to think of it as $z$ but **evil** and **nefarious**.
- $z\bar{z} = a^2 + b^2 = |z|^2$
- The angle $z$ makes with the $\text{Re}(z)$ axis on the complex plane is an **argument** of $z$.
- $\text{Re}(z) = |z| \cos(\phi)$.
- $\text{Im}(z) = |z| \sin(\phi)$.
- $z = |z|(\cos(\phi) + i \sin(\phi))$ is called the **polar form** of $z$.

## Vectors in $C^n$

$\mathbb{C}^n$, or the **complex n-space**, is the set of all sequences of complex numbers.

**5.3.1** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{C}^n$, and if $k$ is a scalar, then:
(a) $\bar{\bar{\mathbf{u}}} = \mathbf{u}$
(b) $\overline{k\mathbf{u}} = \bar{k}\bar{\mathbf{u}}$

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

(d) $\overline{\mathbf{u} - \mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}}$

**5.3.2** If $A$ is an $m \times n$ complex matrix and $B$ is a $n \times p$ complex matrix, then

(a) $\overline{\overline{A}} = A$

(b) $\overline{\left(A^T\right)} = \left(\overline{A}\right)^T$

(c) $\overline{AB} = \overline{A}\,\overline{B}$

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{C}^n$, then the **complex Euclidean inner product** (or **complex dot product**) is $\mathbf{u} \cdot \mathbf{v}$ and defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n$$

The **Euclidean norm** on $\mathbb{C}^n$ is

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}$$

As normal, $\mathbf{v}$ is a **unit vector** if $\|\mathbf{v}\| = 1$, and two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

**5.3.3** If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in $\mathbb{C}^n$, and if $k$ is a scalar, then the complex Euclidean inner product has the following properties:

(a) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ 　　[Antisymmetry property]

(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ 　　[Distributive property]

(c) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$ 　　[Homogeneity property]

(d) $\mathbf{u} \cdot (k\mathbf{v}) = \overline{k}(\mathbf{u} \cdot \mathbf{v})$ 　　[Antihomogeneity property]

(e) $\mathbf{u} \cdot \mathbf{v} \geq 0$ and $\mathbf{u} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = 0$ 　　[Positivity property]

# Vector Concepts in $C^n$

**5.3.4** If $\lambda$ is an eigenvalue of a real $n \times n$ matrix $A$, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of $A$, and $\overline{\mathbf{x}}$ is its corresponding eigenvector.

**5.3.5** If $A$ is a $2 \times 2$ matrix with real entries, then the characteristic equation of $A$ is $\boxed{\lambda^2 - \text{tr}(A)\lambda + \det(A) = 0}$ and

(a) $A$ has two distinct real eigenvalues if $\text{tr}(A)^2 - 4\det(A) > 0$

(b) $A$ has one distinct repeated real eigenvalue if $\text{tr}(A)^2 - 4\det(A) = 0$

(c) $A$ has two complex conjugate eigenvalues if $\text{tr}(A)^2 - 4\det(A) < 0$

^ this comes in very handy even for non-complex matrices!!

**5.3.6** If $A$ is a real symmetric matrix, then $A$ has real eigenvalues.

# A Geometric Interpretation of Complex Eigenvalues

**5.3.7** The eigenvalues of the real matrix

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

are $\lambda = a \pm bi$. If $a$ and $b$ are not both zero, then this matrix can be factored as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

where $\phi$ is the angle from the positive x-axis to the ray that joins the origin to the point $(a, b)$.

**5.3.8** Let $A$ be a real $2 \times 2$ matrix with complex eigenvalues $\lambda = a \pm bi \quad (b \neq 0)$. If $\mathbf{x}$ is an eigenvector of $A$ that corresponds to $\lambda = a - bi$, then the matrix $P = [\text{Re}(\mathbf{x}) \quad \text{Im}(\mathbf{x})]$ is invertible and

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1}$$

# 5.5: Dynamical Systems and Markov Chains

A **dynamical system** is a set of variables that change with time. The value of a vector at any point is the **state**, and a vector formed of these states is a **state vector**.

## Markov Chains

A **probability vector** is a matrix with $n$ possible states, expressed as probabilities.

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \quad \begin{matrix} \text{Probability the system is at State 1} \\ \text{Probability the system is at State 2} \\ \vdots \\ \text{Probability the system is at State } n \end{matrix}$$

A square matrix with probability vectors as columns is called a **stochastic matrix**.

A **Markov Chain** is a dynamical system whose state vectors at successive times are defined by an equation of the form

$$\mathbf{x}(k+1) = P\mathbf{x}(k)$$

in which $P = [p_{ij}]$ is a stochastic matrix and $p_{ij}$ is the probability that the system

will be in state $i$ by time $t = k$. The matrix $P$ is called the **transition matrix** for the system.

A transition matrix ($n \times n$) looks like this:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

where the columns represent the state of the system at $t = n$ and the rows represent the chances of the system moving to that specific state on $t = n + 1$.

You can find the state vector at any time $t$ by manipulating the transition matrix via

$$\mathbf{x}_t = P^t \mathbf{x}_0$$

which is useful for finding any state vector immediately from the initial state vector.

# Long-Term Behavior of a Markov Chain

In some cases, a state vector will stabilize after a sufficiently long number of manipulations by $P$. But, not all chains stabilize. A stochastic matrix $P$ is said to be **regular** if $P$ or some positive power of $P$ has all positive entries. A Markov chain whose transition matrix is regular is said to be a **regular Markov chain**.

**5.5.1** If $P$ is the transition matrix of a regular Markov chain, then:
(a) There is a unique probability vector $\mathbf{q}$ with positive entries such that $P\mathbf{q} = \mathbf{q}$.
(b) For any initial probability vector $\mathbf{x}_0$, the sequence of state vectors $\mathbf{x}_0, P\mathbf{x}_0, \ldots, P^k\mathbf{x}_0, \ldots$ converges to $\mathbf{q}$.
(c) The sequence $P, P^2, \ldots, P^k, \ldots$ converges to the matrix $Q$, each of whose column vectors is $\mathbf{q}$.

To find $\mathbf{q}$:

- Make sure the matrix has a steady-state matrix (it is regular)
- Solve $(I - P)\mathbf{q} = 0$ for a set of parametric equations
- You should get a vector like $\begin{bmatrix} \frac{1}{2}s \\ s \end{bmatrix}$. Use $1 = q_1 + q_2 + \ldots$ for every element in the vector to find the value of the parameter (in this case $s$).
- Substitute back into the vector.

