Chapter 2

Vector Spaces and Matrices

2.1 Definitions

Definition: A (real) vector space consists of a non empty set $\mathbf{V}$ and two operations.

- The first operation, addition, is defined for pairs of elements in $\mathbf{V}$, $\mathbf{x}$ and $\mathbf{y}$, and yields an element in $\mathbf{V}$, denoted by $\mathbf{x} + \mathbf{y}$.
- The second operation, scalar multiplication, is defined for the pair α, a real number, and an element $\mathbf{x} \in \mathbf{V}$, and yields an element in $\mathbf{V}$ denoted by $\alpha\mathbf{x}$.

Eight properties are assumed to hold for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}, \alpha, \beta, 1 \in \mathbf{R}$:

(1) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(2) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

(3) There is an element in $\mathbf{V}$ denoted $\mathbf{0}$ such that

$$\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$$

(4) For each $\mathbf{x} \in \mathbf{V}$ there is an element in $\mathbf{V}$ denoted $-\mathbf{x}$ such that

$$\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0}$$

(5) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for all α

(6) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for all α, β

(7) $1\mathbf{x} = \mathbf{x}$

(8) $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ for all α, β

- The elements of $\mathbf{V}$ are called vectors and the elements in $\mathbf{R}$ are called scalars.
- Elements of $\mathbf{V}$ will be denoted by bold faced roman letters, i.e. $\mathbf{x}$.
- Although a vector space is defined as the triple consisting of the set $\mathbf{V}$ and the two operations, we will abuse notation slightly by referring to $\mathbf{V}$ as a vector space rather than $\mathbf{V}, +, \cdot$, where $\cdot$ refers to scalar multiplication.

If $\mathbf{V}$ is a vector space then it is easy to show that

- $0\mathbf{x} = \mathbf{0}$
- $(-1)\mathbf{x} = -\mathbf{x}$
- $\alpha\mathbf{0} = \mathbf{0}$

example 1: $\mathbf{V} =$ set of all ordered pairs of real numbers i.e.

$$\mathbf{V} = \mathbf{R}^2 = \{(x_1, x_2) : \; x_1 \in R, x_2 \in R\}$$

- "+" is defined by $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2)$
- "$\alpha\mathbf{x}$" is defined by $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2)$

example 2: $V =$ set of all ordered n tuples of real numbers i.e. $\mathbf{V} = \mathbf{R}^n$.

- "+" defined by $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$
- "$\alpha\mathbf{x}$" defined by $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)$

2.2 Geometry of Euclidean Space

When $\mathbf{V} = \mathbf{R}^n$ elements of the vector space and the operations of addition and scalar multiplication may be represented by geometric vectors. Geometric vectors correspond to the notion of a vector in $\mathbf{R}^n$ as a directed line segment.

Definition: If P and Q represent points in $\mathbf{R}^n$ a **located vector** beginning at P and ending at Q is represented by $\overrightarrow{PQ}$.

- Two located vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ are said to be **equivalent** if $Q-P = S-R$
- Every located vector $\overrightarrow{PQ}$ is equivalent to the located vector beginning at the origin $O = (0, 0, \ldots, 0)$ and ending at $Q - P$ i.e. to $\overrightarrow{O(Q-P)}$
- Two located vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ are said to be
  - **Parallel** if there is a real number $c \neq 0$ such that $Q - P = c(S - R)$
  - In the **same direction** if $Q-P = c(S-R)$ for some real number $c > 0$.
  - In the **opposite direction** if $Q - P = c(S - R)$ for some real number $c < 0$

Definition: A geometric vector in R^n corresponding to an algebraic vector in R^n (i.e., an ordered n tuple $\mathbf{x} = (x_1, \ldots, x_n)$) is the equivalence class of all directed line segments $\overrightarrow{PQ}$ in $\mathbf{R}^n$ that are formed by points P and Q whose coordinates $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_n)$ respectively, satisfy the equations

$$p_i - q_i = x_i;\ 1 \leq i \leq n$$

- A single member of this class of directed line segments is typically chosen to represent the entire class namely the one in which $P = (0, \ldots, 0)$.
- The operations "vector addition" and "scalar multiplication" on algebraic vectors correspond to simple operations on geometric vectors.
- An interpretation of the sum of geometric vectors is the vector corresponding to the diagonal of a parallelogram constructed in the "plane" determined by the two vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ having $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ for adjacent sides. The diagonal is chosen with one end at the origin.
- Scalar multiplications of a vector corresponds to a change in length and possibly a reversal in direction of the associated geometric vector. The geometric vector associated with αP is denoted $\alpha\,\overrightarrow{OP}$.

