# Linear Algebra

## SYSTEMS OF LINEAR EQUATIONS AND MATRICES

## 1 Systems of Linear Equations

**Linear Equation.** A linear equation in the variables $x_1, x_2, \dots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, a_2, \dots, a_n, b \in \mathbb{R}$ are constants.

**System of Linear Equations.** A system of $m$ linear equations in $n$ unknowns is a collection

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

This can be written in matrix form as $Ax = b$, where $A = [a_{ij}]$ is the $m \times n$ coefficient matrix, $x = [x_1, \dots, x_n]^T$, and $b = [b_1, \dots, b_m]^T$.

**Augmented Matrix.** The augmented matrix of the system $Ax = b$ is

$$[A \mid b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

**Elementary Row Operations (EROs).**

(i) **Row Swap:** $R_i \leftrightarrow R_j$ — interchange rows $i$ and $j$.

(ii) **Row Scaling:** $R_i \to cR_i$ — multiply row $i$ by a nonzero scalar $c$.

(iii) **Row Replacement:** $R_i \to R_i + cR_j$ — add $c$ times row $j$ to row $i$.

**Row Echelon Form (REF).** A matrix is in REF if:

1. All zero rows are at the bottom.
2. The leading entry (pivot) of each nonzero row is to the right of the pivot in the row above.
3. All entries below a pivot are zero.

**Reduced Row Echelon Form (RREF).** A matrix is in RREF if it is in REF and additionally:

1. Each pivot is 1.
2. Each pivot is the only nonzero entry in its column.

**Theorem 1.** Every matrix has a unique RREF.

**Gaussian Elimination.** To solve $Ax = b$:

1. Form the augmented matrix $[A \mid b]$.
2. Use EROs to reduce to REF (or RREF).
3. Back-substitute (if REF) or read off the solution (if RREF).

**Pivot and Free Variables.** Variables corresponding to pivot columns are called *pivot variables* (or basic variables). All other variables are *free variables*. Free variables can take any value and parametrize the solution set.

**Consistency.** A system $Ax = b$ is *consistent* if it has at least one solution, and *inconsistent* if it has no solution. A system is inconsistent if and only if the RREF of $[A \mid b]$ has a row of the form $[0\ 0\ \cdots\ 0 \mid c]$ with $c \neq 0$.

**Theorem 2.** A system of linear equations has either:

1. No solution (inconsistent),
2. Exactly one solution (consistent, no free variables), or
3. Infinitely many solutions (consistent, at least one free variable).

**Homogeneous Systems.** A system $Ax = \mathbf{0}$ is called homogeneous. It is always consistent since $x = \mathbf{0}$ is always a solution (the *trivial solution*). A homogeneous system has a nontrivial solution if and only if it has a free variable.

**Theorem 3.** If $A$ is an $m \times n$ matrix with $m < n$ (more unknowns than equations), then $Ax = \mathbf{0}$ has infinitely many solutions.

**Example.** Solve the system:

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 3 \\ 2x_1 + 4x_2 + 3x_3 &= 8 \\ 3x_1 + 6x_2 + 5x_3 &= 14 \end{aligned}$$

*Solution.* Form the augmented matrix and row reduce:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 4 & 3 & 8 \\ 3 & 6 & 5 & 14 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 2 \\ 3 & 6 & 5 & 14 \end{array}\right] \xrightarrow{R_3 - 3R_1} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 5 \end{array}\right]$$

$$\xrightarrow{R_3 - 2R_2} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The last row gives $0 = 1$, so the system is **inconsistent** (no solution).

**Example.** Solve $Ax = \mathbf{0}$ where $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{bmatrix}$.

*Solution.* $\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & 0 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$.

Free variables: $x_2 = s$, $x_3 = t$. Solution: $x_1 = -2s + t$, so

$$x = \begin{bmatrix} -2s + t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

**Example.** Solve the system:

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

*Solution.* Row reduce the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{array}\right] \xrightarrow{R_3 - 5R_1} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{array}\right] \xrightarrow{R_3 - 5R_2} \left[\begin{array}{cc} 1 & -2 \\ 0 & 2 \\ 0 & 0 \end{array}\right.$$

Back-substitute: $x_3 = -1$, $x_2 = \frac{8+8(-1)}{2} = 0$, $x_1 = 2(0) - (-1) = 1$. The system has a **unique solution**: $(x_1, x_2, x_3) = (1, 0, -1)$.

**Example.** Solve the system:

$$\begin{aligned} x_1 + 3x_2 + x_3 + x_4 &= 3 \\ 2x_1 + 6x_2 + 4x_3 + 2x_4 &= 6 \end{aligned}$$

*Solution.* $\left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 3 \\ 2 & 6 & 4 & 2 & 6 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 3 \\ 0 & 0 & 2 & 0 & 0 \end{array}\right] \xrightarrow{\frac{1}{2}R_2}$
$\left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 3 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right]$

$$\xrightarrow{R_1 - R_2} \left[\begin{array}{cccc|c} 1 & 3 & 0 & 1 & 3 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right].$$

Free variables: $x_2 = s$, $x_4 = t$. Solution: $x_3 = 0$, $x_1 = 3 - 3s - t$, so

$$x = \begin{bmatrix} 3 - 3s - t \\ s \\ 0 \\ t \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

The system is consistent with **infinitely many solutions**. Note the general solution has the form $x = x_p + x_h$, where $x_p$ is a particular solution and $x_h \in \text{null}(A)$.

## 2 Matrix Algebra

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

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

We write $A = [a_{ij}]_{m\times n}$.

**Matrix Operations.** Let $A, B$ be $m \times n$ matrices and $c \in \mathbb{R}$.

1. **Addition:** $(A + B)_{ij} = a_{ij} + b_{ij}$.
2. **Scalar Multiplication:** $(cA)_{ij} = c \cdot a_{ij}$.
3. **Matrix Multiplication:** If $A$ is $m \times n$ and $B$ is $n \times p$, then $AB$ is $m \times p$ with $(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$.

**Properties of Matrix Arithmetic.** Let $A, B, C$ be matrices of compatible sizes and $r, s \in \mathbb{R}$.

1. $A + B = B + A$ (commutativity of addition).
2. $(A + B) + C = A + (B + C)$ (associativity of addition).
3. $A(BC) = (AB)C$ (associativity of multiplication).
4. $A(B + C) = AB + AC$ (left distributivity).
5. $(A + B)C = AC + BC$ (right distributivity).
6. $r(sA) = (rs)A, \quad r(A + B) = rA + rB, \quad (r + s)A = rA + sA$.

**Warning.** In general, $AB \neq BA$. Matrix multiplication is **not commutative**.

**Transpose.** The transpose of $A = [a_{ij}]_{m\times n}$ is $A^T = [a_{ji}]_{n\times m}$.
Properties: $(A^T)^T = A$, $(A + B)^T = A^T + B^T$, $(cA)^T = cA^T$, $(AB)^T = B^T A^T$.

**Symmetric Matrix.** $A$ is symmetric if $A^T = A$. That is, $a_{ij} = a_{ji}$ for all $i, j$.

**Identity Matrix.** The $n \times n$ identity matrix $I_n$ has 1s on the diagonal and 0s elsewhere: $(I_n)_{ij} = \delta_{ij}$. For any $m \times n$ matrix $A$: $I_m A = A$ and $AI_n = A$.

**Invertible Matrices.** An $n \times n$ matrix $A$ is *invertible* (nonsingular) if there exists an $n \times n$ matrix $B$ such that $AB = BA = I_n$. We write $B = A^{-1}$.
**Theorem 4.** If $A$ is invertible, then $A^{-1}$ is unique.

*Proof.* Suppose $B$ and $C$ are both inverses of $A$, so $AB = BA = I$ and $AC = CA = I$. Then $B = BI = B(AC) = (BA)C = IC = C$. □

**Properties of Inverses.** If $A$ and $B$ are invertible $n \times n$ matrices:

1. $(A^{-1})^{-1} = A$.
2. $(AB)^{-1} = B^{-1}A^{-1}$.
3. $(A^T)^{-1} = (A^{-1})^T$.
4. $(cA)^{-1} = \frac{1}{c}A^{-1}$ for $c \neq 0$.

**Computing $A^{-1}$.** Row reduce $[A \mid I_n]$. If $A$ is invertible, the result is $[I_n \mid A^{-1}]$.

**Example.** Find $A^{-1}$ for $A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$.
*Solution.*

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 7 & 0 & 1 \end{array}\right] \xrightarrow{R_2 - 3R_1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & -3 & 1 \end{array}\right] \xrightarrow{R_1 - 2R_2} \left[\begin{array}{cc|cc} 1 & 0 & 7 & -2 \\ 0 & 1 & -3 & 1 \end{array}\right]$$

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

$2 \times 2$ **Inverse Formula.** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad - bc \neq 0$, then

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

**Elementary Matrices.** An elementary matrix $E$ is obtained by performing one ERO on $I_n$.

1. Every ERO on $A$ can be achieved by left-multiplying $A$ by the corresponding $E$.
2. Every elementary matrix is invertible, and its inverse is also an elementary matrix.

**Theorem 5.** An $n \times n$ matrix $A$ is invertible if and only if $A$ is row equivalent to $I_n$. Equivalently, $A$ is a product of elementary matrices.

## 3 LU and PLU Factorization

**Permutation Matrix.** A permutation matrix $P$ is obtained by rearranging the rows of $I_n$. Properties: $P^T = P^{-1}$, $\det(P) = \pm 1$, and the product of permutation matrices is a permutation matrix.

**LU Factorization.** An $m \times n$ matrix $A$ has an LU factorization if $A = LU$, where $L$ is a unit lower triangular matrix (1s on the diagonal) and $U$ is in REF.
**Theorem 6.** An $m \times n$ matrix $A$ has an LU factorization if and only if $A$ can be reduced to REF using only row replacement operations $R_i \to R_i + cR_j$ with $i > j$ (no row swaps or scaling).

*Proof sketch.* Each row replacement $R_i \to R_i + cR_j$ $(i > j)$ corresponds to left-multiplication by an elementary matrix $E$ that is unit lower triangular. Its inverse $E^{-1}$ is also unit lower triangular (replace $c$ by $-c$). If $E_k \cdots E_2 E_1 A = U$, then $A = E_1^{-1}E_2^{-1} \cdots E_k^{-1}U = LU$, and the product of unit lower triangular matrices is unit lower triangular. □

**Computing LU.** Perform Gaussian elimination (row replacements only) to obtain $U$. Record the multipliers: if we subtract $m_{ij}$ times row $j$ from row $i$, then $L_{ij} = m_{ij}$.

**Example.** Find the LU factorization of $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 3 & 3 \\ 8 & 7 & 9 \end{bmatrix}$.
*Solution.* $R_2 - 2R_1$, then $R_3 - 4R_1$, then $R_3 - 3R_2$:

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

Verify: $LU = A$. ✓

**Solving $Ax = b$ via LU.** Given $A = LU$:

1. Solve $Ly = b$ by forward substitution.
2. Solve $Ux = y$ by back substitution.

This is efficient when solving $Ax = b$ for multiple right-hand sides $b$, since $L$ and $U$ are computed once.

**PLU Factorization.** If row swaps are needed, we use the PLU factorization $PA = LU$ (equivalently $A = P^T LU$), where $P$ is a permutation matrix.
**Theorem 7.** Every $m \times n$ matrix $A$ has a PLU factorization $PA = LU$.

*Proof.* Gaussian elimination with partial pivoting (swapping rows to place the largest entry in the pivot position) always produces REF. Record all row swaps in $P$ and all multipliers in $L$. The permutation matrix $P$ encodes the row interchanges needed before elimination proceeds without further swaps, yielding $PA = LU$. □

**Example.** Find the PLU factorization of $A = \begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix}$.

*Solution.* Swap $R_1 \leftrightarrow R_2$ first: $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $PA = \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}$ is already in REF, so $L = I$ and $U = \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}$.

## 4 Determinants

**Determinant.** The determinant of a square matrix $A$, written $\det(A)$ or $|A|$, is a scalar value that encodes important information about $A$.