# 6: Inner Product Spaces

## 6.1: Inner Products

### General Inner Products

You should assume all vector spaces in this chapter, unless said so implicitly, are real.

An **inner product** on a real vector space $V$ is a function that associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ with a pair of vectors in $V$ in such a way that the following axioms are satisfied for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and all scalars $k$.

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. $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$

   A real vector space with an inner product is called a **real inner product space**.

We can conveniently fill all of these axioms instantly by defining $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$, which is how **dot product** is defined.

If $V$ is a real inner product space, then the **norm** (**length**) of a vector $\mathbf{v}$ in $V$ denoted by $\|\mathbf{v}\|$ and equals

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

and the distance between two vectors is denoted by $d(\mathbf{u}, \mathbf{v})$ and equals

$$d(\mathbf{u}, \mathbf{v} = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle})$$

**6.1.1** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a vector space $V$, and $k$ is a scalar:
(a) $\|\mathbf{v}\| \geq 0$ if and only if $\mathbf{v} = 0$
(b) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$
(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
(d) $d(\mathbf{u}, \mathbf{v}) \geq 0$ if and only if $\mathbf{u} = \mathbf{v}$

You can also add positive real number **weights** to each component-wise part of the dot product formula to modify the resultant product like

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n$$

called the **weighted Euclidean inner product with weights** $w_1, w_2, \ldots, w_n$.

## Unit Circles and Spheres in Inner Product Spaces

If $V$ is an inner product space, the set of points in $V$ that satisfy $\|\mathbf{u}\| = 1$ is called the **unit sphere** in $V$ (or the **unit circle** in $\mathbb{R}^2$)

## Inner Products Generated by Matrices

The Euclidean inner product is a subset of a greater class of inner products named **matrix inner products**. Let $\mathbf{u}$ and $\mathbf{v}$ be column vectors in $\mathbb{R}^n$ and $A$ be an invertible $n \times n$ matrix. Then, this

$$\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}$$

also defines an inner product called the **inner product on $\mathbb{R}^n$ generated by $A$**.

Alternatively, since $\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^t\mathbf{u}$, we can rewrite this as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T A^T A\mathbf{u}$$

Notice that for the **Euclidean inner product**, we simply use $A = I$.

## Algebraic Properties of Inner Products

**6.1.2** If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a vector space $V$, and $k$ is a scalar:
(a) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$
(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
(c) $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$
(d) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$
(e) $k\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$

# 6.2: Angle and Orthogonality in Inner Product Spaces

Recall the Cauchy-Schwarz Inequality from Chapter 3. The following is a more general form of the inequality.

**6.2.1 Cauchy-Schwarz Inequality:** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space $V$, then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

or these two alternate forms

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle\langle \mathbf{v}, \mathbf{v} \rangle$$
$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2$$

Using Cauchy-Schwarz, we can define the $\theta$ between two vectors $\mathbf{u}$ and $\mathbf{v}$ to be $\theta = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)$.

## Length and Distance in Inner Product Spaces

**6.2.2** If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a real inner product space $V$, and if $k$ is any scalar, then:
(a) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$     [Triangle inequality for vectors]
(b) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$     [Triangle inequality for distances]

## Orthogonality

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{v}, \mathbf{u} \rangle = 0$.

**6.2.3 Generalized Theorem of Pythagoras:** If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in an inner product space, then

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

If $W$ is a subspace of a real inner product space $V$, then the set of all vectors in $V$ that are orthogonal to every vector in $W$ is called the **orthogonal complement** of $W$ and is denoted by $W^{\perp}$.

**6.2.4** If $W$ is a subspace of a real inner product space $V$:
(a) $W^{\perp}$ is a subspace of $V$
(b) $W \cap W^{\perp} = \mathbf{0}$

**6.2.5** If $W$ is a subspace of a finite-dimensional inner product space $V$, then the orthogonal complement of $W^{\perp}$ is $W$.

# 6.3: Gram-Schmidt Process; QR-Decomposition

## Orthogonal and Orthonormal Sets

A set of two or more vectors in a real inner product space is said to be **orthogonal** if all pairs of distinct vectors in that set are orthogonal. An orthogonal set in which all vectors have norm 1 is said to be **orthonormal**.

You can make an orthonormal set out of an orthogonal set by using $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ for every vector in the set. (This process is called **normalization**).

**6.3.1** If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space, then $S$ is linearly independent.

A basis consisting of orthonormal and orthogonal vectors are called an **orthonormal** and **orthogonal basis** respectively.

# Coordinates Relative to Orthonormal Bases

**6.3.2**
(a) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for an inner product space $V$, and $\mathbf{u}$ is any vector in $V$,

$$\mathbf{u}(s) = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n$$

(b) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for an inner product space $V$, and $\mathbf{u}$ is any vector in $V$,

$$\mathbf{u}(s) = \langle \mathbf{u}, \mathbf{v}_1 \rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n \rangle\mathbf{v}_n$$

# Orthogonal Projections

The Projection Theorem defined in Chapter 3 is a special case of the following general theorem:

**6.3.3** If $W$ is a finite-dimensional subspace of an inner product space $V$, then every vector $\mathbf{u}$ in $V$ can be expressed in only one way in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is in $W$ and $\mathbf{w}_2$ is in $W^\perp$.

The projection vectors are commonly referred to by:

- $\mathbf{w}_1 = \text{proj}_W\mathbf{u}$
- $\mathbf{w}_2 = \text{proj}_{W^\perp}\mathbf{u}$

$\mathbf{u}$ can also be expressed as $\text{proj}_W\mathbf{u} + \text{proj}_{W^\perp}\mathbf{u} = \text{proj}_W\mathbf{u} + (\mathbf{u} - \text{proj}_W\mathbf{u})$.

**6.3.4**
(a) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthogonal basis for $W$, and $\mathbf{u}$ is any vector in $V$, then

$$\text{proj}_W\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

(b) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for $W$, and $\mathbf{u}$ is any vector in $V$, then

$$\text{proj}_W\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_r \rangle\mathbf{v}_r$$

# The Gram-Schmidt Process

**6.3.5** Every nonzero finite-dimensional inner product space has an orthonormal basis.

The Gram-Schmidt Process is used to convert a basis $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ into an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

1. $\mathbf{v}_1 = \mathbf{u}_1$
2. $\mathbf{v}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1$
3. $\mathbf{v}_3 = \mathbf{u}_3 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2} \mathbf{v}_2$
4. $\mathbf{v}_4 = \mathbf{u}_4 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2} \mathbf{v}_3$
5. $\vdots$
6. repeat as many times as necessary ($n$ times)

**6.3.6** If $W$ is a finite-dimensional inner product space:
(a) Every orthogonal set of nonzero vectors in $W$ can be enlarged to an orthogonal basis for $W$.
(b) Every orthonormal set in $W$ can be enlarged to an orthonormal basis for $W$.