Definition: The **scalar product** of $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ is

$$\overrightarrow{OP} \circ \overrightarrow{OQ} = \sum_{i=1}^{n} p_i q_i$$

Scalar products have the following four properties

- $\overrightarrow{OP} \circ \overrightarrow{OQ} = \overrightarrow{OQ} \circ \overrightarrow{OP}$
- $\overrightarrow{OR} \circ (\overrightarrow{OP} + \overrightarrow{OQ}) = \overrightarrow{OR} \circ \overrightarrow{OP} + \overrightarrow{OR} \circ \overrightarrow{OQ}$
- $(c\, \overrightarrow{OR}) \circ \overrightarrow{OQ} = c(\overrightarrow{OR} \circ \overrightarrow{OQ})$
- $\overrightarrow{OP} \circ \overrightarrow{OP} = 0$ if and only if $\overrightarrow{OP}$ is the vector O. Otherwise $\overrightarrow{OP} \circ \overrightarrow{OP} > 0$

Definition: The **norm** or length of $\overrightarrow{OP}$ is defined as

$$\| \overrightarrow{OP} \| = \sqrt{\overrightarrow{OP} \circ \overrightarrow{OP}}$$

Definition: The **distance** between $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ is defined as

$$\| \overrightarrow{OP} - \overrightarrow{OQ} \|$$

Distances have the following properties:

- $\| \overrightarrow{OP} - \overrightarrow{OQ} \| = \| \overrightarrow{OQ} - \overrightarrow{OP} \|$
- $\|c \overrightarrow{OP} \| = |c| \| \overrightarrow{OP} \|$
- $\| \overrightarrow{OP} + \overrightarrow{OQ} \| = \| \overrightarrow{OP} - \overrightarrow{OQ} \|$ if and only if $\overrightarrow{OP} \circ \overrightarrow{OQ} = 0$

Definition: If $\overrightarrow{OP} \circ \overrightarrow{OQ} = 0$ we say that $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ are perpendicular or **orthogonal**.

Pythagorean Theorem: If $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ are orthogonal then

$$\| \overrightarrow{OP} + \overrightarrow{OQ} \|^2 = \| \overrightarrow{OP} \|^2 + \| \overrightarrow{OQ} \|^2$$

Let $\overrightarrow{OP}$ be any vector and let $\overrightarrow{OQ}$ be any non zero vector. Let C be the point such that $\overrightarrow{CP}$ is orthogonal to $\overrightarrow{OQ}$. Then

- $\overrightarrow{OC}= c\,\overrightarrow{OQ}$ and $\overrightarrow{OP} - \overrightarrow{OC}$ is orthogonal to $\overrightarrow{OQ}$
- The number c is unique and is given by

$$c = \frac{\overrightarrow{OP} \circ \overrightarrow{OQ}}{\| \overrightarrow{OQ} \|}$$

- c is called the **component** of $\overrightarrow{OP}$ along $\overrightarrow{OQ}$
- The **projection** of $\overrightarrow{OP}$ along $\overrightarrow{OQ}$ is $c\,\overrightarrow{OQ}$ where c is the component.

Cauchy-Schartz Inequality For any two vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ we have that

$$| \overrightarrow{OP} \circ \overrightarrow{OQ} | \leq \| \overrightarrow{OP} \| \| \overrightarrow{OQ} \|$$

Triangle Inequality For any two vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ we have that

$$\| \overrightarrow{OP} + \overrightarrow{OQ} \| \leq \| \overrightarrow{OP} \| + \| \overrightarrow{OQ} \|$$

- From the Cauchy-Schwartz Inequality we have that

$$-1 \leq \frac{\overrightarrow{OP} \circ \overrightarrow{OQ}}{\| \overrightarrow{OQ} \| \dot{\|} \| \overrightarrow{OQ} \|} \leq +1$$

- Thus there is a unique θ, $0 \leq \theta \leq \pi$ such that

$$\cos(\theta) = \frac{\overrightarrow{OP} \circ \overrightarrow{OQ}}{\| \overrightarrow{OQ} \| \dot{\|} \| \overrightarrow{OQ} \|}$$

- θ is called the **angle** between $\overrightarrow{OP}$ and $\overrightarrow{OQ}$

Some miscellaneous geometric concepts are:

- The set of points X such that $\| \overrightarrow{OX} - \overrightarrow{OP} \| < k$ is called the **open ball** of radius k centered at $\overrightarrow{OP}$. If $<$ is replaced by $\leq$ we get the **closed ball** and if $<$ is replaced by $=$ we get a **sphere**.
- The **line** through P and Q is defined as the set of all X such that

$$X = \lambda P + (1 - \lambda) Q$$

  The line passing through O and P is the set of all points which satisfy $X = \lambda P$

- The **plane** through points $P_1, P_2, \ldots, P_k$ is the set of all points X such that

$$X = \sum_{j=1}^{k} \lambda_j P_j \quad \text{where} \quad \sum_{j=1}^{k} \lambda_j = 1$$

The plane through $O, P_1, P_2, \ldots, P_k$ is the set of all points X such that

$$X = \sum_{j=1}^{k} \lambda_j P_j$$

2.3 Subspaces of Vector Spaces

Definition: If $\mathbf{W}$ is a non-empty subset of a vector space $\mathbf{V}$ closed under addition and scalar multiplication then $\mathbf{W}$ is called a subspace of $\mathbf{V}$.