**$2\times 2$ Determinant.**

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

**$3\times 3$ Determinant (Sarrus' Rule).**

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

**Cofactor Expansion.** Let $A$ be an $n \times n$ matrix. The $(i,j)$-minor $M_{ij}$ is the determinant of the $(n-1)\times(n-1)$ matrix obtained by deleting row $i$ and column $j$. The $(i,j)$-cofactor is $C_{ij} = (-1)^{i+j} M_{ij}$. Then:

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along row } i\text{)},$$

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along column } j\text{)}.$$

**Example.** Compute $\det(A)$ for $A = \begin{bmatrix} 2 & 1 & 3 \\ 0 & -1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$.

*Solution.* Expanding along row 1:

$$\begin{aligned} \det(A) &= 2\det\begin{bmatrix} -1 & 2 \\ 4 & -1 \end{bmatrix} - 1\det\begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} + 3\det\begin{bmatrix} 0 & -1 \\ 1 & 4 \end{bmatrix} \\ &= 2(1-8) - 1(0-2) + 3(0+1) \\ &= 2(-7) - 1(-2) + 3(1) = -14 + 2 + 3 = -9. \end{aligned}$$

**Properties of Determinants.**

1. $\det(A^T) = \det(A)$.
2. If $A$ has a row (or column) of zeros, then $\det(A) = 0$.
3. If $A$ has two identical rows (or columns), then $\det(A) = 0$.
4. If $A$ is triangular (upper or lower), then $\det(A) = \prod_{i=1}^{n} a_{ii}$.
5. $\det(AB) = \det(A)\det(B)$.
6. $\det(A^{-1}) = \frac{1}{\det(A)}$ (when $A$ is invertible).
7. $\det(cA) = c^n \det(A)$ for an $n \times n$ matrix $A$.

**Example.** Compute $\det(A)$ using EROs for $A = \begin{bmatrix} 2 & -1 & 3 \\ 4 & 1 & 0 \\ -2 & 3 & -3 \end{bmatrix}$.

*Solution.* $R_2 - 2R_1$, $R_3 + R_1$: $\begin{bmatrix} 2 & -1 & 3 \\ 0 & 3 & -6 \\ 0 & 2 & 0 \end{bmatrix}$. $R_3 - \frac{2}{3}R_2$: $\begin{bmatrix} 2 & -1 & 3 \\ 0 & 3 & -6 \\ 0 & 0 & 4 \end{bmatrix}$.
Triangular matrix $\Rightarrow \det = 2 \cdot 3 \cdot 4 = 24$. No row swaps, so $\det(A) = 24$.

**Effect of EROs on Determinants.**

1. $R_i \leftrightarrow R_j$: det changes sign.
2. $R_i \to cR_i$: det is multiplied by $c$.
3. $R_i \to R_i + cR_j$: det is unchanged.

**Theorem 8.** *$A$ is invertible if and only if $\det(A) \neq 0$.*

*Proof.* ($\Rightarrow$) If $A$ is invertible, then $1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1})$, so $\det(A) \neq 0$. ($\Leftarrow$) If $\det(A) \neq 0$, then $A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$ exists (verified by direct computation: $A \cdot \operatorname{adj}(A) = \det(A)I$). ☐

**Remark.** The proof of $\det(AB) = \det(A)\det(B)$ is more involved. One approach: each ERO corresponds to left-multiplication by an elementary matrix $E$ with known determinant. Write $A$ as a product of elementary matrices (if invertible), and use the multiplicativity of determinants for elementary matrices. If $A$ is singular, then $AB$ is also singular, so both sides are 0.

**Adjugate (Classical Adjoint).** The adjugate of $A$ is $\operatorname{adj}(A) = [C_{ij}]^T$, where $C_{ij}$ are the cofactors. Then:

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

**Cramer's Rule.** If $A$ is an invertible $n \times n$ matrix, the unique solution to $Ax = b$ is

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad i = 1, 2, \ldots, n,$$

where $A_i$ is the matrix $A$ with column $i$ replaced by $b$.

**Example.** Use Cramer's Rule to solve $\begin{cases} 2x + y = 5 \\ x - y = 1 \end{cases}$.

*Solution.* $A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$, $\det(A) = -3$.

$$x = \frac{\det\begin{bmatrix} 5 & 1 \\ 1 & -1 \end{bmatrix}}{-3} = \frac{-6}{-3} = 2, \quad y = \frac{\det\begin{bmatrix} 2 & 5 \\ 1 & 1 \end{bmatrix}}{-3} = \frac{-3}{-3} = 1.$$

**Geometric Interpretation.** For a $2 \times 2$ matrix $A$, $|\det(A)|$ equals the area of the parallelogram formed by the column vectors. For a $3 \times 3$ matrix $A$, $|\det(A)|$ equals the volume of the parallelepiped formed by the column vectors.

# VECTOR SPACES

## 5 Vector Spaces

**Vector Space.** Let $V$ be a nonempty set of objects called vectors, and let two operations be defined: vector addition and scalar multiplication, where addition of two vectors in $V$ gives a vector in $V$, and multiplication of a scalar with a vector in $V$ gives a vector in $V$. Then, $V$ is a vector space if and only if the following axioms hold for all $u, v, w \in V$ and for all scalars $c, d$:

1. $u + v \in V$ (closure under addition).
2. $u + v = v + u$ (commutativity).
3. $u + (v + w) = (u + v) + w$ (associativity of addition).
4. $\exists\, \mathbf{0} \in V$ s.t. $u + \mathbf{0} = \mathbf{0} + u = u$ (additive identity).
5. $\forall u \in V$, $\exists\, (-u) \in V$ s.t. $u + (-u) = (-u) + u = \mathbf{0}$ (additive inverse).
6. $cu \in V$ (closure under scalar multiplication).
7. $c(u + v) = cu + cv$ (distributive over vector addition).
8. $(c + d)u = cu + du$ (distributive over scalar addition).
9. $c(du) = (cd)u$ (associativity of scalar multiplication).
10. $1u = u$ (scalar multiplicative identity).

**Example.** Important vector spaces:

- $\mathbb{R}^n$: the set of all $n$-tuples of real numbers.
- $M_{m\times n}(\mathbb{R})$: the set of all $m \times n$ matrices with real entries.
- $P_n(\mathbb{R})$: the set of all polynomials of degree $\leq n$ with real coefficients.
- $C^n[a, b]$: the set of all functions with continuous derivatives up to order $n$ on $[a, b]$.

**Theorem 9.** *Let $V$ be a vector space. Then,*

1. $0v = \mathbf{0}$, for all $v \in V$.
2. $c\mathbf{0} = \mathbf{0}$, for all scalars $c$.
3. $(-1)v = -v$, for all $v \in V$.
4. The zero vector $\mathbf{0}$ is unique.
5. For each $v \in V$, the additive inverse $-v$ is unique.
6. If $cv = \mathbf{0}$, then $c = 0$ or $v = \mathbf{0}$.

*Proof of (1).* We have $0v = (0 + 0)v = 0v + 0v$ by Axiom 8. Adding $-(0v)$ to both sides: $\mathbf{0} = 0v$. ☐

*Proof of (3).* We must show $(-1)v$ is the additive inverse of $v$, i.e., $v + (-1)v = \mathbf{0}$. By Axiom 10 and 8: $v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = \mathbf{0}$ by (1). Since additive inverses are unique, $(-1)v = -v$. ☐

*Proof of (4).* Suppose $\mathbf{0}_1$ and $\mathbf{0}_2$ are both zero vectors. Then $\mathbf{0}_1 = \mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2$, where the first equality uses $\mathbf{0}_2$ as the identity and the second uses $\mathbf{0}_1$ as the identity. ☐

*Proof of (6).* Suppose $cv = \mathbf{0}$ and $c \neq 0$. Then $c^{-1}$ exists, so $v = 1v = (c^{-1}c)v = c^{-1}(cv) = c^{-1}\mathbf{0} = \mathbf{0}$ by (2). ☐

## 6 Subspaces

**Subspace.** Let $V$ be a vector space. A subset $W \subseteq V$ is a subspace of $V$ if $W$ is itself a vector space under the same operations of $V$.

**Subspace Test.** Let $W \subseteq V$. Then $W$ is a subspace of $V$ if and only if:

1. $\mathbf{0} \in W$ (contains the zero vector).
2. For all $u, v \in W$, $u + v \in W$ (closed under addition).
3. For all $u \in W$ and $c \in \mathbb{R}$, $cu \in W$ (closed under scalar multiplication).

Equivalently (combining 2 and 3): $W \neq \varnothing$, $\mathbf{0} \in W$, and for all $u, v \in W$ and $c, d \in \mathbb{R}$, $cu + dv \in W$.

**Example.** Show that $W = \{(x, y, z) \in \mathbb{R}^3 \mid x + 2y - z = 0\}$ is a subspace of $\mathbb{R}^3$.
*Solution.* (1) $(0,0,0)$: $0 + 2(0) - 0 = 0$. ✓ (2) Let $u = (x_1, y_1, z_1)$, $v = (x_2, y_2, z_2) \in W$. Then $u + v = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$ and $(x_1 + x_2) + 2(y_1 + y_2) - (z_1 + z_2) = (x_1 + 2y_1 - z_1) + (x_2 + 2y_2 - z_2) = 0 + 0 = 0$, so $u + v \in W$. ✓ (3) For $c \in \mathbb{R}$: $cu = (cx_1, cy_1, cz_1)$ and $cx_1 + 2(cy_1) - (cz_1) = c(x_1 + 2y_1 - z_1) = 0$, so $cu \in W$. ✓
Alternatively, $W = \text{null}\begin{bmatrix} 1 & 2 & -1 \end{bmatrix}$, which is automatically a subspace.

**Null Space.** Let $A$ be an $m \times n$ matrix. The null space of $A$ is

$$\text{null}(A) = \{x \in \mathbb{R}^n \mid Ax = \mathbf{0}\}.$$

**Theorem 10.** If $A$ is an $m \times n$ matrix, then null$(A)$ is a subspace of $\mathbb{R}^n$.

*Proof.* We verify the three conditions of the subspace test.

1. $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in \text{null}(A)$.
2. Let $u, v \in \text{null}(A)$. Then $A(u + v) = Au + Av = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $u + v \in \text{null}(A)$.
3. Let $u \in \text{null}(A)$ and $c \in \mathbb{R}$. Then $A(cu) = c(Au) = c\mathbf{0} = \mathbf{0}$, so $cu \in \text{null}(A)$.

Therefore null$(A)$ is a subspace of $\mathbb{R}^n$. □

**Theorem 11.** If $W_1$ and $W_2$ are subspaces of $V$, then $W_1 \cap W_2$ is also a subspace of $V$.

*Proof.* Since $\mathbf{0} \in W_1$ and $\mathbf{0} \in W_2$, we have $\mathbf{0} \in W_1 \cap W_2$. If $u, v \in W_1 \cap W_2$ and $c \in \mathbb{R}$, then $u, v \in W_1$ implies $cu + v \in W_1$ (since $W_1$ is a subspace), and similarly $cu + v \in W_2$. So $cu + v \in W_1 \cap W_2$. □

**Note.** $W_1 \cup W_2$ is generally **not** a subspace of $V$. For example, in $\mathbb{R}^2$, let $W_1 = \text{span}\{e_1\}$ (the $x$-axis) and $W_2 = \text{span}\{e_2\}$ (the $y$-axis). Then $e_1 \in W_1 \cup W_2$ and $e_2 \in W_1 \cup W_2$, but $e_1 + e_2 = (1, 1) \notin W_1 \cup W_2$.

## 7 Spanning Sets

**Linear Combination.** Let $V$ be a vector space and $v_1, v_2, \ldots, v_n \in V$. A vector $v \in V$ is a linear combination of $v_1, \ldots, v_n$ if there exist scalars $c_1, c_2, \ldots, c_n$ such that

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

**Span.** The span of $\{v_1, v_2, \ldots, v_n\}$ is the set of all linear combinations:

$$\text{span}\{v_1, v_2, \ldots, v_n\} = \{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n \mid c_1, \ldots, c_n \in \mathbb{R}\}.$$

**Spanning Set.** If $V = \text{span}\{v_1, \ldots, v_n\}$, we say that $\{v_1, \ldots, v_n\}$ is a spanning set for $V$, or that $\{v_1, \ldots, v_n\}$ spans $V$.

**Theorem 12.** Let $V$ be a vector space and $S = \{v_1, \ldots, v_n\} \subseteq V$. Then span$(S)$ is a subspace of $V$.