# QR-Decomposition

**6.3.7** **QR-Decomposition:** If $A$ is an $n \times n$ matrix with linearly independent columns, then $A$ can be factored as

$$A = QR$$

where $Q$ is an $m \times n$ matrix with orthonormal columns, and $R$ is an $n \times n$ invertible upper triangular matrix.

Since a matrix will have linearly independent columns if and only if it is invertible, all **invertible matrices can be converted into a QR-decomposition**.

1. Separate $A$ into column vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.
2. Apply the Gram-Schmidt process to each.
3. Normalize the orthogonal vectors to create $n$ orthonormal column vectors.
4. It follows that

$$R = \begin{bmatrix} \langle u_1, q_1 \rangle & \langle u_2, q_1 \rangle & \langle u_3, q_1 \rangle \\ 0 & \langle u_2, q_2 \rangle & \langle u_3, q_2 \rangle \\ 0 & 0 & \langle u_3, q_3 \rangle \end{bmatrix}$$

5. $Q$ is equal to the reassembled orthonormal column vectors.

# 6.4: Best Approximation; Least Squares

# Least Squares Solutions of Linear Systems

Sometimes, when we have a system $A\mathbf{x} = \mathbf{b}$ that is **inconsistent**, we will try and find the closest $\mathbf{x}$ to a solution. To do this, we find the $\|\mathbf{b} - A\mathbf{x}\|$. The vector $\mathbf{x}$ in this expression is the **least squares solution**, $\mathbf{b} - A\mathbf{x}$ is the **least squares vector**, and $\|\mathbf{b} - A\mathbf{x}\|$ is the **least squares error**.

It's called "least squares" because minimizing the least squares error has the side effect of minimizing the component vectors of $\mathbf{b} - A\mathbf{x}$, which are squares (recall that $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots}$).

**6.4.1 Best Approximation Theorem**: If $W$ is a finite-dimensional subspace of an inner product space $V$, and $\mathbf{b}$ is a vector in $V$, then $\operatorname{proj}_W \mathbf{b}$ is the **best approximation to b from** $W$ such that $\|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$ for every vector $\mathbf{w}$ in $W$ that isn't $\operatorname{proj}_W \mathbf{b}$.

# Finding Least Squares Solutions

The **normal equation** associated with a system $A\mathbf{x} = \mathbf{b}$ is

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

and as such,

**6.4.2** For every linear system $A\mathbf{x} = \mathbf{b}$, the associated normal system

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

is consistent, and all solutions are least squares solutions of $A\mathbf{x} = \mathbf{b}$. Moreover, if $\mathbf{x}$ is any least squares solution and $W$ is the column space of $A$, then

$$A\mathbf{x} = \operatorname{proj}_W \mathbf{b}$$

# Conditions for Uniqueness of Least Squares Solutions

**6.4.3** If $A$ is an $m \times n$ matrix, then the following are equivalent.
(a) The column vectors of $A$ are linearly independent.
(b) $A^T A$ is invertible.

**6.4.4** If $A$ is an $m \times n$ matrix with linearly independent column vectors, then for every $m \times 1$ matrix $\mathbf{b}$, $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution. This solution is given by

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$$

and if $W$ is the column space of $A$, then

$$A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b} = \text{proj}_W \mathbf{b}$$

**6.4.5** For square matrix $A$ with no duplicate rows/columns, all of these statements are equivalent:

(a) $A$ is invertible.
(b) $A\mathbf{x} = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ can be expressed as a product of elementary matrices.
(e) $A\mathbf{x} = b$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(f) $A\mathbf{x} = b$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(g) $\det(A) \neq 0$.
(h) The column vectors of $A$ are linearly independent.
(i) The row vectors of $A$ are linearly independent.
(j) The column vectors of $A$ span $\mathbb{R}^n$.
(k) The row vectors of $A$ span $\mathbb{R}^n$.
(l) The column vectors of $A$ form a basis for $\mathbb{R}^n$.
(m) The row vectors of $A$ form a basis for $\mathbb{R}^n$.
(n) $A$ has rank $n$.
(o) $A$ has nullity 0.
(p) The orthogonal complement of the null space of $A$ is $\mathbb{R}^n$.
(q) The orthogonal complement of the row space of $A$ is $\{0\}$.
(r) $\lambda = 0$ is not an eigenvalue of $A$.
(s) $A^T A$ is invertible.

# 6.5: Mathematical Modeling Using Least Squares

A **mathematical model** is an equation that can be used to extrapolate values past collected or measured data. A **polynomial model** uses polynomials as the equation form.

## Least Squares Fit of a Straight Line

A **regression line** is a straight line that minimizes the sum of the squares of the data point errors of a set of data.

Ideally, if all of the data points can be placed on one line, then we can transform the equations of the data points

$$y_1 = a + bx_1$$
$$y_2 = a + bx_2$$
$$\cdots$$

into a system

$$\begin{bmatrix} 1 & x \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

Let us associate these matrices with variables, such that this equation becomes $M\mathbf{v} = \mathbf{y}$. When the data points do not exactly sit on the line (practically always), we instead try to solve the equation

$$\begin{bmatrix} 1 & x \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix}^T \begin{bmatrix} 1 & x \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & x \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

for the quantity

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

or as we established, as $M^T M \mathbf{v} = M^T \mathbf{y}$.

Placing the elements of this matrix into $y = ax + b$ gives us the regression line. This line minimizes the quantity $\|\mathbf{y} - M\mathbf{v}\|^2$, which can be expressed as

$$(y_1 - (a + bx_1))^2 + (y_2 - (a + bx_2))^2 + \cdots + (y_n - (a + bx_n))^2$$

where the quantities
$d_1 = |y_1 - (a + bx_1)|, \quad d_2 = |y_2 - (a + bx_2)|, \quad \ldots \quad d_2 = |y_2 - (a + bx_2)|, \quad d_n = |y_n - $
are called **residuals**.

The residual $d_n$ is the distance between the data point $x_n, y_n$ and the regression line. Since the regression line minimizes the squares of the residuals (the 'error'), the regression line is sometimes called the **line of best fit**.