Note: If $\mathbf{W}$ is a subspace of $\mathbf{V}$, then $\mathbf{W}$ is itself a vector space.

example 1: If $\mathbf{V}$ is a vector space, both $\{\mathbf{0}\}$ and $\mathbf{V}$ are subspaces. $\mathbf{V}$ is the largest subspace in that it contains every other subspace and $\{\mathbf{0}\}$ is the smallest in the sense that it is contained in every subspace of $\mathbf{V}$.

example 2: Any line in $\mathbf{R}^2$ (or $\mathbf{R}^3$) which passes through the origin is a subspace of $\mathbf{R}^2$ (respectively $\mathbf{R}^3$).

example 3: Let $\mathbf{V} = \mathbf{R}^n$ and let

$$\mathbf{W}_i = \{(x_1, x_2, \ldots, x_i, 0, \ldots, 0) : \; x_k \in R\,; \; 1 \leq k \leq i\}$$

Then $\mathbf{W}_i$ is a subspace of $\mathbf{V}$, and for $1 \leq i \leq n-1$, we have $\mathbf{W}_i \subset \mathbf{W}_{i+1}$
.

Definition: Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of vectors in a vector space $\mathbf{V}$. The vector $\mathbf{x} \in \mathbf{V}$ is a **linear combination** of the vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ if there are scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_n \mathbf{x}_n$$

Theorem 2.1: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be elements of a vector space $\mathbf{V}$ and let $\mathbf{W}$ be the subset of $\mathbf{V}$ consisting of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ i.e.

$$\mathbf{W} = \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{x}_i : \ \alpha_i \in R\,; \ 1 \leq i \leq n \right\}$$

Then $\mathbf{W}$ is a subspace of $\mathbf{V}$.

Definition: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors of the space $\mathbf{V}$. The subspace of all linear combinations of these vectors will be denoted $\mathbf{Sp}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and the $\mathbf{x}_i$'s will be called generators of this subspace.

Definition: If $\mathbf{W}$ is a subspace of $\mathbf{V}$ and $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are vectors in $\mathbf{W}$ such that $\mathbf{W} = \mathbf{Sp}(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ then we say that the $\mathbf{y}_i$'s are generators of $\mathbf{W}$ or that $\mathbf{W}$ is generated or **spanned** by the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$.

example 1: Let $\mathbf{V} = \mathbf{R}^n$ and $\mathbf{e}_i = (0, 0, \ldots, 1, \ldots, 0);\ 1 \leq i \leq n$. (i.e., $\mathbf{e}_i$ is the ith standard unit vector in $\mathbf{R}^n$). Then $\mathbf{V} = \mathbf{Sp}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

example 2: Let $\mathbf{V} = \mathbf{R}^3$ and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be the standard unit vectors in $\mathbf{R}^3$ Let $\mathbf{f}_1 = (1, 1, 0)$ and $\mathbf{f}_2 = (1, 1, 1)$. Consider the following subspaces:

$$\mathbf{W}_1 = \mathbf{Sp}(\mathbf{0});\ \mathbf{W}_2 = \mathbf{Sp}(\mathbf{f}_1);\ \mathbf{W}_3 = \mathbf{Sp}(\mathbf{f}_2)$$

$$\mathbf{W}_4 = \mathbf{Sp}(\mathbf{e}_1, \mathbf{e}_2);\ \mathbf{W}_5 = \mathbf{Sp}(\mathbf{f}_1, \mathbf{e}_1);\ \mathbf{W}_6 = \mathbf{Sp}(\mathbf{f}_1, \mathbf{e}_3)$$

Now $\mathbf{W}_2$ and $\mathbf{W}_3$ are lines through the origin in the direction of $\mathbf{f}_1$ and $\mathbf{f}_2$, respectively, and $\mathbf{W}_1 = \mathbf{W}_2 \cap \mathbf{W}_3$. Also, $\mathbf{W}_4 = \mathbf{W}_5$ and $\mathbf{W}_2 = \mathbf{W}_4 \cap \mathbf{W}_6$.

- The idea of a subspace being spanned by a set of vectors can be extended to a subspace being spanned by a set of other subspaces.
- Select a (finite) set of generators from each of the subspaces, form their set union and consider the subspace of linear combinations of these vectors.