*Proof.* (1) $\mathbf{0} = 0v_1 + \cdots + 0v_n \in \text{span}(S)$. (2) Let $u = \sum a_i v_i$ and $w = \sum b_i v_i$ be in span$(S)$. Then $u + w = \sum (a_i + b_i) v_i \in \text{span}(S)$. (3) For $c \in \mathbb{R}$: $cu = \sum (ca_i) v_i \in \text{span}(S)$. □

**Example.** span$\{(1, 0), (0, 1)\} = \mathbb{R}^2$.
span$\{1, x, x^2, \ldots, x^n\} = P_n(\mathbb{R})$.
$\text{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} = M_{2\times 2}(\mathbb{R}).$

## 8 Linear Dependence and Independence

**Linearly Dependent (LD).** A set $\{v_1, v_2, \ldots, v_n\}$ is LD if there exist scalars $c_1, c_2, \ldots, c_n$, **not all zero**, such that

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

**Linearly Independent (LI).** A set $\{v_1, v_2, \ldots, v_n\}$ is LI if

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

That is, the only linear combination giving $\mathbf{0}$ is the trivial one.

**Checking LI/LD in $\mathbb{R}^n$.** Form the matrix $A = [v_1 \; v_2 \; \cdots \; v_n]$ and row reduce. The set is LI if and only if every column is a pivot column. Equivalently, the set is LI if and only if $Ax = \mathbf{0}$ has only the trivial solution.

**Example.** Determine whether $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$ is LI or LD.

*Solution.* $\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \xrightarrow{R_2 - 2R_1,\ R_3 - 3R_1} \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -6 \end{bmatrix} \xrightarrow{R_3 - 2R_2} \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$

Column 3 has no pivot ⇒ the set is **LD**. The free variable gives $v_3 = v_1 - v_2$.

**Important Lemmas.**

1. A set containing the zero vector is always LD.
2. A set $\{v\}$ with one nonzero vector is LI.
3. If $\{v_1, \ldots, v_n\}$ is LD, then at least one vector is a linear combination of the others.
4. If $\{v_1, \ldots, v_n\}$ is LI and $\{v_1, \ldots, v_n, v_{n+1}\}$ is LD, then $v_{n+1} \in \text{span}\{v_1, \ldots, v_n\}$.
5. In $\mathbb{R}^n$, any set of more than $n$ vectors is LD.

**Wronskian.** Let $f_1, f_2, \ldots, f_n \in C^{n-1}(I)$. The Wronskian of $f_1, \ldots, f_n$ is

$$W(f_1, \ldots, f_n)(x) = \det \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{bmatrix}.$$

If $W(f_1, \ldots, f_n)(x_0) \neq 0$ for some $x_0 \in I$, then $\{f_1, \ldots, f_n\}$ is LI on $I$.

**Example.** Show $\{e^x, e^{2x}\}$ is LI.
*Solution.* $W(e^x, e^{2x}) = \det \begin{bmatrix} e^x & e^{2x} \\ e^x & 2e^{2x} \end{bmatrix} = 2e^{3x} - e^{3x} = e^{3x} \neq 0.$

## 9 Bases and Dimension

**Basis.** A set $B = \{v_1, v_2, \ldots, v_n\} \subseteq V$ is a basis for $V$ if:

1. $B$ spans $V$: span$(B) = V$.
2. $B$ is linearly independent.

**Standard Bases.**

- $\mathbb{R}^n$: $\{e_1, e_2, \ldots, e_n\}$, where $e_i$ has 1 in position $i$ and 0 elsewhere.
- $P_n(\mathbb{R})$: $\{1, x, x^2, \ldots, x^n\}$.
- $M_{2\times 2}(\mathbb{R})$: $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$.

**Dimension.** If $V$ has a basis $B = \{v_1, \ldots, v_n\}$, then $\dim(V) = n$. By convention, $\dim(\{\mathbf{0}\}) = 0$.

**Example.** $\dim(\mathbb{R}^n) = n$, $\quad \dim(P_n(\mathbb{R})) = n + 1$, $\quad \dim(M_{m\times n}(\mathbb{R})) = mn$.
¡br¿ $\dim(\text{Sym}_2(\mathbb{R})) = 3$ (the space of $2 \times 2$ symmetric matrices).
$\dim(C^n[a, b]) = \infty$.

**Theorem 13.** Let $V$ be a vector space with $\dim(V) = n$.

1. Any LI set in $V$ has at most $n$ vectors.
2. Any spanning set for $V$ has at least $n$ vectors.
3. Any LI set of exactly $n$ vectors is a basis.
4. Any spanning set of exactly $n$ vectors is a basis.

**Extending to a Basis.** If $S = \{v_1, \ldots, v_k\}$ is LI and $k < n = \dim(V)$, then $S$ can be extended to a basis by adding $n - k$ vectors.

**Reducing to a Basis.** If $S = \{v_1, \ldots, v_m\}$ spans $V$ and $m > n = \dim(V)$, then we can remove vectors from $S$ to obtain a basis.

**Finding a Basis for** $\text{null}(A)$**.** Row reduce $A$ to RREF. Express pivot variables in terms of free variables to get the general solution. The vectors corresponding to each free variable form a basis for $\text{null}(A)$.

**Example.** Find a basis for $\text{null}(A)$ where $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 2 & 4 & 1 & 0 \\ 3 & 6 & 1 & -1 \end{bmatrix}$.

*Solution.* RREF of $A$: $\begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.
Free variables: $x_2 = s$, $x_4 = t$. Then $x_1 = -2s + t$, $x_3 = -2t$.

$$\text{null}(A) = \text{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}, \quad \dim(\text{null}(A)) = 2.$$

**Theorem 14.** Let $A$ be an $n \times n$ matrix. The following are equivalent:

1. The columns of $A$ form a basis for $\mathbb{R}^n$.
2. $A$ is invertible.
3. $\det(A) \neq 0$.
4. $Ax = \mathbf{0}$ has only the trivial solution.
5. $Ax = b$ has a unique solution for every $b$.
6. $\text{rank}(A) = n$.
7. $\text{null}(A) = \{\mathbf{0}\}$.
8. The RREF of $A$ is $I_n$.

## 10 Direct Sums

**Sum of Subspaces.** If $W_1, W_2$ are subspaces of $V$, their sum is

$$W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1,\ w_2 \in W_2\}.$$

This is a subspace of $V$ and is the smallest subspace containing both $W_1$ and $W_2$.

**Direct Sum.** The sum $W_1 + W_2$ is a *direct sum*, written $W_1 \oplus W_2$, if $W_1 \cap W_2 = \{\mathbf{0}\}$. Equivalently, every vector $v \in W_1 + W_2$ can be written *uniquely* as $v = w_1 + w_2$ with $w_1 \in W_1$, $w_2 \in W_2$.
**Theorem 15.** $W_1 + W_2$ is a direct sum if and only if every vector in $W_1 + W_2$ has a unique representation as $w_1 + w_2$ with $w_i \in W_i$.

*Proof.* ($\Rightarrow$) Suppose $W_1 \cap W_2 = \{\mathbf{0}\}$ and $v = w_1 + w_2 = w_1' + w_2'$. Then $w_1 - w_1' = w_2' - w_2 \in W_1 \cap W_2 = \{\mathbf{0}\}$, so $w_1 = w_1'$ and $w_2 = w_2'$. ($\Leftarrow$) If $v \in W_1 \cap W_2$, then $v = v + \mathbf{0} = \mathbf{0} + v$ gives two representations of $v \in W_1 + W_2$. By uniqueness, $v = \mathbf{0}$. ☐

**Theorem 16 (Dimension Formula).** If $W_1, W_2$ are subspaces of a finite-dimensional vector space $V$, then

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

In particular, if $V = W_1 \oplus W_2$, then $\dim(V) = \dim(W_1) + \dim(W_2)$.

**Multiple Direct Sums.** More generally, $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$ if every $v \in V$ can be written uniquely as $v = w_1 + \cdots + w_k$ with $w_i \in W_i$. This holds iff $V = W_1 + \cdots + W_k$ and $W_i \cap (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_k) = \{\mathbf{0}\}$ for each $i$.

**Eigenspace Decomposition.** If $A$ is diagonalizable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then

$$\mathbb{R}^n = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k}.$$

This decomposition is the key structural result behind diagonalization: each vector decomposes uniquely into eigenvector components.

**Example.** In $\mathbb{R}^3$, let $W_1 = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ (the $x$-axis) and $W_2 = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ (the $yz$-plane). Then $W_1 \cap W_2 = \{\mathbf{0}\}$ and $W_1 + W_2 = \mathbb{R}^3$, so $\mathbb{R}^3 = W_1 \oplus W_2$.

## 11 Dual Spaces

**Linear Functional.** A linear functional (or linear form) on a vector space $V$ over $\mathbb{F}$ is a linear transformation $f : V \to \mathbb{F}$. That is, $f(au + bv) = af(u) + bf(v)$ for all $u, v \in V$ and $a, b \in \mathbb{F}$.

**Examples.**

- The trace $\text{tr} : \mathbb{R}^{n \times n} \to \mathbb{R}$ defined by $\text{tr}(A) = \sum_{i=1}^n a_{ii}$ is a linear functional on the space of $n \times n$ matrices.
- For a fixed $a \in \mathbb{R}^n$, the map $f(x) = a^T x$ is a linear functional on $\mathbb{R}^n$.
- The evaluation map $\text{ev}_c : P_n(\mathbb{R}) \to \mathbb{R}$ defined by $\text{ev}_c(p) = p(c)$ is a linear functional on the space of polynomials of degree $\leq n$.

**Dual Space.** The dual space of $V$, denoted $V^*$, is the vector space of all linear functionals on $V$:

$$V^* = \mathcal{L}(V, \mathbb{F}) = \{f : V \to \mathbb{F} \mid f \text{ is linear}\}.$$

$V^*$ is itself a vector space under pointwise addition and scalar multiplication: $(f + g)(v) = f(v) + g(v)$ and $(cf)(v) = c \cdot f(v)$.
**Theorem 17.** If $V$ is finite-dimensional, then $\dim(V^*) = \dim(V)$, so $V^* \cong V$.

**Dual Basis.** Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for $V$. The dual basis $\beta^* = \{f_1, \ldots, f_n\}$ for $V^*$ is defined by

$$f_i(v_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Each $f_i$ is the unique linear functional satisfying this condition. Then $\beta^*$ is a basis for $V^*$.

*Proof.* Each $f_i$ is well-defined since a linear map is determined by its action on a basis. To show $\beta^*$ is a basis: if $\sum c_i f_i = 0$, apply to $v_j$ to get $c_j = 0$, so $\beta^*$ is linearly independent. Since $|\beta^*| = n = \dim(V^*)$, it is a basis. ☐

**Example.** Let $V = \mathbb{R}^2$ with standard basis $\beta = \{e_1, e_2\}$. The dual basis is $\{f_1, f_2\}$ where $f_1(a, b) = a$ and $f_2(a, b) = b$ (the coordinate functions). For $\beta' = \{(1, 1), (1, -1)\}$, the dual basis is $f_1'(a, b) = \frac{a+b}{2}$ and $f_2'(a, b) = \frac{a-b}{2}$.

**Double Dual.** The double dual of $V$ is $V^{**} = (V^*)^*$. There is a natural (canonical) isomorphism $\Phi : V \to V^{**}$ defined by $\Phi(v)(f) = f(v)$ for all $f \in V^*$. This map is injective (hence an isomorphism when $V$ is finite-dimensional) and does not depend on any choice of basis, unlike the isomorphism $V \cong V^*$.

**Annihilator.** For a subspace $W \subseteq V$, the annihilator of $W$ is

$$W^0 = \{f \in V^* \mid f(w) = 0 \ \forall w \in W\}.$$

$W^0$ is a subspace of $V^*$.
**Theorem 18.** If $V$ is finite-dimensional and $W$ is a subspace of $V$, then $\dim(W^0) = \dim(V) - \dim(W)$.