**6.5.1** Given a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and two matrices $M$ and $\mathbf{y}$, there exists a unique least squares line fit of the form $y = a^* + b^* x$ that best fits these data points. Furthermore, the coefficients $a^*$ and $b^*$ that define this line can be found by solving the equation $\mathbf{v} = (M^T M)^{-1} M^T \mathbf{y}$, where $\mathbf{v} = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$, and the solution $\mathbf{v}$ is also unique.

# Least Squares Fit of a Polynomial

We can generalize the method used for a straight line to a polynomial. Use $\mathbf{v} = (M^T M)^{-1} M^T \mathbf{y}$ to find $a^*, b^*$.

# 6.6: Function Approximation; Fourier Series

## Best Approximations

An **approximation** is a function of a specific form that is the closest possible to a function of another form. For example, approximation of $\sin(x)$ over $[1, \frac{\pi}{2}]$ through a function of form $ax^2 + bx + c$.

## Measurements of Error

At any point $x_0$ between two functions $f(x)$ and $g(x)$, the distance between the two functions (**deviation**) is $|f(x_0) - g(x_0)|$. To extend this across an area, we integrate:

$$\text{error} = \int_b^a |f(x) - g(x)| \, dx$$

(that's right, Calc 1 lied to you, it doesn't matter which one is on top, just use an absolute value)

The value $\int_a^b [f(x) - g(x)]^2$ is called the **mean square error** and is the value that needs to be minimized.

## Least Squares Approximations

Let $f(x)$ be a function continuous on $[a, b]$. Let $C[a, b]$ have the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x) \, dx$$

and let $W$ be a finite-dimensional subspace of $C[a, b]$. Thus, we must find a function $\mathbf{g}$ that minimizes

$$\|\mathbf{f} - \mathbf{g}\|^2 = \int_a^b [f(x) - g(x)]^2$$

**6.6.1** If $\mathbf{f}$ is a continuous function on $[a, b]$, and $W$ is a finite-dimensional subspace of $C[a, b]$, then the function $g$ in $W$ that minimizes the mean square error

$$\int_a^b |f(x) - g(x)|^2 \, dx$$

is $\mathbf{g} = \text{proj}_W\mathbf{f}$, where the orthogonal projection is relative to the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx$$

The function $\mathbf{g} = \text{proj}_W\mathbf{f}$ is called the least squares approximation to $f$ from $W$.

# Fourier Series

A function of the form $T(x) = c_0 + c_1 f(x) + c_2 f(2x) + \cdots + c_n f(nx)$ where $f(x)$ is one of the trigonometric functions $\sin, \cos$ is called a **trigonometric polynomial**. If $c_n$ and $d_n$ are both nonzero, then $T(x)$ has **order** $n$.

If we attempt to find a least squares approximation of a function $f(x)$ by a trigonometric polynomial $T(x)$ of order $n$ or less over $[0, 2\pi]$, we end up with

$$\text{proj}_W f = \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots b_n \sin nx].$$

where

$$a_0 = \frac{2}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g_0}\rangle = \frac{2}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx$$

$$a_1 = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g_1}\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\cos x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx$$

$$\vdots$$

$$a_n = \frac{1}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g_n}\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\cos(nx)\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(nx)\,dx$$

$$b_1 = \frac{1}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g_{n+1}}\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\sin x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin x\,dx$$

$$\vdots$$

$$b_n = \frac{1}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g_{2n}}\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\sin(nx)\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(nx)\,dx$$

Or in other words,

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(kx)\,dx, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(kx)\,dx$$

and all $a, b$ are known as the **Fourier coefficients** of $\mathbf{f}$.

# 7. Diagonalization and Quadratic Forms

## 7.1: Orthogonal Matrices

### Orthogonal Matrices

We say a square matrix is **orthogonal** if its transpose is its inverse, such that $AA^T = I$. A transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is an **orthogonal transformation** if $A$ is orthogonal.

**7.1.1** The following are equivalent for an $n \times n$ matrix $A$.
(a) $A$ is orthogonal.
(b) The column vectors of $A$ form an orthonormal set in $\mathbb{R}^n$ with the Euclidean inner product.
(c) The row vectors of $A$ form an orthonormal set in $\mathbb{R}^n$ with the Euclidean inner product.

**7.1.2** Transferability of orthogonality:
(a) The transpose of an orthogonal matrix is orthogonal.
(b) The inverse of an orthogonal matrix is orthogonal.
(c) A product of orthogonal matrices is orthogonal.
(c) If $A$ is orthogonal, then $\det(A) = 1$ or $\det(A) = -1$.

### Properties of Orthogonal Transformations

**7.1.3** The following are equivalent for an $n \times n$ matrix $A$.
(a) $A$ is orthogonal.
(b) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.
(c) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$.

### Change of Orthonormal Basis

**7.1.4** If $S$ is an orthonormal basis for an $n$-dimensional inner product space $V$, and

$$(\mathbf{u})_S = (u_1, u_2, \ldots, u_n) \text{ and } (\mathbf{v})_S = (v_1, v_2, \ldots, v_n)$$

then these familiar formulas are valid:
(a) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \ldots u_n^2}$
(b) $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$
(c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$

**7.1.5** Let $V$ be a finite-dimensional vector space. If $P$ is the transition matrix from one orthonormal basis for $V$ to another orthonormal basis for $V$, then $P$ is an orthogonal matrix.

# 7.2: Orthogonal Diagonalization

If $A$ and $B$ are square matrices, then $B$ is **orthogonally similar** to $A$ if there is an orthogonal matrix $P$ such that $B = P^T AP$. If $A$ is similar to some diagonal matrix $D$, as in $P^T AP = D$, then $A$ is **orthogonally diagonalizable** and $P$ **orthogonally diagonalizes** $A$.

## Conditions for Orthogonal Diagonalizability

**7.2.1** The following are equivalent for an $n \times n$ matrix $A$ with real entries.
(a) $A$ is orthogonally diagonalizable.
(b) $A$ has an orthonormal set of $n$ eigenvectors.
(c) $A$ is symmetric.

## Properties of Symmetric Matrices

**7.2.2** If $A$ is a symmetric matrix with real entries:
(a) The eigenvalues of $A$ are all real numbers.
(b) Eigenvectors from different eigenspaces are orthogonal.

How to orthogonally diagonalize a $n \times n$ symmetric matrix:

1. Find a basis for each eigenspace of $A$
2. Apply Gram-Schmidt to each base to obtain an orthonormal basis for each eigenspace
3. Form a matrix $P$ whose columns are the vectors from Step 2. This matrix will orthogonally diagonalize $A$. The eigenvalues on the diagonal of $D = P^T AP$ will be in the same order as the corresponding eigenvectors in $P$.

