• Let \( A = (a_{ij}) \) be an \( n \times n \) matrix and let \( M_{ij} \) denote the \((n-1) \times (n-1)\) matrix obtained from \( A \) by deleting the row and column containing \( a_{ij} \). The determinant of \( M_{ij} \) is called the minor of \( a_{ij} \). We define the cofactor \( A_{ij} \) of \( a_{ij} \) by
\[
A_{ij} = (-1)^{i+j} \det(M_{ij})
\]

• Subspace: If \( S \) is a nonempty subset of a vector space \( V \), and \( S \) satisfies the conditions
  (i) \( \alpha x \in S \) whenever \( x \in S \) for any scalar \( \alpha \)
  (ii) \( x + y \in S \) whenever \( x \in S \) and \( y \in S \)
then \( S \) is said to be a subspace of \( V \).

• Let \( v_1, v_2, \ldots, v_n \) be vectors in a vector space \( V \). The set
\[
\{ \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{R} \}
\]
is called the span of \( v_1, v_2, \ldots, v_n \) and is denoted by \( \text{Span}(v_1, v_2, \ldots, v_n) \).

• The vectors \( v_1, v_2, \ldots, v_n \) in a vector space \( V \) are said to be linearly independent if
\[
c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0
\]
implies that all the scalars \( c_1 = c_2 = \cdots = c_n = 0 \).

**Theorem 0.1.** Let \( x_1, x_2, \ldots, x_n \) be \( n \) vectors in \( \mathbb{R}^n \) and let
\[
X = (x_1, x_2, \ldots, x_n)
\]
The vectors \( x_1, x_2, \ldots, x_n \) will be linearly independent and span \( \mathbb{R}^n \) if and only if \( X \) is nonsingular.

• If vectors \( v_1, v_2, \ldots, v_n \) are linearly independent and span \( V \), then \( v_1, v_2, \ldots, v_n \) form a basis for \( V \) and \( V \) has dimension \( n \).

**Theorem 0.2.** If vectors \( v_1, v_2, \ldots, v_n \) form a basis for \( V \), then any collection of (strictly) more than \( n \) vectors in \( V \), is linearly dependent.

• The rank of a matrix \( A \), denoted \( \text{rank}(A) \), is the number of non-zero rows in the reduced echelon form of \( A \). The dimension of the null space of a matrix is called the nullity of the matrix.

**Theorem 0.3.** If \( A \) is an \( m \times n \) matrix, then the rank of \( A \) plus the nullity of \( A \) equals \( n \).

• For an \( n \times n \) matrix \( A = (a_{ij}) \), \( p(\lambda) = \det(A - \lambda I) \) is called the characteristic polynomial of \( A \).
\[
\text{tr}(A) = \sum_{i=1}^{n} a_{ii}
\]
is called the trace of \( A \).

**Theorem 0.4.** If \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are the eigenvalues of \( A \), then
\[
\lambda_1 \lambda_2 \cdots \lambda_n = \det(A) = p(0) \tag{0.1}
\]
\[
\lambda_1 + \lambda_2 + \cdots + \lambda_n = \text{tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{0.2}
\]

• An \( n \times n \) matrix \( A \) is said to be diagonalizable if there exists a nonsingular matrix \( X \) and a diagonal matrix \( D \) such that \( X^{-1}AX = D \). We say that \( X \) diagonalizes \( A \).

**Theorem 0.5.** An \( n \times n \) matrix \( A \) is diagonalizable if and only if \( A \) has \( n \) linearly independent eigenvectors.

• A mapping \( L \) from a vector space \( V \) into a vector space \( W \) is said to be a linear transformation if for all \( v_1, v_2 \in V \) and all scalars \( \alpha \)
  (i) \( L(v_1 + v_2) = L(v_1) + L(v_2) \)
  (ii) \( L(\alpha v_1) = \alpha L(v_1) \)
Let \( L : V \rightarrow W \) be a linear transformation. Let \( 0_V \) and \( 0_W \) be the zero vectors in \( V \) and \( W \), respectively. The kernel of \( L \), denoted \( \ker(L) \), is defined by
\[
\ker(L) = \{ v \in V \mid L(v) = 0_W \}
\]
Let \( S \) be a subspace of \( V \). The image of \( S \), denoted \( L(S) \), is defined by
\[
L(S) = \{ L(v) \mid v \in S \}
\]
The image of the entire vector space, \( L(V) \), is called the range of \( L \).

• Let \( L : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be a linear transformation. An \( m \times n \) matrix \( A \) is called the (standard) matrix representation of \( A \) if
\[
L(x) = Ax, \quad x \in \mathbb{R}^n
\]

**Theorem 0.6.** For any linear transformation \( L : \mathbb{R}^n \rightarrow \mathbb{R}^m \), \( L \) has an \( m \times n \) matrix representation \( A \). Moreover,
\[
A = (L(e_1), L(e_2), \ldots, L(e_n))
\]
where \( e_1, \ldots, e_n \) is the standard basis of \( \mathbb{R}^n \).
Let $\langle x, y \rangle = x^T y$ be the usual inner product in $\mathbb{R}^n$. Then

1. $\langle x, x \rangle \geq 0$ with equality if and only if $x = 0$.
2. $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in \mathbb{R}^n$.
3. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ for all $x, y, z \in \mathbb{R}^n$ and all scalars in $\alpha$ and $\beta$.