*Proof.* The inclusion map $\iota : W \hookrightarrow V$ induces the dual map $\iota^* : V^* \to W^*$ defined by $\iota^*(f) = f|_W$. Then $\ker(\iota^*) = W^0$ and $\iota^*$ is surjective (extend any functional on $W$ to $V$). By the rank-nullity theorem, $\dim(W^0) = \dim(V^*) - \dim(W^*) = \dim(V) - \dim(W)$. ☐

**Transpose (Dual Map).** If $T : V \to W$ is linear, the transpose (or dual map) $T^* : W^* \to V^*$ is defined by $T^*(g) = g \circ T$ for $g \in W^*$. If $A$ is the matrix of $T$ with respect to bases $\beta, \gamma$, then the matrix of $T^*$ with respect to $\gamma^*, \beta^*$ is $A^T$.

## 12 Change of Basis

**Coordinate Vector.** Let $B = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for $V$. If $v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$, then the coordinate vector of $v$ relative to $B$ is

$$[v]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

**Example.** Let $B = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix} \right\}$ be a basis for $\mathbb{R}^2$. Find $[v]_B$ for $v = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$.

*Solution.* We need $c_1$ and $c_2$ such that $c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$. Solving gives

$c_1 = -2$, $c_2 = 3$. So $[v]_B = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

**Transition Matrix.** Let $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_n\}$ be two bases for $V$. The transition matrix from $B$ to $C$, written $P_{C \leftarrow B}$, satisfies

$$[v]_C = P_{C \leftarrow B}[v]_B \quad \text{for all } v \in V.$$

The columns of $P_{C \leftarrow B}$ are $[v_1]_C, [v_2]_C, \ldots, [v_n]_C$.

**Computing $P_{C \leftarrow B}$ in $\mathbb{R}^n$.** Row reduce $[C \mid B]$ to $[I \mid P_{C \leftarrow B}]$. Here $C = [w_1 \ \cdots \ w_n]$ and $B = [v_1 \ \cdots \ v_n]$.

**Properties.**

1. $P_{C \leftarrow B}$ is invertible, and $(P_{C \leftarrow B})^{-1} = P_{B \leftarrow C}$.
2. If $D$ is a third basis, then $P_{D \leftarrow B} = P_{D \leftarrow C} \cdot P_{C \leftarrow B}$.

**Example.** Let $B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ and $C = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ be bases for $\mathbb{R}^2$. Find $P_{C \leftarrow B}$.
*Solution.* Row reduce $[C \mid B]$:

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 2 & 1 & 1 & 1 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 0 & -1 & -1 & 3 \end{array}\right]$$

$$\xrightarrow{-R_2} \left[\begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & -3 \end{array}\right] \xrightarrow{R_1 - R_2} \left[\begin{array}{cc|cc} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & -3 \end{array}\right].$$

So $P_{C \leftarrow B} = \begin{bmatrix} 0 & 2 \\ 1 & -3 \end{bmatrix}$. Verify: $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. ✓

## 13 Row Space and Column Space

**Column Space.** The column space of $A$ is

$$\text{col}(A) = \text{span}\{\text{columns of } A\} = \{Ax \mid x \in \mathbb{R}^n\}.$$

**Row Space.** The row space of $A$ is

$$\text{row}(A) = \text{span}\{\text{rows of } A\} = \text{col}(A^T).$$

**Rank.** $\text{rank}(A) = \dim(\text{col}(A)) = \dim(\text{row}(A)) =$ number of pivot columns.

**Theorem 19 (Rank-Nullity Theorem for Matrices).** If $A$ is an $m \times n$ matrix, then

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

where $\text{nullity}(A) = \dim(\text{null}(A))$.

This is a direct consequence of the Rank-Nullity Theorem for linear transformations applied to $T(x) = Ax$. Since $\text{rank}(A) = \dim(\text{col}(A)) = \dim(\text{rng}\, T)$ and $\text{nullity}(A) = \dim(\ker T)$, the result follows immediately. Intuitively, the $n$ columns of $A$ contribute either to the rank (independent directions in the output) or to the nullity (directions collapsed to zero), and these two account for all $n$ dimensions of the domain.

**Finding a Basis for $\text{col}(A)$.** Row reduce $A$ to RREF. The columns of the **original** $A$ that correspond to pivot columns form a basis for $\text{col}(A)$.

**Finding a Basis for $\text{row}(A)$.** Row reduce $A$ to REF (or RREF). The nonzero rows of the REF form a basis for $\text{row}(A)$.
**Theorem 20.** Row equivalent matrices have the same row space.

**Example.** Find bases for $\text{col}(A)$ and $\text{row}(A)$ where

$$A = \begin{bmatrix} 1 & 3 & 5 & 4 \\ 3 & 1 & 6 & 8 \\ 4 & 4 & 11 & 2 \\ 7 & 5 & 17 & 20 \end{bmatrix}.$$

*Solution.* RREF of $A$: $\begin{bmatrix} 1 & 0 & 13/8 & 5/2 \\ 0 & 1 & 9/8 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Pivots in columns 1 and 2.

Basis for $\text{col}(A)$: $\left\{ \begin{bmatrix} 1 \\ 3 \\ 4 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 4 \\ 5 \end{bmatrix} \right\}$ (columns 1 and 2 of original $A$).
Basis for $\text{row}(A)$: $\left\{ \left(1, 0, \frac{13}{8}, \frac{5}{2}\right), \left(0, 1, \frac{9}{8}, \frac{1}{2}\right) \right\}$.
$\text{rank}(A) = 2$.

**LINEAR TRANSFORMATIONS**

## 14 Linear Transformations

**Linear Transformation.** Let $V$ and $W$ be vector spaces. A function $T : V \to W$ is a linear transformation if:

(a) $\forall v_1, v_2 \in V$, $T(v_1 + v_2) = T(v_1) + T(v_2)$.
(b) $\forall k \in \mathbb{F}$ and $v \in V$, $T(kv) = kT(v)$.

Equivalently, $T(c_1 v_1 + c_2 v_2) = c_1 T(v_1) + c_2 T(v_2)$ for all $v_1, v_2 \in V$ and $c_1, c_2 \in \mathbb{F}$.

**Examples.**

- $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(a, b) = (a + 3b, 5a + 7b)$ is a LT.
- The differential operator: $D : C^n(I) \to C^{n-1}(I)$, $D(f) = f'$, is a LT. More generally,
$$(D^n + a_{n-1}(x)D^{n-1} + \cdots + a_1(x)D + a_0(x))(f)$$
is a linear transformation.
- Integration: $T : C(\mathbb{R}) \to \mathbb{R}$ defined by $T(f) = \int_a^b f(t)\, dt$ is a LT.
- Reflection about the $x$-axis: $T(x, y) = (x, -y)$.
- Projection onto the $y$-axis: $T(x, y) = (0, y)$.
- The zero transformation: $T(v) = \mathbf{0}$ for all $v$.
- The identity transformation: $T(v) = v$ for all $v$.
- Matrix transformation: $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(x) = Ax$ for an $m \times n$ matrix $A$.

**Theorem 21.** Let $V$ and $W$ be vector spaces, and let $T : V \to W$ be a LT. Then,

1. $T(\mathbf{0}_V) = \mathbf{0}_W$.
2. $T(c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n)$.
3. $T(v_1 - v_2) = T(v_1) - T(v_2)$.

**Fact.** $T$ is determined by its values on basis vectors. If $B = \{v_1, \ldots, v_n\}$ is a basis for $V$ and $v = c_1 v_1 + \cdots + c_n v_n$, then

$$T(v) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n).$$

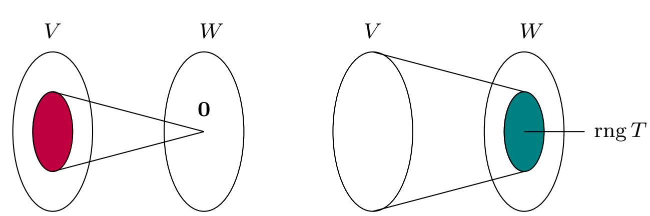
**Non-Examples.** A function $T$ is **not** a LT if:

- $T(\mathbf{0}) \neq \mathbf{0}$. For example, $T(x) = x + 1$ is not a LT since $T(0) = 1 \neq 0$.
- $T$ involves nonlinear operations. For example, $T(x, y) = (x^2, y)$ is not a LT.

## 15 Kernel and Range

**Kernel and Range.** Let $T : V \to W$ be a LT.

1. The kernel of $T$: $\ker T = \{v \in V \mid T(v) = \mathbf{0}\}$.
2. The range of $T$: $\text{rng}\, T = \{w \in W \mid w = T(v) \text{ for some } v \in V\}$.




(a) $\ker T$ (b) $\text{rng}\, T$

**Theorem 22.** Let $T : V \to W$ be a LT. Then $\ker T$ is a subspace of $V$, and $\text{rng}\, T$ is a subspace of $W$.

*Proof ($\ker T$ is a subspace).* (1) $T(\mathbf{0}_V) = \mathbf{0}_W$, so $\mathbf{0}_V \in \ker T$. (2) If $u, v \in \ker T$, then $T(u + v) = T(u) + T(v) = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $u + v \in \ker T$. (3) If $u \in \ker T$ and $c \in \mathbb{F}$, then $T(cu) = cT(u) = c\mathbf{0} = \mathbf{0}$, so $cu \in \ker T$. □

*Proof ($\text{rng}\, T$ is a subspace).* (1) $\mathbf{0}_W = T(\mathbf{0}_V) \in \text{rng}\, T$. (2) If $w_1, w_2 \in \text{rng}\, T$, then $w_1 = T(v_1)$ and $w_2 = T(v_2)$ for some $v_1, v_2 \in V$. So $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2) \in \text{rng}\, T$. (3) If $w = T(v) \in \text{rng}\, T$ and $c \in \mathbb{F}$, then $cw = cT(v) = T(cv) \in \text{rng}\, T$. □

**Connection to Null Space and Column Space.** Let $A$ be an $m \times n$ matrix and $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(v) = Av$. Then:

1. $\ker T = \text{null}(A)$.
2. $\text{rng}\, T = \text{col}(A)$.

**Theorem 23 (Rank-Nullity Theorem).** Let $T : V \to W$ be a LT with $\dim V = n$. Then
$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim V,$$
i.e., $\dim(\operatorname{rng} T) + \dim(\ker T) = n$.

*Proof.* Let $\dim(\ker T) = k$ and let $\{u_1, \dots, u_k\}$ be a basis for $\ker T$. Extend this to a basis $\{u_1, \dots, u_k, v_1, \dots, v_{n-k}\}$ of $V$. We claim $\{T(v_1), \dots, T(v_{n-k})\}$ is a basis for $\operatorname{rng} T$.

*Spanning:* Let $w \in \operatorname{rng} T$, so $w = T(v)$ for some $v \in V$. Write $v = \sum a_i u_i + \sum b_j v_j$. Then $w = T(v) = \sum a_i T(u_i) + \sum b_j T(v_j) = \sum b_j T(v_j)$ since $T(u_i) = \mathbf{0}$.

*LI:* Suppose $\sum c_j T(v_j) = \mathbf{0}$. Then $T(\sum c_j v_j) = \mathbf{0}$, so $\sum c_j v_j \in \ker T$. Thus $\sum c_j v_j = \sum d_i u_i$ for some scalars $d_i$. This gives $\sum d_i u_i - \sum c_j v_j = \mathbf{0}$, and since $\{u_1, \dots, u_k, v_1, \dots, v_{n-k}\}$ is a basis (hence LI), all $c_j = 0$.

So $\dim(\operatorname{rng} T) = n - k$, giving $\operatorname{rank}(T) + \operatorname{nullity}(T) = n$. □

**Example.** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x, y, z) = (x + y, y - z)$. The matrix is $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$. RREF: $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$. Free variable $x_3 = t$:
$$\ker T = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad \dim(\ker T) = 1.$$
Since $\operatorname{rank}(A) = 2$, $\operatorname{rng} T = \mathbb{R}^2$. Verify: $2 + 1 = 3 = \dim(\mathbb{R}^3)$. ✓

## 16 One-to-One and Onto

**One-to-one (Injective).** $T : V \to W$ is one-to-one if $T(v_1) = T(v_2) \implies v_1 = v_2$.

**Onto (Surjective).** $T : V \to W$ is onto if for every $w \in W$, there exists $v \in V$ with $T(v) = w$.

**Isomorphism.** $T$ is an isomorphism if it is both one-to-one and onto (bijective).

**Theorem 24.** Let $T : V \to W$ be a LT.

1. $T$ is one-to-one if and only if $\ker T = \{\mathbf{0}\}$.
2. $T$ is onto if and only if $\operatorname{rng} T = W$, i.e., $\dim(\operatorname{rng} T) = \dim W$.