## Spectral Decomposition

By multiplying through $A = PDP^T$, we get

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

This expansion is called a **spectral decomposition** of $A$. The matrix $\mathbf{u}\mathbf{u}^T$ is the standard matrix for the projection of $\mathbb{R}^n$ on the subspace spanned by $\mathbf{u}$. Essentially, the spectral decomposition says that the result of transforming a vector $\mathbf{x}$ by multiplication by a matrix $A$ can be found by projecting $\mathbf{x}$ orthogonally

on the one-dimensional eigenvectors of $A$, scaling those projections by the corresponding eigenvalues, then adding them.

# The Nondiagonalizable Case

*Note: For this part, $\times$ will be used to represent an unspecified value.*

**7.2.3 Schur's Theorem:** If $A$ is an $n \times n$ matrix with real entries and real eigenvalues, then there is an orthogonal matrix $P$ such that $P^T A P$ is an upper triangular matrix of the form

$$P^T A P = \begin{bmatrix} \lambda_1 & \times & \times & \ldots & \times \\ 0 & \lambda_2 & \times & \ldots & \times \\ 0 & 0 & \lambda_3 & \ldots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $A$, potentially repeated according to multiplicity.

This matrix is often called $S$, so that the diagonalization equation is rewritten as $A = P^T S P$.

**7.2.3 Hessenberg's Theorem:** If $A$ is an $n \times n$ matrix with real entries, then there is an orthogonal matrix $P$ such that $P^T A P$ is a matrix of the form

$$P^T A P = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \ddots & \times & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \times & \times & \times \\ 0 & 0 & \cdots & 0 & \times & \times \end{bmatrix}$$

where unlike with Schur's Theorem, the eigenvalues are not guaranteed to be in $P$. Essentially, Hessenburg's Theorem says that every square matrix with real entries is orthogonally similar to a matrix in which each entry below the first **subdiagonal** is zero. The above matrix is more specifically known as **upper Hessenburg form**; the formula becomes $A = PHP^T$, and is called an **upper Hessenburg decomposition** of $A$.

# 7.3: Quadratic Forms

> ⚠️ **Warning**
>
> Be warned that this section may be lacking! I recommend using the textbook instead, because explaining 3D geometry without images is really hard.

# Definition of a Quadratic Form

Expressions of form $a_1x_2 + a_2x_2 + \cdots + a_nx_n$ are called **linear forms** of the **variables** $(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$. **Quadratic forms** are of the form $a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2 + $ (all terms of the form $a_kx_ix_j$ where $i \neq j$). The terms $a_kx_ix_j$ are known as **cross product terms**. Since there is one combination $ax_ix_j$ and one combination $ax_jx_i$ for two values of $i, j$, these are usually doubled to avoid repetition. Thus, a general quadratic form may look like this: $a_1x_1^2 + a_2x_2^2 + 2a_3x_1x_2$

For a symmetric $n \times n$ matrix $A$ and $n \times 1$ vector $\mathbf{x}$, $Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is the **quadratic form associated with** $A$. Expansion of this product will result in the quadratic form. Using properties of dot products, $\mathbf{x}^T A\mathbf{x}$ is also equal to $\mathbf{x} \cdot A\mathbf{x}$ and $A\mathbf{x} \cdot \mathbf{x}$.

If $A$ is diagonal, then the expanded quadratic form will have no cross product terms.

# Change of Variable in Quadratic Form

There are three main problems concerning quadratic forms:

1. If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on $R^2$ or $R^3$, what kind of curve or surface is represented by the equation $\mathbf{x}^T A\mathbf{x} = k$?
2. If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on $R^n$, what conditions must $A$ satisfy for $\mathbf{x}^T A\mathbf{x} = k$ to have positive values for $\mathbf{x} \neq 0$?
3. If $\mathbf{x}^T A\mathbf{x}$ is a quadratic form on $R^n$, what are its maximum and minimum values if $\|\mathbf{x}\| = 0$?

We can make a substitution $\mathbf{x} = P\mathbf{y}$ in order to express the variables $x_1, x_2, \ldots, x_n$ as $y_1, y_2, \ldots, y_n$. If $P$ is invertible, this is called **change of variable**. If $P$ is orthogonal, this is called **orthogonal change of variable**.

**7.2.4  The Principal Axes Theorem**: If $A$ is a symmetric $n \times n$ matrix, then there is an orthogonal change of variable that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross product terms. Specifically, if $P$ orthogonally

diagonalizes $A$, then making the change of variable $\mathbf{x} = P\mathbf{y}$ in the quadratic form $\mathbf{x}^T A\mathbf{x}$ yields the quadratic form

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $A$ corresponding to the eigenvectors that form the successive columns of $P$.

# Quadratic Forms in Geometry

A **conic section** is the curve that results when cutting a double-napped cone with a plane. The following sections may result:

○ Ellipse

○ Parabola

○ Hyperbola

A quadratic form in $\mathbb{R}^2$ arises in the equations of conic sections. If a conic section's equation has no linear terms, it is called a **central conic**. Parabolas cannot be central conics. Furthermore, if there are also no cross product terms, it is called a **central conic in standard position**. The general equation of a conic in standard position is $ax^2 + by^2 + f = 0$.

These are the important central conics in standard position:

○ $\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1, \alpha > 0, \beta > 0$
○ $\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1, \alpha > 0, \beta > 0$
○ $\frac{y^2}{\beta^2} + \frac{x^2}{\alpha^2} = 1, \alpha > 0, \beta > 0$

# Identifying Conic Sections

Approaching the first question stated earlier, to identify a conic from its quadratic form:

0. Assumes quadratic form is in general form $ax^2 + 2bxy + cy^2 = d$

1. Write conic in matrix form $\mathbf{x}^T A\mathbf{x} + d = 0$ where $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$

2. Find eigenvalues of $A$

3. Find orthonormal bases for the eigenspaces and arrange as columns of a matrix $P$.

4. If the determinant is -1, swap columns.

5. Arrange the eigenvalues, in corresponding order of their orthonormal bases, as columns in a diagonal matrix.

6. Substitute as change of base $\mathbf{x}'^T A \mathbf{x}'$ where $A$ is the matrix of eigenvalues, and expand.

7. Place in quadratic form without cross product $ax'^2 + by'^2 = -d$.

8. To find the angle rotated through to find this result, substitute
$$P = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$ This matrix is very useful.

# Positive Definite Quadratic Forms

We now consider the second question.

