- 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) $\alpha\mathbf{x} \in S$ whenever $\mathbf{x} \in S$ for any scalar α

  (ii) $\mathbf{x} + \mathbf{y} \in S$ whenever $\mathbf{x} \in S$ and $\mathbf{y} \in S$

  then S is said to be a subspace of V.

- Let $v_1, v_2, \cdots, v_n$ be vectors in a vector space V. The set

$$\{\alpha_1 v_1 + \alpha_2 v_2 + \cdots \alpha_n v_n \,|\, \alpha_1, \cdots, \alpha_n \in \mathbb{R}\}$$

  is called the span of $v_1, v_2, \cdots, v_n$ and is denoted by $\text{Span}(v_1, v_2, \cdots, v_n)$.

- The vectors $v_1, v_2, \cdots, v_n$ in a vector space V are said to be linearly independent if

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

  implies that all the scalars $c_1 = c_2 = \cdots = c_n = 0$.

  Theorem 0.1. *Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ be n vectors in $\mathbb{R}^n$ and let*

$$X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$$

  The vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ will be linearly independent and span $\mathbb{R}^n$ if and only if X is nonsingular.

- If vectors $v_1, v_2, \cdots, v_n$ are linearly independent and span V, then $v_1, v_2, \cdots, v_n$ form a basis for V and V has dimension n.

  Theorem 0.2. *If vectors $v_1, v_2, \cdots, v_n$ 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 $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of A, then*

$$\lambda_1 \cdot \lambda_2 \cdots \lambda_n = p(0) = \det(A) \qquad (0.1)$$

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \text{tr}(A) = \sum_{i=1}^{n} a_{ii} \qquad (0.2)$$

- 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 α

  (i) $L(v_1 + v_2) = L(v_1) + L(v_2)$

  (ii) $L(\alpha v_1) = \alpha L(v_1)$

- Let $L : V \to W$ be a linear transformation. Let $\mathbf{0}_V$ and $\mathbf{0}_W$ be the zero vectors in V and W, respectively. The kernel of L, denoted $\ker(L)$, is defined by

$$ker(L) = \{v \in V | L(v) = \mathbf{0}_W\}$$

  Let S be a subspace of V. The image of S, denoted $L(S)$, is defined by

$$L(S) = \{L(v) | v \in S\}$$

  The image of the entire vector space, $L(V)$, is called the range of L.

- Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. An $m \times n$ matrix A is called the (standard) matrix representation of A if

$$L(\mathbf{x}) = A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n$$

  Theorem 0.6. *For any linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$, L has an $m \times n$ matrix representation A. Moreover,*

$$A = (L(\mathbf{e}_1), L(\mathbf{e}_2), \cdots, L(\mathbf{e}_n))$$

  where $\mathbf{e_1}, \cdots, \mathbf{e_n}$ is the standard basis of $\mathbb{R}^n$.

- Let $< \mathbf{x}, \mathbf{y} >= \mathbf{x}^T\mathbf{y}$ be the usual inner product in $\mathbb{R}^n$. Then
    1. $< \mathbf{x}, \mathbf{x} > \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$.
    2. $< \mathbf{x}, \mathbf{y} >=< \mathbf{y}, \mathbf{x} >$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$
    3. $< \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} >= \alpha < \mathbf{x}, \mathbf{z} > +\beta < \mathbf{y}, \mathbf{z} >$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathbb{R}^n$ and all scalars in α and β.
- Let $\|\mathbf{x}\| = \sqrt{< \mathbf{x}, \mathbf{x} >}$ be the usual 2-norm in $\mathbb{R}^n$. Then
    1. $\|\mathbf{x}\| \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$.
    2. $\|\alpha\mathbf{x}\| = |\alpha| \, \|\mathbf{x}\|$ for any scalar α
    3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$.

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

- If $\mathbf{y} \neq \mathbf{0}$, then the scalar projection of $\mathbf{x}$ onto $\mathbf{y}$ is given by

$$\alpha = \frac{< \mathbf{x}, \mathbf{y} >}{\|\mathbf{y}\|}$$

  and the vector projection of $\mathbf{x}$ onto $\mathbf{y}$ is given by

$$\mathbf{p} = \alpha \left(\frac{\mathbf{y}}{\|\mathbf{y}\|} \right) = \frac{< \mathbf{x}, \mathbf{y} >}{\|\mathbf{y}\|^2}\mathbf{y}$$

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

- Two subspaces X and Y of $\mathbb{R}^n$ are said to be orthogonal if $< \mathbf{x}, \mathbf{y} >= 0$ for every $\mathbf{x} \in X$ and every $\mathbf{y} \in Y$. 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,

$$Y^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid < \mathbf{x}, \mathbf{y} >= 0 \ , \quad \text{for every} \quad \mathbf{y} \in Y\}$$

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

  Theorem 0.9. *Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ be m vectors in $\mathbb{R}^n$. Let $A = \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T \end{pmatrix}$ be an $m \times n$ matrix. If $V = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)$, then $V^\perp = N(A)$.*

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

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

  have a unique solution

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

  and $\widehat{\mathbf{x}}$ is the unique least squares solution of the system $A\mathbf{x} = \mathbf{b}$.

- Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be non-zero vectors in $\mathbb{R}^m$. If $< \mathbf{v}_i, \mathbf{v}_j >= 0$ whenever $i \neq j$, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 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 $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ is an orthonormal set in $\mathbb{R}^n$, then $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ is an orthonormal basis for $\mathbb{R}^n$. And if $\mathbf{v} = \sum_{i=1}^n c_i u_i$, then $c_i =< \mathbf{v}, \mathbf{u}_i >$. Moreover,*

$$\|\mathbf{v}\|^2 = \sum_{i=1}^n c_i^2. \quad \textit{(Parseval's Formula)}$$

- 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. *$< Q\mathbf{x}, Q\mathbf{y} >=< \mathbf{x}, \mathbf{y} >$*
    3. *$\|Q\mathbf{x}\| = \|\mathbf{x}\|$*

- **Theorem 0.13.** *If $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$ is an orthonormal set in $\mathbb{R}^n$, then the orthogonal (vector) projection of a vector $\mathbf{x} \in \mathbb{R}^n$ onto* Span$(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$ *is given by*

$$\mathbf{p} = \sum_{i=1}^k < \mathbf{x}, \mathbf{u}_i > \mathbf{u}_i$$

- **Theorem 0.14** (The Gram–Schmidt Process)**.** *Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ be a basis for $\mathbb{R}^n$. In Step 1, let*

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1$$

  and define $\mathbf{u}_2, \mathbf{u}_2, \cdots, \mathbf{u}_n$ recursively by:

  in Step $k + 1$, let

$$\mathbf{p}_k = \sum_{i=1}^k < \mathbf{x}_{k+1}, \mathbf{u}_i > \mathbf{u}_i$$

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \mathbf{p}_k, \quad \text{and} \quad \mathbf{u}_{k+1} = \frac{1}{\|\mathbf{v}_{k+1}\|}\mathbf{v}_{k+1}$$

  Then $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ is an orthonormal basis for $\mathbb{R}^n$.