*Proof of (1).* ($\Rightarrow$) Suppose $T$ is one-to-one. Let $v \in \ker T$, so $T(v) = \mathbf{0} = T(\mathbf{0})$. Since $T$ is one-to-one, $v = \mathbf{0}$. So $\ker T = \{\mathbf{0}\}$. ($\Leftarrow$) Suppose $\ker T = \{\mathbf{0}\}$. If $T(v_1) = T(v_2)$, then $T(v_1 - v_2) = T(v_1) - T(v_2) = \mathbf{0}$, so $v_1 - v_2 \in \ker T = \{\mathbf{0}\}$, giving $v_1 = v_2$. □

**Theorem 25.** Let $T : V \to W$ be a LT with $\dim V = \dim W = n$. Then:
$$T \text{ is one-to-one} \iff T \text{ is onto} \iff T \text{ is an isomorphism.}$$

**Isomorphic Vector Spaces.** $V$ and $W$ are isomorphic (written $V \cong W$) if there exists an isomorphism $T : V \to W$.

**Theorem 26.** Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic if and only if they have the same dimension.

**Example.** $P_2(\mathbb{R}) \cong \mathbb{R}^3$ since $\dim(P_2(\mathbb{R})) = 3 = \dim(\mathbb{R}^3)$. An explicit isomorphism is $T(a + bx + cx^2) = (a, b, c)$.

## 17 Matrix Representation of Linear Transformations

**Matrix of a Linear Transformation.** Let $T : V \to W$ be a LT, let $B = \{v_1, \dots, v_n\}$ be a basis for $V$, and let $C = \{w_1, \dots, w_m\}$ be a basis for $W$. The matrix of $T$ relative to $B$ and $C$ is
$$[T]_{C \leftarrow B} = \Big[ [T(v_1)]_C \ \ [T(v_2)]_C \ \ \cdots \ \ [T(v_n)]_C \Big].$$
This is the $m \times n$ matrix whose columns are the coordinate vectors of $T(v_1), \dots, T(v_n)$ relative to $C$.

The key property is: $[T(v)]_C = [T]_{C \leftarrow B}[v]_B$ for all $v \in V$.

When $V = W$ and $B = C$, we write $[T]_B = [T]_{B \leftarrow B}$.

**Example.** Let $T : P_2(\mathbb{R}) \to P_1(\mathbb{R})$ be the derivative $T(p) = p'$. Let $B = \{1, x, x^2\}$ and $C = \{1, x\}$. Then:
$$T(1) = 0 = 0 \cdot 1 + 0 \cdot x, \quad T(x) = 1 = 1 \cdot 1 + 0 \cdot x, \quad T(x^2) = 2x = 0 \cdot 1 + 2 \cdot x.$$
So $[T]_{C \leftarrow B} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

**Example.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be $T(x, y) = (x + y, 2x - y)$. With the standard basis $E = \{e_1, e_2\}$:
$$T(e_1) = T(1, 0) = (1, 2), \quad T(e_2) = T(0, 1) = (1, -1).$$
So $[T]_E = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$.

**Change of Basis for Linear Transformations.** If $B$ and $B'$ are two bases for $V$, and $T : V \to V$ is a LT, then
$$[T]_{B'} = P_{B' \leftarrow B}[T]_B P_{B \leftarrow B'} = (P_{B' \leftarrow B})[T]_B (P_{B' \leftarrow B})^{-1}.$$

**Similar Matrices.** Two $n \times n$ matrices $A$ and $B$ are similar if there exists an invertible matrix $P$ such that $B = P^{-1}AP$. Similar matrices represent the same linear transformation with respect to different bases.

**Theorem 27.** Similar matrices have the same:

1. Determinant.
2. Rank.
3. Trace.
4. Eigenvalues (with same multiplicities).
5. Characteristic polynomial.

# EIGENVALUES AND DIAGONALIZATION

## 18 Eigenvalues and Eigenvectors

**Definition.** Let $A$ be an $n \times n$ matrix. A scalar $\lambda$ is an *eigenvalue* of $A$ if there exists a nonzero vector $v$ such that
$$Av = \lambda v.$$
The nonzero vector $v$ is called an *eigenvector* of $A$ corresponding to $\lambda$.

**Eigenspace.** The eigenspace of $A$ corresponding to eigenvalue $\lambda$ is
$$E_\lambda = \operatorname{null}(A - \lambda I) = \{v \in \mathbb{R}^n \mid (A - \lambda I)v = \mathbf{0}\}.$$
This is a subspace of $\mathbb{R}^n$. The eigenvectors corresponding to $\lambda$ are the nonzero vectors in $E_\lambda$.

**Characteristic Polynomial.** The characteristic polynomial of $A$ is
$$p(\lambda) = \det(A - \lambda I).$$
The eigenvalues of $A$ are the roots of $p(\lambda) = 0$.

**Finding Eigenvalues and Eigenvectors.**

1. Compute $\det(A - \lambda I) = 0$ and solve for $\lambda$.
2. For each eigenvalue $\lambda_i$, solve $(A - \lambda_i I)v = \mathbf{0}$ to find the eigenspace $E_{\lambda_i}$.

**Example.** Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$.

*Solution.* $\det(A - \lambda I) = \det \begin{bmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{bmatrix} = (4 - \lambda)(3 - \lambda) - 2 = \lambda^2 - 7\lambda + 10 = (\lambda - 5)(\lambda - 2) = 0$.
Eigenvalues: $\lambda_1 = 5$, $\lambda_2 = 2$.

For $\lambda_1 = 5$: $(A - 5I)v = \mathbf{0} \implies \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} v = \mathbf{0} \implies v = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
$E_5 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

For $\lambda_2 = 2$: $(A - 2I)v = \mathbf{0} \implies \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} v = \mathbf{0} \implies v = t \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.
$E_2 = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$.

**Example.** Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & 0 & 1 \end{bmatrix}$.

*Solution.* $\det(A - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 2 & 0 \\ 0 & 3 - \lambda & 0 \\ 2 & 0 & 1 - \lambda \end{bmatrix} = (3 - \lambda) \det \begin{bmatrix} 1 - \lambda & 0 \\ 2 & 1 - \lambda \end{bmatrix} = (3 - \lambda)(1 - \lambda)^2 - (3 - \lambda) \cdot 4$. Expanding along row 2: $(3 - \lambda)[(1 - \lambda)^2 - 4] = (3 - \lambda)(\lambda^2 - 2\lambda - 3) = (3 - \lambda)(\lambda - 3)(\lambda + 1) = -(3 - \lambda)^2(\lambda + 1) = 0$.
Eigenvalues: $\lambda_1 = 3$ (a.m. $= 2$), $\lambda_2 = -1$ (a.m. $= 1$).

For $\lambda_1 = 3$: $(A-3I)v = \mathbf{0} \implies \begin{bmatrix} -2 & 2 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & -2 \end{bmatrix} v = \mathbf{0} \implies v = s\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. So $E_3 = \text{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$, g.m. = 1.

For $\lambda_2 = -1$: $(A+I)v = \mathbf{0} \implies \begin{bmatrix} 2 & 2 & 0 \\ 0 & 4 & 0 \\ 2 & 0 & 2 \end{bmatrix} v = \mathbf{0} \implies v = t\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$. $E_{-1} = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right\}$.

Since g.m.(3) $= 1 < 2 =$a.m.(3), this matrix is **not** diagonalizable.

**Algebraic and Geometric Multiplicity.**

- The *algebraic multiplicity* of $\lambda$, denoted a.m.($\lambda$), is the multiplicity of $\lambda$ as a root of the characteristic polynomial.
- The *geometric multiplicity* of $\lambda$, denoted g.m.($\lambda$), is $\dim(E_\lambda) = \text{nullity}(A - \lambda I)$.

**Theorem 28.** For each eigenvalue $\lambda$: $1 \le \text{g.m.}(\lambda) \le \text{a.m.}(\lambda)$.

**Properties of Eigenvalues.**

1. $\det(A) = \prod_{i=1}^n \lambda_i$ (product of eigenvalues).
2. $\text{tr}(A) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i$ (sum of eigenvalues).
3. $A$ is invertible if and only if 0 is not an eigenvalue of $A$.
4. If $\lambda$ is an eigenvalue of $A$, then $\lambda^k$ is an eigenvalue of $A^k$.
5. If $A$ is invertible and $\lambda$ is an eigenvalue, then $\lambda^{-1}$ is an eigenvalue of $A^{-1}$.
6. Eigenvalues of $A$ and $A^T$ are the same.

**Theorem 29.** Eigenvectors corresponding to distinct eigenvalues are linearly independent.

*Proof.* We prove this by induction. The base case ($k = 1$) is clear: a single eigenvector is nonzero, hence LI. Suppose the result holds for $k-1$ eigenvectors. Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues with eigenvectors $v_1, \ldots, v_k$. Suppose $c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}$. Multiplying by $A$:

$$c_1\lambda_1v_1 + c_2\lambda_2v_2 + \cdots + c_k\lambda_kv_k = \mathbf{0}.$$

Subtracting $\lambda_k$ times the original equation:

$$c_1(\lambda_1 - \lambda_k)v_1 + c_2(\lambda_2 - \lambda_k)v_2 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = \mathbf{0}.$$

By the inductive hypothesis, $\{v_1, \ldots, v_{k-1}\}$ is LI, so $c_i(\lambda_i - \lambda_k) = 0$ for $i = 1, \ldots, k-1$. Since $\lambda_i \ne \lambda_k$, we get $c_i = 0$ for all $i < k$. The original equation then gives $c_kv_k = \mathbf{0}$, and since $v_k \ne \mathbf{0}$, $c_k = 0$. □

**Trace.** The trace of an $n \times n$ matrix $A$ is $\text{tr}(A) = \sum_{i=1}^n a_{ii}$.
Properties: $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$, $\text{tr}(cA) = c\,\text{tr}(A)$, $\text{tr}(AB) = \text{tr}(BA)$.

**Cayley-Hamilton Theorem.** Every square matrix satisfies its own characteristic polynomial. That is, if $p(\lambda) = \det(A - \lambda I)$, then $p(A) = \mathbf{0}$ (the zero matrix).

*Proof sketch (diagonalizable case).* If $A = PDP^{-1}$ with $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then $p(A) = Pp(D)P^{-1}$. Since $p(D) = \text{diag}(p(\lambda_1), \ldots, p(\lambda_n))$ and each $\lambda_i$ is a root of $p$, we get $p(D) = \mathbf{0}$, hence $p(A) = \mathbf{0}$. The non-diagonalizable case requires a continuity argument or Jordan form. □

**Example.** If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $p(\lambda) = \lambda^2 - 5\lambda - 2$. By Cayley-Hamilton: $A^2 - 5A - 2I = \mathbf{0}$, so $A^2 = 5A + 2I$.

## 19 Diagonalization

**Diagonal Matrix.** A diagonal matrix is a square matrix with all off-diagonal entries equal to zero:

$$D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

**Diagonalizable Matrix.** An $n \times n$ matrix $A$ is diagonalizable if there exists an invertible matrix $P$ and a diagonal matrix $D$ such that $A = PDP^{-1}$, or equivalently $P^{-1}AP = D$.

The columns of $P$ are the eigenvectors of $A$, and the diagonal entries of $D$ are the corresponding eigenvalues.

**Theorem 30.** An $n \times n$ matrix $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors.

*Proof.* ($\Rightarrow$) If $A = PDP^{-1}$, then $AP = PD$. Writing $P = [v_1 \ \cdots \ v_n]$ and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, the $i$-th column gives $Av_i = \lambda_iv_i$. Since $P$ is invertible, $\{v_1, \ldots, v_n\}$ is LI. ($\Leftarrow$) If $v_1, \ldots, v_n$ are LI eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$, set $P = [v_1 \ \cdots \ v_n]$ (invertible since the columns are LI). Then $AP = PD$ where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, so $A = PDP^{-1}$. □

**Theorem 31.** $A$ is diagonalizable if and only if $\text{g.m.}(\lambda) = \text{a.m.}(\lambda)$ for every eigenvalue $\lambda$.