A quadratic form $\mathbf{x}^T A \mathbf{x}$ is said to be **positive definite** if $\mathbf{x}^T A \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$, **negative definite** if $\mathbf{x}^T A \mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$, and **indefinite** if $\mathbf{x}^T A \mathbf{x}$ has both positive and negative values.

**7.3.2** If $A$ is a symmetric matrix, then:
(a) $\mathbf{x}^T A \mathbf{x}$ is positive definite if and only if all eigenvalues of $A$ are positive.
(b) $\mathbf{x}^T A \mathbf{x}$ is negative definite if and only if all eigenvalues of $A$ are negative.
(c) $\mathbf{x}^T A \mathbf{x}$ is indefinite if and only if it has at least one eigenvalue of $A$ is positive and at least one is negative.

There are two more classifications, positive semidefinite if $\mathbf{x}^T A \mathbf{x} \geq 0$ for $\mathbf{x} \neq \mathbf{0}$ and negative semidefinite if $\mathbf{x}^T A \mathbf{x} \leq 0$ for $\mathbf{x} \neq \mathbf{0}$, but are less important.

Note: positive definite and negative definite matrices are invertible.

# Classifying Conic Sections Using Eigenvalues

If $\mathbf{x}^T A \mathbf{x} = k$ is an equation of a quadratic form, we can divide both sides by $k$ to make $\mathbf{x}^T B \mathbf{x} = 1$ where $B = (\frac{1}{k})A$. The general equation of the conic of this new form is $\lambda_1 x'^2 + \lambda_2 y'^2 = 1$, and by rearranging, we see that the lengths of the axes (from origin to collision of axis with the conic) are $2/\sqrt{\lambda_1}, 2/\sqrt{\lambda_2}$.

**7.3.3** If $A$ is a symmetric matrix, then:
(a) $\mathbf{x}^T A \mathbf{x} = 1$ represents an ellipse if $A$ is positive definite.
(b) $\mathbf{x}^T A \mathbf{x} = 1$ has no graph if $A$ is negative definite.
(c) $\mathbf{x}^T A \mathbf{x} = 1$ represents a hyperbola if $A$ is indefinite.

# Identifying Positive Definite Matrices

The **k-th principal submatrix** of an $n \times n$ matrix $A$ is the matrix with dimensions $k \times k$ consisting of the first $k$ rows and columns of $A$.

**7.3.4** If $A$ is a symmetric matrix, then:

(a) $A$ is positive definite if and only if the determinant of all principal submatrices are positive.

(b) $A$ is negative definite if and only if the determinants of the principal submatrices alternate between positive and negative, starting with negative.

(c) $A$ is indefinite if and only if it is neither positive definite or negative definite, and at least one principal submatrix has a positive determinant and one has a negative determinant.

# 7.4: Optimization Using Quadratic Forms

## Constrained Extremum Problems

This type of problem revolves around finding the maximum and minimum values of a quadratic form $\mathbf{x}^T A \mathbf{x}$ with the constraint $\|\mathbf{x}\| = 1$. Imagine an $xyz$ 3 dimensional system with the quadratic form on it, and imagine the constraint as a circle with radius 1 on the $xy$ plane. The maximum and minimum values of the form are the maximum and minimum $z$-values on the circle.

**7.3.4 Constrained Extremum Theorem:** Let $A$ be a symmetric $n \times n$ matrix whose eigenvalues in order of decreasing size are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:

(a) The quadratic form $\mathbf{x}^T A \mathbf{x}$ has a maximum value of $\lambda_1$ and a minimum value of $\lambda_n$, both of which are obtained on the set of vectors for which $\|\mathbf{x}\| = 1$.

(b) The maximum value of $\mathbf{x}^T A \mathbf{x}$ occurs at an eigenvector corresponding to the eigenvalue $\lambda_1$.

(c) The minimum value of $\mathbf{x}^T A \mathbf{x}$ occurs at an eigenvector corresponding to the eigenvalue $\lambda_n$.

## Constrained Extrema and Level Curves

Imagine an x-y-z plane with a function $f(x, y)$. At a plane $z = k$ for a constant $k$, the curve on the x-y plane formed through the intersection of the plane and the function is given by $f(x, y) = k$. These are called **level curves**.

The level curves of a quadratic form in $\mathbb{R}^2$ are given by $\mathbf{x}^T A \mathbf{x} = k$. The largest values of the quadratic form that intersect the unit circle in $\mathbb{R}^2$ will be the largest and smallest values of $k$ for which the graph of $\mathbf{x}^T A \mathbf{x} = k$ intersects the unit circle.

## Relative Extrema of Functions of Two Variables

If a function $f(x, y)$ has first order partial derivatives, then its relative maxima and minima occur wherever $f_x(x, y)$ and $f_y(x, y)$ are both equal to zero. These are

called **critical points** of $f$.

Let us define a function $D(x, y) = f(x, y) - f(x_0, y_0)$ where $(x, y)$ are very close, but different from $(x_0, y_0)$. It is hard to find the sign of $D$.

**7.4.3  Second Derivative Test:** Suppose that $(x_0, y_0)$ is a critical point of $f(x, y)$ and that $f$ has continuous second-order partial derivatives in some circular region centered at $(x_0, y_0)$. Then:

(a)  $f$ has a relative minimum at $(x_0, y_0)$ if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \text{ and } f_{xx}(x_0, y_0) > 0$$

(b)  $f$ has a relative maximum at $(x_0, y_0)$ if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \text{ and } f_{xx}(x_0, y_0) < 0$$

(c)  $f$ has a saddle point at $(x_0, y_0)$ if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) < 0$$

(d)  The test is inconclusive if

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) = 0$$

We can make this easier with the properties of symmetric matrices. We form a symmetric matrix

$$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

which is called the **Hessian matrix** of $f$. The determinant of the Hessian is

$$f_{xx}(x, y)f_{yy}(x, y) - f_{xy}(x, y)^2$$

which appears in **7.4.2**.

**7.4.4  Hessian Form of the Second Derivative Test:** Suppose that $(x_0, y_0)$ is a critical point of $f(x, y)$ and that $f$ has continuous second-order partial derivatives in some circular region centered at $(x_0, y_0)$. If $H(x_0, y_0)$ is the Hessian of $f$ at $(x_0, y_0)$, then:
(a)  $f$ has a relative minimum at $(x_0, y_0)$ if $H(x_0, y_0)$ is positive definite.
(b)  $f$ has a relative maximum at $(x_0, y_0)$ if $H(x_0, y_0)$ is negative definite.
(c)  $f$ has a saddle point at $(x_0, y_0)$ if $H(x_0, y_0)$ is indefinite.
(d)  The test is inconclusive otherwise.