Let $\|x\| = \sqrt{\langle x, x \rangle}$ be the usual 2-norm in $\mathbb{R}^n$. Then

1. $\|x\| \geq 0$ with equality if and only if $x = 0$.
2. $\|\alpha x\| = |\alpha| \|x\|$ for any scalar $\alpha$.
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

**Theorem 0.7** (The Pythagorean Law). If $x, y$ are orthogonal vectors in $\mathbb{R}^n$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

- If $y \neq 0$, then the scalar projection of $x$ onto $y$ is given by
  $$ \alpha = \frac{\langle x, y \rangle}{\|y\|} $$
  and the vector projection of $x$ onto $y$ is given by
  $$ p = \alpha \left( \frac{y}{\|y\|} \right) = \frac{\langle x, y \rangle}{\|y\|^2} y $$

**Theorem 0.8.** If $y \neq 0$ and $p$ is the vector projection of $x$ onto $y$, then $x - p$ and $p$ are orthogonal.

Two subspaces $X$ and $Y$ of $\mathbb{R}^n$ are said to be orthogonal if $\langle x, y \rangle = 0$ for every $x \in X$ and every $y \in Y$. If $X$ and $Y$ are orthogonal, we write $X \perp Y$.

Let $Y$ a subspace of $\mathbb{R}^n$. The set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in $Y$ will be denoted $Y^\perp$. Thus,

$$ Y^\perp = \{x \in \mathbb{R}^n \mid \langle x, y \rangle = 0 \text{, for every } y \in Y \} $$

The set $Y^\perp$ is called the orthogonal complement of $Y$.

**Theorem 0.9.** Let $v_1, v_2, \cdots, v_m$ be $m$ vectors in $\mathbb{R}^n$. Let $A = \begin{pmatrix} v_1^T \\ \vdots \\ v_m^T \end{pmatrix}$ be an $m \times n$ matrix. If $V = \text{span}(v_1, v_2, \cdots, v_m)$, then $V^\perp = N(A)$.

**Theorem 0.10.** If $A$ is an $m \times n$ matrix of rank $n$, the equations

$$ A^T Ax = A^T b $$

have a unique solution

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

and $\hat{x}$ is the unique least squares solution of the system $Ax = b$.

- Let $v_1, v_2, \cdots, v_n$ be non-zero vectors in $\mathbb{R}^m$. If $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$, then $\{v_1, v_2, \cdots, v_n\}$ is said to be an orthogonal set in $\mathbb{R}^m$. An orthonormal set of vectors is an orthogonal set of unit vectors.

**Theorem 0.11.** If $u_1, u_2, \cdots, u_n$ is an orthonormal set in $\mathbb{R}^n$, then $u_1, u_2, \cdots, u_n$ is an orthonormal basis for $\mathbb{R}^n$. And if $v = \sum_{i=1}^n c_i u_i$, then $c_i = \langle v, u_i \rangle$. Moreover,

$$ \|v\|^2 = \sum_{i=1}^n c_i^2. \text{ (Parseval’s Formula)} $$

- An $n \times n$ matrix $Q$ is said to be orthogonal matrix if the column vectors of $Q$ form an orthonormal set in $\mathbb{R}^n$.

**Theorem 0.12.** $Q$ is an orthogonal matrix if and only if $Q^T Q = I$. If $Q$ is an orthogonal matrix, then

1. $Q^T = Q^{-1}$
2. $\langle Qx, Qy \rangle = \langle x, y \rangle$
3. $\|Qx\| = \|x\|$

**Theorem 0.13.** If $u_1, u_2, \cdots, u_k$ is an orthonormal set in $\mathbb{R}^n$, then the orthogonal (vector) projection of a vector $x \in \mathbb{R}^n$ onto $\text{Span}(u_1, u_2, \cdots, u_k)$ is given by

$$ p = \sum_{i=1}^k \langle x, u_i \rangle u_i $$

**Theorem 0.14** (The Gram–Schmidt Process). Let $x_1, x_2, \cdots, x_n$ be a basis for $\mathbb{R}^n$. In Step 1, let

$$ u_1 = \frac{1}{\|x_1\|} x_1 $$

and define $u_2, u_3, \cdots, u_n$ recursively by:

in Step $k + 1$, let

$$ p_k = \sum_{i=1}^k \langle x_{k+1}, u_i \rangle u_i $$

and define $v_{k+1} = x_{k+1} - p_k$, and $u_{k+1} = \frac{1}{\|v_{k+1}\|} v_{k+1}$

Then $u_1, u_2, \cdots, u_n$ is an orthonormal basis for $\mathbb{R}^n$. 