**Theorem 32.** If $A$ has $n$ distinct eigenvalues, then $A$ is diagonalizable (the converse is false).

**Diagonalization Procedure.**

1. Find the eigenvalues $\lambda_1, \ldots, \lambda_k$ of $A$.
2. For each $\lambda_i$, find a basis for $E_{\lambda_i}$.
3. If the total number of basis vectors equals $n$, form $P$ using these eigenvectors as columns and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$.
4. If not, $A$ is not diagonalizable.

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

*Solution.* From the earlier example, $\lambda_1 = 5$ with eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\lambda_2 = 2$ with eigenvector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$. So:

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}, \quad A = PDP^{-1}.$$

**Powers of Matrices.** If $A = PDP^{-1}$, then $A^k = PD^kP^{-1}$, where

$$D^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}.$$

**Example.** Compute $A^{100}$ for $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$.

*Solution.* $A^{100} = PD^{100}P^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 5^{100} & 0 \\ 0 & 2^{100} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}^{-1}$.

Since $P^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$:

$$A^{100} = \frac{1}{3}\begin{bmatrix} 2 \cdot 5^{100} + 2^{100} & 5^{100} - 2^{100} \\ 2 \cdot 5^{100} - 2 \cdot 2^{100} & 5^{100} + 2 \cdot 2^{100} \end{bmatrix}.$$

**Non-Diagonalizable Example.** $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has $\lambda = 2$ with a.m. $= 2$ but g.m. $= 1$ $\left(E_2 = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}\right)$. Since g.m. $<$ a.m., $A$ is not diagonalizable.

**Example.** Diagonalize $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 1 & 4 \end{bmatrix}$.

*Solution.* $p(\lambda) = \det(A - \lambda I) = (1-\lambda)[(1-\lambda)(4-\lambda)+2] = (1-\lambda)(\lambda^2 - 5\lambda + 6) = (1-\lambda)(\lambda-2)(\lambda-3)$.

Eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ (all distinct $\Rightarrow$ diagonalizable).

$E_1$: $(A - I)v = \mathbf{0} \implies v = t\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. $E_2$: $(A - 2I)v = \mathbf{0} \implies v = t\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

$E_3$: $(A - 3I)v = \mathbf{0} \implies v = t\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad A = PDP^{-1}.$$

## 20 Minimal Polynomial

**Definition.** The *minimal polynomial* $m_A(\lambda)$ of an $n \times n$ matrix $A$ is the monic polynomial of *least degree* such that $m_A(A) = \mathbf{0}$ (the zero matrix).

**Existence.** The minimal polynomial exists and is unique: the set of polynomials $p$ with $p(A) = \mathbf{0}$ is nonempty (by Cayley-Hamilton, $p_A(\lambda)$ is such a polynomial) and contains a unique monic polynomial of minimum degree.

**Theorem 33.** Let $p_A(\lambda) = \det(A - \lambda I)$ be the characteristic polynomial and $m_A(\lambda)$ the minimal polynomial of $A$.

1. $m_A(\lambda)$ divides $p_A(\lambda)$.
2. $m_A(\lambda)$ and $p_A(\lambda)$ have the same roots (i.e., the same eigenvalues).
3. $m_A(\lambda)$ divides every polynomial $q(\lambda)$ such that $q(A) = \mathbf{0}$.

*Proof of (2).* If $\lambda_0$ is a root of $m_A$, then $m_A(\lambda) = (\lambda - \lambda_0)q(\lambda)$, so $\mathbf{0} = m_A(A) = (A - \lambda_0 I)q(A)$. Since $\deg(q) < \deg(m_A)$, we have $q(A) \neq \mathbf{0}$, so there exists $v$ with $q(A)v \neq \mathbf{0}$. Then $(A - \lambda_0 I)(q(A)v) = \mathbf{0}$, so $q(A)v$ is an eigenvector for $\lambda_0$; hence $\lambda_0$ is an eigenvalue. Conversely, if $\lambda_0$ is an eigenvalue with eigenvector $v$, then $\mathbf{0} = m_A(A)v = m_A(\lambda_0)v$, so $m_A(\lambda_0) = 0$ since $v \neq \mathbf{0}$. □

**Theorem 34.** $A$ is diagonalizable if and only if $m_A(\lambda)$ has no repeated roots; i.e., $m_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_k)$ with distinct $\lambda_i$.

This provides a powerful diagonalizability test: instead of computing all geometric multiplicities, one can check whether the minimal polynomial is a product of distinct linear factors.

**Computing $m_A(\lambda)$.** Factor the characteristic polynomial $p_A(\lambda) = \prod(\lambda - \lambda_i)^{n_i}$. The minimal polynomial has the form $m_A(\lambda) = \prod(\lambda - \lambda_i)^{k_i}$ where $1 \le k_i \le n_i$. Test candidates starting with $k_i = 1$ (all exponents 1), then increase exponents until $m_A(A) = \mathbf{0}$.

**Examples.**

- $A = \operatorname{diag}(2,3,3)$: $p_A(\lambda) = (\lambda-2)(\lambda-3)^2$. Test $m = (\lambda-2)(\lambda-3)$: $m(A) = (A-2I)(A-3I) = \operatorname{diag}(0,1,1)\cdot\operatorname{diag}(-1,0,0) = \mathbf{0}$. So $m_A(\lambda) = (\lambda-2)(\lambda-3)$ and $A$ is diagonalizable.
- $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$: $p_A(\lambda) = (\lambda-2)^2$. Test $m = (\lambda-2)$: $m(A) = A - 2I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \neq \mathbf{0}$. So $m_A(\lambda) = (\lambda-2)^2$ (repeated root) and $A$ is **not** diagonalizable.

## 21 Jordan Normal Form

When a matrix is not diagonalizable, the Jordan normal form is the "closest to diagonal" form achievable by a similarity transformation.

**Jordan Block.** A $k \times k$ Jordan block for eigenvalue $\lambda$ is the matrix

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda & 1 \\ 0 & \cdots & 0 & 0 & \lambda \end{bmatrix}_{k\times k}.$$

A $1 \times 1$ Jordan block $J_1(\lambda) = [\lambda]$ is just a diagonal entry.

**Generalized Eigenspace.** The *generalized eigenspace* of $A$ for eigenvalue $\lambda$ is

$$G_\lambda = \operatorname{null}((A - \lambda I)^{n_\lambda}) = \{v \in \mathbb{R}^n \mid (A - \lambda I)^{n_\lambda} v = \mathbf{0}\},$$

where $n_\lambda = \text{a.m.}(\lambda)$. We always have $E_\lambda \subseteq G_\lambda$ and $\dim(G_\lambda) = n_\lambda$.

**Theorem 35 (Jordan Normal Form).** Let $A$ be an $n \times n$ matrix (over $\mathbb{C}$, or over $\mathbb{R}$ if all eigenvalues are real). Then there exists an invertible matrix $P$ such that

$$P^{-1}AP = J = \begin{bmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_r}(\lambda_r) \end{bmatrix},$$

where $J$ is the *Jordan normal form* of $A$. The Jordan blocks are unique up to reordering.

**Properties.**

- The number of Jordan blocks for $\lambda$ equals $\text{g.m.}(\lambda) = \dim(E_\lambda)$.
- The sum of the sizes of all Jordan blocks for $\lambda$ equals $\text{a.m.}(\lambda)$.
- $A$ is diagonalizable $\iff$ every Jordan block is $1 \times 1$ $\iff$ $m_A$ has no repeated roots.
- The largest Jordan block for $\lambda$ has size equal to the exponent of $(\lambda - \lambda_i)$ in $m_A(\lambda)$.

**Computing Jordan Form.** For each eigenvalue $\lambda$ with $\text{a.m.}(\lambda) = n_\lambda$:

1. Compute $r_k = \operatorname{rank}((A - \lambda I)^k)$ for $k = 1, 2, \ldots$ until it stabilizes.
2. The number of blocks of size $\ge k$ is $n_\lambda - r_k + r_{k-1}$ (where $r_0 = n$). Equivalently, the number of Jordan blocks of size exactly $k$ is
$$\operatorname{nullity}((A - \lambda I)^k) - \operatorname{nullity}((A - \lambda I)^{k-1}).$$

**Example.** Let $A = \begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix}$ with $p_A(\lambda) = (\lambda-1)(\lambda-2)(\lambda-4)^2$.

For $\lambda = 4$: a.m. $= 2$. Compute $\operatorname{rank}(A - 4I) = 3$, so nullity $= 1 =$ g.m. Since g.m. $<$ a.m., there is one $2 \times 2$ Jordan block for $\lambda = 4$. For $\lambda = 1$ and $\lambda = 2$: a.m. $= 1$, so each yields a $1 \times 1$ block.

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

**Powers via Jordan Form.** If $A = PJP^{-1}$, then $A^k = PJ^kP^{-1}$. For a single Jordan block:

$$J_m(\lambda)^k = \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \lambda^k \end{bmatrix},$$

where the $(i,j)$ entry is $\binom{k}{j-i}\lambda^{k-(j-i)}$ for $j \ge i$ and 0 for $j < i$.

**Matrix Exponential.** For a Jordan block, $e^{J_m(\lambda)t} = e^{\lambda t}\begin{bmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{m-1}}{(m-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{m-2}}{(m-2)!} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & t \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$. If $A = PJP^{-1}$, then $e^{At} = Pe^{Jt}P^{-1}$.

**Example.** Find the Jordan form and a Jordan basis for $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

*Solution.* $p(\lambda) = (\lambda-2)^2(\lambda-3)$. Eigenvalues: $\lambda = 2$ (a.m. $= 2$) and $\lambda = 3$ (a.m. $= 1$).

For $\lambda = 2$: $\operatorname{nullity}(A - 2I) = \operatorname{nullity}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 1$. Since g.m. $= 1 < 2 =$ a.m., one $2 \times 2$ Jordan block.

Eigenvector: $v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Generalized eigenvector: solve $(A - 2I)v_2 = v_1$, i.e.,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \implies v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

For $\lambda = 3$: $v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I, \quad J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = A.$$

$A$ is already in Jordan form. For a non-trivial case, if $B = \begin{bmatrix} 3 & 1 \\ -1 & 5 \end{bmatrix}$, then $p(\lambda) = (\lambda-4)^2$ and $B - 4I = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$ with nullity 1. Eigenvector: $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Generalized eigenvector: $(B - 4I)v_2 = v_1 \implies v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. So $P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $P^{-1}BP = \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$.

**INNER PRODUCT SPACES**

## 22 Inner Product Spaces

**Inner Product.** Let $V$ be a vector space over $\mathbb{R}$. An inner product on $V$ is a function $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ satisfying:

1. **Positive Definiteness:** $\forall v \in V$, $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0$ iff $v = \mathbf{0}$.
2. **Symmetry:** $\forall v, w \in V$, $\langle v, w\rangle = \langle w, v\rangle$.
3. **Linearity (first argument):** $\forall u, v, w \in V$ and $c \in \mathbb{R}$:
   - $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.
   - $\langle cu, w\rangle = c\langle u, w\rangle$.

An *inner product space* is a vector space equipped with an inner product.

**Examples.**

- The dot product on $\mathbb{R}^n$: $\langle u, v\rangle = u \cdot v = \sum_{i=1}^n u_i v_i$ is the standard inner product on $\mathbb{R}^n$.
- On $V = C[a,b]$: $\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$ is the standard inner product on $C[a,b]$.
- On $V = C[0,1]$: $\langle f, g\rangle = \int_0^{1/2} f(x)g(x)\,dx$ is **not** an inner product. Consider $f(x) = \begin{cases} 0 & x \leq \frac{1}{2} \\ (x-\frac{1}{2})^2 & x > \frac{1}{2} \end{cases}$. Then $f \neq 0$ but $\langle f, f\rangle = \int_0^{1/2} f^2(x)\,dx = 0$, violating positive definiteness.