# 7.5: Hermitian, Unitary, and Normal Matrices

# Hermitian and Unitary Matrices

If $A$ is a complex matrix, the **conjugate transpose** of $A$ is given by $A^* = \overline{A^T}$

**7.5.1** If $k$ is a complex scalar and $A$ and $B$ are complex matrices whose sizes are such that the stated operations can be performed, then:
(a) $(A^*)^* = A$
(b) $(A + B)^* = A^* + B^*$
(c) $(A - B)^* = A^* - B^*$
(d) $(kA)^* = \bar{k}A^*$
(e) $(AB)^* = B^*A^*$

A square matrix is said to be **unitary** if $AA^* = A^*A = I$ or equivalently $A^* = A^{-1}$, and **Hermitian** if $A^* = A$.

The diagonal entries of a Hermitian matrix must always be real, and corresponding entries across the main diagonal must be complex conjugates. The direct way to tell if a matrix is unitary is to see if it satisfies $A^* = A^{-1}$.

The relationship $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v}^T}\mathbf{u}$ from Chapter 5 can be said as $\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^*\mathbf{u}$.

**7.5.2** If $A$ is a Hermitian matrix, then:
(a) The eigenvalues of $A$ must be real numbers.
(b) Eigenvectors from different eigenspaces are orthogonal.

The following theorem provides a method to ascertain if a matrix is unitary without finding the inverse.

**7.5.3** If $A$ is a $n \times n$ matrix with complex entries, then the following are equivalent:
(a) $A$ is unitary.
(b) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{C}^n$.
(c) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{C}^n$.
(d) The column vectors of $A$ form a orthonormal set in $\mathbb{C}^n$ with respect to the complex Euclidean inner product.
(e) The row vectors of $A$ form a orthonormal set in $\mathbb{C}^n$ with respect to the complex Euclidean inner product.

# Unitary Diagonalizability

A square complex matrix $A$ is **unitarily diagonalizable** if there is a unitary matrix $P$ such that $P^*AP = D$ is a complex diagonalizable matrix. Any such matrix $P$ is said to **unitarily diagonalize** $A$.

**7.5.4** Every $n \times n$ Hermitian matrix $A$ has an orthonormal set of $n$ eigenvectors. $A$ is unitarily diagonalized by any $n \times n$ matrix $P$ whose column vectors form an orthonormal set of eigenvectors of $A$.

Unitarily diagonalizing a Hermitian matrix:

1. Find a basis for each eigenspace of $A$.
2. Apply the Gram–Schmidt process to each of these bases to obtain orthonormal bases for the eigenspaces.
3. Form the matrix $P$ whose column vectors are the basis vectors obtained in Step 2. It will be unitary and unitarily diagonalize $A$.

## Skew-Symmetric and Skew-Hermitian Matrices

A square real matrix $A$ is **skew-symmetric** if $A^T = -A$, and a square complex matrix $A$ is **skew-Hermitian** if $A^* = -A$.

Skew-symmetric matrices must have zeroes on the main diagonal, and skew-Hermitian matrices must have zeroes or pure complex numbers on the main diagonal.

## Normal Matrices

A square complex matrix $A$ is unitarily diagonalizable if and only if $AA^* = A^*A$. Matrices with this property are called **normal**. The following types of matrices are normal:

- Complex matrices
  - Hermitian matrices
  - Skew-Hermitian matrices
  - Unitary matrices
- Real matrices
  - Symmetric matrices
  - Skew-symmetric matrices
  - Orthogonal matrices

# 8. General Linear Transformations

## 8.1: General Linear Transformations

If $T : V \to W$ is a mapping from a vector space $V$ to a vector space $W$, then $T$ is called a **linear transformation** from $V$ to $W$ if the following two properties hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $V$ and all scalars $k$:

- $T(k\mathbf{u}) = kT(\mathbf{u})$    [**Homogenity property**]
- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$    [**Additivity property**]

  In the special case $V = W$, the linear transformation $T$ is called a **linear operator** on $V$.

**8.1.1** If $T : V \to W$ is a linear transformation, then:
(a) $T(\mathbf{0}) = \mathbf{0}$
(b) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in $V$.
(c) $T(-\mathbf{u}) = -T(\mathbf{u})$ for all $\mathbf{u}$ in $V$.

The mapping $T : V \to W$ defined by $T(\mathbf{v}) = \mathbf{0}$ for every $\mathbf{v}$ in $V$ is called the **zero transformation**. $I : V \to V$ defined by $I(\mathbf{v}) = \mathbf{v}$ is called the **identity operator** on $V$.

The linear operator $T : V \to V$ defined by $T(\mathbf{x}) = c\mathbf{x}$ for every $\mathbf{x}$ in $V$ and scalar $c$ is called the **contraction of $V$** if $0 < c < 1$, and the **dilation of $V$** if $c > 1$.

# Finding Linear Transformations from Images of Basis Vectors

**8.1.2** Let $T : V \to W$ is a linear transformation, for which the vector space $V$ is finite-dimensional. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for $V$, then the image of any vector $\mathbf{v}$ in $V$ can be expressed as:

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \cdots + c_n T(\mathbf{v}_n)$$

where $c_1, c_2, \ldots, c_n$ are the coefficients required to express $\mathbf{v}$ as a linear transformation of the vectors in basis $S$.

# Kernel and Range

If $T : V \to W$ is a linear transformation, then the set of vectors in $V$ that $T$ maps into $\mathbf{0}$ is called the **kernel** of $T$ and is denoted by $\ker(T)$. The set of all vectors in $W$ that are images under $T$ of at least one vector in $V$ is called the **range** of $T$ and is denoted by $R(T)$.

- If $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is multiplication by the $m \times n$ matrix $A$, then the kernel of $T_A$ is the null space of $A$ and the range is the column space of $A$.
- Let $T : V \to W$ be the zero transformation. $\ker(T) = V$, and $R(T) = \{\mathbf{0}\}$.

- Let $I : V \to V$ be the identity operator. $\ker(I) = \{\mathbf{0}\}$, and $R(I) = V$.
- Let $T : R^3 \to R^3$ be the orthogonal projection on the $xy$ plane. $\ker(T)$ is the set of points of form $(0, 0, z)$, and $R(T)$ is the set of points of form $(x, y, 0)$.
- Let $T : R^2 \to R^2$ be the linear operator that rotates every vector in the $xy$ plane through the angle $\theta$. $\ker(T) = \{\mathbf{0}\}$, and $R(T) = R^2$.