**Norm.** The norm of $v \in V$ is $\|v\| = \sqrt{\langle v, v\rangle}$.

**Distance.** $d(u,v) = \|u - v\|$.

**Angle.** The angle $\theta$ between $v$ and $w$ is $\cos\theta = \frac{\langle v,w\rangle}{\|v\|\,\|w\|}$.

**Orthogonality.** $u$ and $v$ are orthogonal if $\langle u, v\rangle = 0$, written $u \perp v$.

**Theorem 36 (Cauchy-Schwarz Inequality).** For all $u, v \in V$: $|\langle u, v\rangle| \leq \|u\| \cdot \|v\|$.

*Proof.* If $v = \mathbf{0}$, both sides are 0. Assume $v \neq \mathbf{0}$. For any $t \in \mathbb{R}$, consider $\langle u - tv, u - tv\rangle \geq 0$ (positive definiteness). Expanding:

$$\langle u, u\rangle - 2t\langle u, v\rangle + t^2\langle v, v\rangle \geq 0.$$

This is a quadratic in $t$ that is always $\geq 0$, so its discriminant must be $\leq 0$:

$$4\langle u, v\rangle^2 - 4\langle u, u\rangle\langle v, v\rangle \leq 0 \implies \langle u, v\rangle^2 \leq \|u\|^2\|v\|^2.$$

Taking square roots gives the result. Equality holds iff $u = tv$ for some $t$. □

**Theorem 37 (Triangle Inequality).** For all $u, v \in V$: $\|u+v\| \leq \|u\| + \|v\|$.

*Proof.* $\|u+v\|^2 = \langle u+v, u+v\rangle = \|u\|^2 + 2\langle u, v\rangle + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$, where the inequality uses Cauchy-Schwarz. □

**Theorem 38 (Pythagorean Theorem).** If $u \perp v$, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

*Proof.* $\|u+v\|^2 = \langle u+v, u+v\rangle = \|u\|^2 + 2\langle u, v\rangle + \|v\|^2 = \|u\|^2 + \|v\|^2$ since $\langle u, v\rangle = 0$. □

**Theorem 39 (Extended Linearity).** Let $V$ be an inner product space. For all $c_1, \ldots, c_n \in \mathbb{R}$, $v_1, \ldots, v_n, w \in V$:

$$\langle c_1v_1 + \cdots + c_nv_n,\ w\rangle = c_1\langle v_1, w\rangle + \cdots + c_n\langle v_n, w\rangle.$$

Similarly for the second argument (by symmetry).

## 23 Orthogonal Sets and Orthogonal Projection

**Orthogonal and Orthonormal Sets.** Let $V$ be an inner product space and $S = \{v_1, v_2, \ldots, v_n\} \subseteq V$.

1. $S$ is an *orthogonal set* if $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$.
2. $S$ is an *orthonormal set* if $S$ is orthogonal and $\|v_i\| = 1$ for all $i$.
3. An orthogonal basis / orthonormal basis is a basis that is orthogonal / orthonormal.

To make an orthogonal set orthonormal, normalize each vector:

$$\hat{v}_i = \frac{v_i}{\|v_i\|}.$$

**Theorem 40.** Any orthogonal set not containing $\mathbf{0}$ is LI.

*Proof.* Let $S = \{v_1, \ldots, v_n\}$ be orthogonal with no zero vectors. Suppose $c_1v_1 + \cdots + c_nv_n = \mathbf{0}$. Taking the inner product with $v_j$:

$$0 = \langle \mathbf{0}, v_j\rangle = \sum_{i=1}^n c_i\langle v_i, v_j\rangle = c_j\langle v_j, v_j\rangle = c_j\|v_j\|^2.$$

Since $v_j \neq \mathbf{0}$, $\|v_j\|^2 > 0$, so $c_j = 0$ for all $j$. □

**Orthogonal Projection.** The projection of $b$ onto $a$ is

$$\text{proj}_a(b) = \frac{\langle a, b\rangle}{\|a\|^2}a = \frac{\langle a, b\rangle}{\langle a, a\rangle}a.$$

The scalar $\frac{\langle a,b\rangle}{\|a\|}$ is called the *scalar projection* (or component) of $b$ onto $a$.

**Theorem 41 (Fourier Representation Theorem).** Let $V$ be an inner product space.

1. If $S = \{v_1, \ldots, v_n\}$ is an orthogonal basis, then for all $v \in V$:
$$v = \sum_{i=1}^n \frac{\langle v, v_i\rangle}{\|v_i\|^2}v_i = \frac{\langle v, v_1\rangle}{\langle v_1, v_1\rangle}v_1 + \cdots + \frac{\langle v, v_n\rangle}{\langle v_n, v_n\rangle}v_n.$$
2. If $S = \{v_1, \ldots, v_n\}$ is an orthonormal basis, then for all $v \in V$:
$$v = \sum_{i=1}^n \langle v, v_i\rangle v_i = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_n\rangle v_n.$$

## 24 The Gram-Schmidt Process

**Gram-Schmidt Process.** Let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis for an inner product space $V$. Define:

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \frac{\langle v_2, w_1\rangle}{\|w_1\|^2}w_1 \\ w_3 &= v_3 - \frac{\langle v_3, w_1\rangle}{\|w_1\|^2}w_1 - \frac{\langle v_3, w_2\rangle}{\|w_2\|^2}w_2 \\ &\vdots \\ w_k &= v_k - \sum_{i=1}^{k-1}\frac{\langle v_k, w_i\rangle}{\|w_i\|^2}w_i \end{aligned}$$

Then $\{w_1, w_2, \ldots, w_n\}$ is an orthogonal basis for $V$. To obtain an orthonormal basis, normalize:

$$\left\{\frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}, \ldots, \frac{w_n}{\|w_n\|}\right\}.$$

**Example.** Apply Gram-Schmidt to $B = \left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}\right\}$ in $\mathbb{R}^3$ (with the dot product).

*Solution.* $w_1 = \begin{bmatrix}1\\1\\0\end{bmatrix}$.

$$w_2 = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{\langle\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}\rangle}{\|\begin{bmatrix}1\\1\\0\end{bmatrix}\|^2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1/2\\-1/2\\1\end{bmatrix}.$$

$$w_3 = \begin{bmatrix}0\\1\\1\end{bmatrix} - \frac{\langle\begin{bmatrix}0\\1\\1\end{bmatrix}, w_1\rangle}{\|w_1\|^2}w_1 - \frac{\langle\begin{bmatrix}0\\1\\1\end{bmatrix}, w_2\rangle}{\|w_2\|^2}w_2 = \begin{bmatrix}0\\1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} - \frac{1/2}{3/2}\begin{bmatrix}1/2\\-1/2\\1\end{bmatrix} = \begin{bmatrix}-2/3\\2/3\\2/3\end{bmatrix}.$$

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

## 25 QR Factorization

**QR Factorization.** If $A$ is an $m \times n$ matrix with linearly independent columns, then $A$ can be factored as $A = QR$, where:

- $Q$ is an $m \times n$ matrix with orthonormal columns.
- $R$ is an $n \times n$ upper triangular invertible matrix with positive diagonal entries.

**Method.** Apply Gram-Schmidt to the columns of $A$, then normalize to get the columns of $Q$. Then $R = Q^T A$.

**Example.** Find the QR factorization of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

*Solution.* Gram-Schmidt on the columns $a_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $a_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$:

$w_1 = a_1$, $w_2 = a_2 - \frac{a_2 \cdot w_1}{\|w_1\|^2} w_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$.

Normalize: $q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $q_2 = \frac{1}{\sqrt{6/4}} \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 2/\sqrt{6} \end{bmatrix}$, $R = Q^T A = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & 3/\sqrt{6} \end{bmatrix}$.

## 26 Orthogonal Complements and Projections

**Orthogonal Complement.** Let $W$ be a subspace of an inner product space $V$. The orthogonal complement of $W$ is

$$W^\perp = \{v \in V \mid \langle v, w \rangle = 0 \text{ for all } w \in W\}.$$

**Theorem 42.** Let $W$ be a subspace of a finite-dimensional inner product space $V$.

1. $W^\perp$ is a subspace of $V$.
2. $\dim(W) + \dim(W^\perp) = \dim(V)$.
3. $(W^\perp)^\perp = W$.
4. $W \cap W^\perp = \{\mathbf{0}\}$.

**Fundamental Subspaces.** For an $m \times n$ matrix $A$:

1. $\text{col}(A)^\perp = \text{null}(A^T)$ (left null space).
2. $\text{row}(A)^\perp = \text{null}(A)$.
3. $\mathbb{R}^n = \text{row}(A) \oplus \text{null}(A)$.
4. $\mathbb{R}^m = \text{col}(A) \oplus \text{null}(A^T)$.

**Theorem 43 (Orthogonal Decomposition Theorem).** Let $W$ be a subspace of $V$ and $v \in V$. Then there exist unique vectors $w \in W$ and $w^\perp \in W^\perp$ such that

$$v = w + w^\perp.$$

The vector $w = \text{proj}_W(v)$ is the *orthogonal projection* of $v$ onto $W$.

*Proof (existence).* Let $\{u_1, \ldots, u_k\}$ be an orthogonal basis for $W$. Set $w = \sum_{i=1}^k \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} u_i \in W$ and $w^\perp = v - w$. For each basis vector $u_j$: $\langle w^\perp, u_j \rangle = \langle v, u_j \rangle - \sum_i \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} \langle u_i, u_j \rangle = \langle v, u_j \rangle - \langle v, u_j \rangle = 0$. So $w^\perp \perp W$.

*Proof (uniqueness).* Suppose $v = w_1 + w_1^\perp = w_2 + w_2^\perp$. Then $w_1 - w_2 = w_2^\perp - w_1^\perp$. The left side is in $W$ and the right side is in $W^\perp$. Since $W \cap W^\perp = \{\mathbf{0}\}$, both sides equal $\mathbf{0}$, so $w_1 = w_2$ and $w_1^\perp = w_2^\perp$. □

**Orthogonal Projection onto a Subspace.** If $\{u_1, \ldots, u_k\}$ is an orthogonal basis for $W$, then

$$\text{proj}_W(v) = \frac{\langle v, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 + \frac{\langle v, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2 + \cdots + \frac{\langle v, u_k \rangle}{\langle u_k, u_k \rangle} u_k.$$

The component of $v$ orthogonal to $W$ is $v - \text{proj}_W(v) \in W^\perp$.

**Example.** Let $W = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ (already orthogonal) and $v = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$.

$\text{proj}_W(v) = \frac{6}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{3}{1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$.

$v - \text{proj}_W(v) = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \in W^\perp$.

## 27 Least Squares Approximation

**Motivation.** If $Ax = b$ has no solution (the system is inconsistent), we seek $\hat{x}$ that minimizes $\|Ax - b\|$, i.e., the vector $\hat{x}$ such that $A\hat{x}$ is as close to $b$ as possible.

**Normal Equations.** The least squares solution $\hat{x}$ satisfies

$$A^T A \hat{x} = A^T b.$$

If $A$ has linearly independent columns, then $A^T A$ is invertible and

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

**Geometric Interpretation.** $A\hat{x} = \text{proj}_{\text{col}(A)}(b)$. The least squares solution projects $b$ onto the column space of $A$. The error vector $b - A\hat{x} \in \text{col}(A)^\perp = \text{null}(A^T)$.

**Theorem 44.** Let $A$ be an $m \times n$ matrix. The equation $A^T A \hat{x} = A^T b$ is always consistent, and its solution set gives all least squares solutions of $Ax = b$.

**Least Squares and QR.** If $A = QR$, then $\hat{x} = R^{-1} Q^T b$.

**Example.** Find the least squares solution to $Ax = b$ where $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$.

*Solution.* $A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$, $A^T b = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$.

$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \hat{x} = \begin{bmatrix} 5 \\ 12 \end{bmatrix} \implies \hat{x} = \begin{bmatrix} 1/3 \\ 1 \end{bmatrix}$.

The best-fit line is $y = \frac{1}{3} + x$.

**Least Squares Regression.** To fit $y = \beta_0 + \beta_1 x$ to data points $(x_1, y_1), \ldots, (x_m, y_m)$, let

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Solve $A^T A \hat{x} = A^T b$ for $\hat{x} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$.

## 28 Orthogonal Matrices

**Orthogonal Matrix.** An $n \times n$ matrix $Q$ is orthogonal if $Q^T Q = Q Q^T = I$, i.e., $Q^{-1} = Q^T$.

Equivalently, $Q$ is orthogonal if and only if its columns form an orthonormal basis for $\mathbb{R}^n$.

**Theorem 45.** Let $Q$ be an $n \times n$ orthogonal matrix. Then:

1. $\det(Q) = \pm 1$.
2. $\|Qx\| = \|x\|$ for all $x \in \mathbb{R}^n$ (preserves lengths).
3. $\langle Qx, Qy \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{R}^n$ (preserves inner products).
4. The rows of $Q$ also form an orthonormal set.
5. If $Q_1, Q_2$ are orthogonal, then $Q_1 Q_2$ is orthogonal.

*Proof of (1).* $1 = \det(I) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2$, so $\det(Q) = \pm 1$. □

*Proof of (2).* $\|Qx\|^2 = (Qx)^T (Qx) = x^T Q^T Q x = x^T I x = \|x\|^2$. □

*Proof of (3).* $\langle Qx, Qy \rangle = (Qx)^T (Qy) = x^T Q^T Q y = x^T y = \langle x, y \rangle$. □

*Proof of (5).* $(Q_1 Q_2)^T (Q_1 Q_2) = Q_2^T Q_1^T Q_1 Q_2 = Q_2^T Q_2 = I$. □

**Examples.** Rotation by angle $\theta$: $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Reflection: $Q = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$.

Permutation matrices are orthogonal. The identity matrix $I$ is orthogonal.

**SPECTRAL THEORY AND MATRIX DECOMPOSITIONS**

## 29 Symmetric Matrices and the Spectral Theorem

**Properties of Symmetric Matrices.** Let $A$ be a real symmetric matrix ($A = A^T$).

1. All eigenvalues of $A$ are real.
2. Eigenvectors corresponding to distinct eigenvalues are orthogonal.
3. $A$ is always diagonalizable.

*Proof of (1).* Suppose $A\bar{v} = \lambda\bar{v}$ where $\bar{v} \neq \mathbf{0}$ (working over $\mathbb{C}$). Then $\bar{v}^*A\bar{v} = \lambda\bar{v}^*\bar{v}$, where $\bar{v}^*$ is the conjugate transpose. Since $A = A^T = \overline{A}$ (real entries), $(\bar{v}^*A\bar{v})^* = \bar{v}^*A^*\bar{v} = \bar{v}^*A\bar{v}$, so $\bar{v}^*A\bar{v}$ is real. Since $\bar{v}^*\bar{v} = \|\bar{v}\|^2 > 0$ is real, $\lambda$ must be real. □

*Proof of (2).* Let $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$ with $\lambda_1 \neq \lambda_2$. Then:

$$\lambda_1\langle v_1, v_2\rangle = \langle Av_1, v_2\rangle = v_1^T A^T v_2 = v_1^T A v_2 = \langle v_1, Av_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

So $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0$. Since $\lambda_1 \neq \lambda_2$, $\langle v_1, v_2\rangle = 0$. □

**Orthogonal Diagonalization.** $A$ is orthogonally diagonalizable if there exists an orthogonal matrix $Q$ such that $Q^TAQ = D$ (diagonal), i.e., $A = QDQ^T$.

**Theorem 46 (Spectral Theorem).** An $n \times n$ real matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Procedure for Orthogonal Diagonalization.**

1. Find the eigenvalues and eigenspaces of $A$.
2. For each eigenspace, find an orthonormal basis (use Gram-Schmidt if $\dim(E_\lambda) > 1$).
3. Form $Q$ with these orthonormal eigenvectors as columns, and $D$ with eigenvalues on the diagonal.

**Spectral Decomposition.** If $A = QDQ^T$ with $Q = [q_1 \ \cdots \ q_n]$ and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then

$$A = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T.$$

Each $q_i q_i^T$ is a rank-1 projection matrix.

**Example.** Orthogonally diagonalize $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

*Solution.* $\det(A - \lambda I) = (\lambda - 3)(\lambda - 1) = 0$. Eigenvalues: $\lambda_1 = 3$, $\lambda_2 = 1$.

$E_3$: $(A - 3I)v = \mathbf{0} \implies v = t\begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies q_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$E_1$: $(A - I)v = \mathbf{0} \implies v = t\begin{bmatrix} -1 \\ 1 \end{bmatrix} \implies q_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$, $A = QDQ^T$.

## 30 Quadratic Forms

**Definition.** A quadratic form on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ of the form

$$Q(x) = x^TAx,$$

where $A$ is an $n \times n$ symmetric matrix. $A$ is called the matrix of the quadratic form.

**Example.** $Q(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 3x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Note: the off-diagonal entries are half the coefficient of the cross term $x_ix_j$.

**Classification.** A quadratic form $Q(x) = x^TAx$ is:

- **Positive definite** if $Q(x) > 0$ for all $x \neq \mathbf{0}$ $\iff$ all eigenvalues of $A$ are positive.
- **Positive semidefinite** if $Q(x) \geq 0$ for all $x$ $\iff$ all eigenvalues $\geq 0$.
- **Negative definite** if $Q(x) < 0$ for all $x \neq \mathbf{0}$ $\iff$ all eigenvalues are negative.
- **Negative semidefinite** if $Q(x) \leq 0$ for all $x$ $\iff$ all eigenvalues $\leq 0$.
- **Indefinite** if $Q$ takes both positive and negative values $\iff$ $A$ has both positive and negative eigenvalues.

**Change of Variable.** If $A = PDP^T$ (orthogonal diagonalization), the substitution $x = Py$ transforms $Q(x) = x^TAx$ into

$$Q = y^TDy = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

which has no cross terms. This is called the *principal axis theorem.*

**Example.** Classify $Q(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 2x_2^2$. The matrix is $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Eigenvalues: $\lambda_1 = 3 > 0$, $\lambda_2 = 1 > 0$. Since all eigenvalues are positive, $Q$ is positive definite.

## 31 Positive Definite Matrices

**Definition.** A symmetric $n \times n$ matrix $A$ is *positive definite* (PD) if $x^TAx > 0$ for all $x \neq \mathbf{0}$. It is *positive semidefinite* (PSD) if $x^TAx \geq 0$ for all $x$.

**Theorem 47 (Equivalent Characterizations).** For a symmetric matrix $A$, the following are equivalent:

1. $A$ is positive definite.
2. All eigenvalues of $A$ are positive.
3. All leading principal minors of $A$ are positive (Sylvester's criterion).
4. There exists an invertible matrix $R$ such that $A = R^TR$.
5. $A$ has a Cholesky factorization $A = LL^T$.

*Proof of (1)⇔(2).* ($\Rightarrow$) If $\lambda$ is an eigenvalue with eigenvector $v$, then $0 < v^TAv = v^T(\lambda v) = \lambda\|v\|^2$, so $\lambda > 0$. ($\Leftarrow$) Since $A$ is symmetric, write $A = PDP^T$ with $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_i > 0$. For $x \neq \mathbf{0}$, let $y = P^Tx \neq \mathbf{0}$. Then $x^TAx = y^TDy = \sum \lambda_i y_i^2 > 0$. □

**Sylvester's Criterion.** The $k$-th *leading principal minor* $\Delta_k$ is the determinant of the top-left $k \times k$ submatrix of $A$. A symmetric matrix $A$ is PD iff $\Delta_k > 0$ for all $k = 1, \ldots, n$. It is ND iff $(-1)^k\Delta_k > 0$.

**Example.** Is $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ positive definite?

$\Delta_1 = 2 > 0$, $\Delta_2 = \det\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 3 > 0$, $\Delta_3 = \det(A) = 4 > 0$. Yes, $A$ is PD.

**Cholesky Factorization.** Every PD matrix $A$ has a unique factorization $A = LL^T$ where $L$ is lower triangular with *positive* diagonal entries.

**Theorem 48.** If $A$ is positive definite, then the Cholesky factorization $A = LL^T$ exists and is unique.

*Proof sketch.* Proceed by induction on $n$. Write $A = \begin{bmatrix} a_{11} & w^T \\ w & A' \end{bmatrix}$ with $a_{11} > 0$ (since $e_1^TAe_1 > 0$). Set $l_{11} = \sqrt{a_{11}}$ and $l = w/l_{11}$. Then $A = \begin{bmatrix} l_{11} & 0 \\ l & I \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & A' - ll^T \end{bmatrix}\begin{bmatrix} l_{11} & l^T \\ 0 & I \end{bmatrix}$. The Schur complement $A' - ll^T$ is again PD (by the PD characterization for submatrices), so by induction it has a Cholesky factor $L'$. Combining yields $L = \begin{bmatrix} l_{11} & 0 \\ l & L' \end{bmatrix}$. □

**Computing Cholesky.** The entries of $L$ are computed as:

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}, \qquad l_{ij} = \frac{1}{l_{jj}}\left(a_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk}\right) \quad \text{for } i > j.$$

**Example.** Find the Cholesky factorization of $A = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}$.

*Solution.* $l_{11} = \sqrt{4} = 2$, $l_{21} = \frac{2}{2} = 1$, $l_{22} = \sqrt{5 - 1} = 2$. So $L = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$, $A = LL^T$.

**Applications.**

- **Optimization:** A critical point $x^*$ of $f(x)$ is a local minimum if the Hessian $H_f(x^*)$ is PD.
- **Covariance matrices:** Every covariance matrix is PSD; it is PD when no variable is a linear combination of the others.
- **Efficient solving:** $Ax = b$ with PD $A$ is solved efficiently via Cholesky: $LL^Tx = b$ requires roughly half the operations of general LU.

## 32 Singular Value Decomposition

**Singular Values.** Let $A$ be an $m \times n$ matrix. The singular values of $A$ are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, where $\sigma_i = \sqrt{\lambda_i}$ and $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues of $A^T A$.

**Theorem 49** (**SVD**)**.** Let $A$ be an $m \times n$ matrix with $\text{rank}(A) = r$. Then there exist:

- An $m \times m$ orthogonal matrix $U$ (left singular vectors),
- An $n \times n$ orthogonal matrix $V$ (right singular vectors),
- An $m \times n$ "diagonal" matrix $\Sigma$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$ on the diagonal,

such that $A = U\Sigma V^T$.

**Computing the SVD.**

1. Compute $A^T A$ and find its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Set $\sigma_i = \sqrt{\lambda_i}$.
2. Find orthonormal eigenvectors $v_1, \ldots, v_n$ of $A^T A$. These form columns of $V$.
3. For $i = 1, \ldots, r$: $u_i = \frac{1}{\sigma_i} A v_i$. Extend $\{u_1, \ldots, u_r\}$ to an orthonormal basis of $\mathbb{R}^m$ to get $U$.
4. $\Sigma$ is $m \times n$ with $\Sigma_{ii} = \sigma_i$ for $i = 1, \ldots, r$ and 0 elsewhere.

**Properties.**

- $\text{rank}(A) =$ number of nonzero singular values.
- $\|A\| = \sigma_1$ (the largest singular value, in the operator 2-norm).
- The columns of $V$ corresponding to zero singular values form an orthonormal basis for $\text{null}(A)$.
- The columns of $U$ corresponding to nonzero singular values form an orthonormal basis for $\text{col}(A)$.

**Example.** Find the SVD of $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.

*Solution.* $A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Eigenvalues: $\lambda_1 = 2$, $\lambda_2 = 0$. So $\sigma_1 = \sqrt{2}$.

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

$u_1 = \frac{1}{\sqrt{2}} A v_1 = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. $u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$, $\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}$, $A = U\Sigma V^T$.

**Pseudoinverse.** The Moore-Penrose pseudoinverse of $A$ is $A^+ = V\Sigma^+ U^T$, where $\Sigma^+$ is obtained by taking the reciprocal of each nonzero entry of $\Sigma$ and transposing. The least squares solution with minimum norm is $\hat{x} = A^+ b$.

## 33 The Invertible Matrix Theorem

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

1. $A$ is invertible.
2. $A$ is row equivalent to $I_n$ (RREF of $A$ is $I_n$).
3. $A$ has $n$ pivot positions.
4. $Ax = \mathbf{0}$ has only the trivial solution.
5. The columns of $A$ are linearly independent.
6. $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.
7. The columns of $A$ span $\mathbb{R}^n$.
8. There exists $C$ such that $CA = I_n$ (left inverse).
9. There exists $D$ such that $AD = I_n$ (right inverse).
10. $A^T$ is invertible.
11. $\det(A) \neq 0$.
12. $\text{col}(A) = \mathbb{R}^n$.
13. $\text{rank}(A) = n$.
14. $\text{null}(A) = \{\mathbf{0}\}$.
15. $\text{nullity}(A) = 0$.
16. The eigenvalues of $A$ are all nonzero.
17. 0 is not a singular value of $A$.