**8.1.3** Let $T : V \to W$ is a linear transformation.
(a) The kernel of $T$ is a subspace of $W$.
(b) The range of $T$ is a subspace of $W$.

# Rank and Nullity of Linear Transformations

Let $T : V \to W$ be a linear transformation. In the case that the range of $T$ is finite-dimensional its dimension is called the **rank of $T$**, and if the kernel is finite-dimensional, its dimension is called the **nullity of $T$**.

**8.1.4** Let $T : V \to W$ is a linear transformation with two finite-dimensional vector spaces $V$ and $W$, then the range of $T$ is finite-dimensional, and $\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V)$.

# 8.2: Compositions and Inverse Transformations

## One-to-One and Onto

If $T : V \to W$ is a linear transformation, then $T$ is **one-to-one** if $T$ maps distinct vectors in $V$ into distinct vectors in $W$.

If $T : V \to W$ is a linear transformation, then $T$ is **onto** $W$ if every vector in $W$ is an image of at least one vector in $V$.

**8.2.1** If $T : V \to W$ is a linear transformation, the following statements are equivalent.
(a) $T$ is one-to-one.
(b) $\ker(T) = \{\mathbf{0}\}$.

|  | One-to-One | Onto |
|---|---|---|
| Rotation operators on $\mathbb{R}^n$ | Yes | Yes |
| Orthogonal projections on $\mathbb{R}^n$ | No | No |
| $T : P_3 \to R^4$ $T(a + bx + cx^2 + dx^3) = (a, b, c, d)$ | Yes | Yes |

| | One-to-One | Onto |
|---|---|---|
| $T : M_{22} \to R^4$ $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a, b, c, d)$ | Yes | Yes |
| $T : P_n \to P_{n+1}$ $T(\mathbf{p}) = T(p(x)) = xp(x)$ | Yes | No |
| $D : \mathbb{C}^1(-\infty, \infty) \to F(-\infty, \infty)$ | No | Yes |

*proving these results is left as an exercise, or just read the examples in the textbook if you want. i'm not a cop*

**8.2.2** If $T : V \to W$ is a linear transformation, the following statements are equivalent.
(a) $T$ is one-to-one.
(b) $\ker(T) = \{\mathbf{0}\}$.
(c) $T$ is onto $(R(T) = W)$.

# Matrix Transformations Revisited

**8.2.3** If $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then
(a) $T_A$ is one-to-one if and only if the columns of $A$ are linearly independent.
(b) $T_A$ is onto if and only if the columns of $A$ span $\mathbb{R}^n$.

**8.2.4** For square matrix $A$ with no duplicate rows/columns, all of these statements are equivalent:
(a) $A$ is invertible.
(b) $A\mathbf{x} = 0$ has only the trivial solution.
(c) The reduced row echelon form of $A$ is $I_n$.
(d) $A$ can be expressed as a product of elementary matrices.
(e) $A\mathbf{x} = b$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
(f) $A\mathbf{x} = b$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
(g) $\det(A) \neq 0$.
(h) The column vectors of $A$ are linearly independent.
(i) The row vectors of $A$ are linearly independent.
(j) The column vectors of $A$ span $\mathbb{R}^n$.
(k) The row vectors of $A$ span $\mathbb{R}^n$.
(l) The column vectors of $A$ form a basis for $\mathbb{R}^n$.
(m) The row vectors of $A$ form a basis for $\mathbb{R}^n$.

(n) $A$ has rank $n$.

(o) $A$ has nullity 0.

(p) The orthogonal complement of the null space of $A$ is $\mathbb{R}^n$.

(q) The orthogonal complement of the row space of $A$ is $\{0\}$.

(r) $\lambda = 0$ is not an eigenvalue of $A$.

(s) $A^T A$ is invertible.

(t) The kernel of $T_A$ is $\{\mathbf{0}\}$.

(u) The range of $T_A$ is $\mathbb{R}^n$.

(v) $T_A$ is one-to-one.

## Inverse Linear Transformations

The inverse of multiplication by $A$ is multiplication by $A^{-1}$.

If $T : V \to W$ is a linear transformation and $\mathbf{w}$ is a vector in $R(T)$ (the range of $T$), the fact that $T$ is one to one means that there exists a transformation using a vector $\mathbf{v}$ from $V$ such that $T(\mathbf{v}) = \mathbf{w}$.

Now we define a transformation $T^{-1}$, that maps $\mathbf{w} \to \mathbf{v}$. Thus, $T^{-1} : R(T) \to V$ is a linear transformation. Note that $T$ and $T^{-1}$ cancel one another out when used in succession.

## Composition of Linear Transformations

If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then the **composition** of $T_2$ with $T_1$ is denoted $T_2 \circ T_1$ and defined by $(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u}))$ where $\mathbf{u}$ is a function in $U$.

**8.2.5** If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then $(T_2 \circ T_1) : U \to W$ is also a linear transformation.

**8.2.6** If $T_1 : U \to V$ and $T_2 : V \to W$ are one-to-one linear transformations, then

(a) $T_2 \circ T_1$ is one-to-one.

(b) $(T_2 \circ T_1)^{-1} = T_2^{-1} \circ T_1^{-1}$.

# 8.3: Isomorphism

## Isomorphism

A linear transformation $T : V \to W$ that is both one-to-one and onto is said to be an **isomorphism**, and $W$ is **isomorphic** to $V$.

Essentially, isomorphism means that two different vector spaces share the same algebraic form and their vector operations are conducted in the same fashion, even

if the types of objects in the spaces may be different.

**8.3.1** Every $n$-dimensional vector space is isomorphic to $\mathbb{R}^n$.

**8.3.2** If $S$ is an ordered basis for a vector space $V$, then the coordinate map $\mathbf{u} \xrightarrow{T} (\mathbf{u})_S$ is an isomorphism between $V$ and $\mathbb{R}^n$.

## Inner Product Space Isomorphisms

Regular isomorphisms only preserve algebraic structure. To preserve geometric structure, an isomorphism will also preserve inner products. We define $T : V \to W$ an **inner product space isomorphism** if $\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v}$ in $V$.

**8.3.3** If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an ordered orthonormal basis for a real $n$-dimensional inner product space $V$, then the coordinate map

$$\mathbf{u} \xrightarrow{T} (\mathbf{u})_s$$

is an inner product space isomorphism between $V$ and the vector space $\mathbb{R}^n$ with the Euclidean inner product.

# 8.4: Matrices for General Linear Transformations