Definition: Let $\mathbf{W}_1, \ldots, \mathbf{W}_n$ be subspaces of $\mathbf{V}$. The set spanned by $\mathbf{W}_1, \ldots, \mathbf{W}_n$ (called the sum of $\mathbf{W}_1, \ldots, \mathbf{W}_n$) is denoted $\mathbf{W}_1 + \ldots + \mathbf{W}_n$ and is defined as the set

$$\mathbf{W}_1 + \ldots + \mathbf{W}_n = \{\mathbf{y} \in \mathbf{V} : \ \mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 + \cdots + \mathbf{y}_n \ \text{ for } \ \mathbf{y}_i \in \mathbf{W}_i; \ 1 \leq i \leq n\}$$

Note that if $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$ are subspaces of $\mathbf{V}$ then

$$\mathbf{W}_1 + \mathbf{W}_2 + \cdots + \mathbf{W}_n$$

is also a subspace of $\mathbf{V}$.

example 1: Let $\mathbf{V} = \mathbf{R}^3$ and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be the standard unit vectors in $\mathbf{R}^3$. If $\mathbf{W}_1 = \mathbf{Sp}(\mathbf{e}_1)$ and $\mathbf{W}_2 = \mathbf{Sp}(\mathbf{e}_2, \mathbf{e}_3)$ then $\mathbf{V} = \mathbf{W}_1 + \mathbf{W}_2$.

Note that the sum of subspaces is not their set union. $\mathbf{W}_1 + \mathbf{W}_2$ always contains $\mathbf{W}_1 \cup \mathbf{W}_2$ but $\mathbf{W}_1 \cup \mathbf{W}_2$ equals $\mathbf{W}_1 + \mathbf{W}_2$ only in the special case in which $\mathbf{W}_1 \subset \mathbf{W}_2$ or $\mathbf{W}_2 \subset \mathbf{W}_1$.

Theorem 2.2 Let $\mathbf{V}$ be a vector space and $\mathbf{W}_1, \mathbf{W}_2$ be subspaces of $\mathbf{V}$ spanned by $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$, respectively. If each $\mathbf{x}_i$ is a linear combination of the $\mathbf{y}_j$'s, then $\mathbf{W}_1 \subset \mathbf{W}_2$.

Corollary 2.3 Let $\mathbf{W}_1 = \mathbf{Sp}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{W}_2 = \mathbf{Sp}(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ be subspaces of $\mathbf{V}$. Then if each $\mathbf{y}_j$ is a linear combination of the $\mathbf{x}_i$'s and each $\mathbf{x}_i$ is a linear combination of the $\mathbf{y}_j$'s, we have $\mathbf{W}_1 = \mathbf{W}_2$.

2.4 Linear Dependence

Definition: The vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in $\mathbf{V}$ are said to be **linearly dependent** if real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ exist such that at least one is different from zero and

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{0}$$

If this is not the case, we say that the vectors are **linearly independent**.

Note that the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent if and only if any equation of the form

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{0} \ \text{ implies } \ \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

Theorem 2.4 The vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly dependent if and only if at least one of the vectors is a linear combination of the others.

Corollary 2.5 Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a linearly independent set of vectors in $\mathbf{V}$ and let $\mathbf{x} \in \mathbf{V}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{x}\}$ are linearly dependent if and only $\mathbf{x}$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

example Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be standard unit vectors in $\mathbf{R}^3$ and $\mathbf{x} = \mathbf{e}_1 + \mathbf{e}_2$. The sets $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, $\{\mathbf{e}_1, \mathbf{e}_3, \mathbf{x}\}$ are linearly independent. However, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{x}\}$ is a linearly dependent set.

2.5 Direct Sums of Subspaces

The concept of linear independence can also be extended to subspaces. Suppose $\mathbf{W}_1$ and $\mathbf{W}_2$ are two subspaces of $\mathbf{V}$ each different than $\mathbf{0}$ such that any pair of vectors $\mathbf{x}_1 \in \mathbf{W}_1$ and $\mathbf{x}_2 \in \mathbf{W}_2$ are linearly independent.

Theorem 2.6 Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be subspaces of $\mathbf{V}$. Then the following conditions are equivalent:

1. $\mathbf{W}_1 \cap \mathbf{W}_2 = \{\mathbf{0}\}$
2. $\mathbf{x}_1 \in \mathbf{W}_1, \mathbf{x}_2 \in \mathbf{W}_2$ and $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{0}$ implies $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{0}$. (i.e., any pair of non-zero vectors $\mathbf{x}_1 \in \mathbf{W}_1, \mathbf{x}_2 \in \mathbf{W}_2$ are linearly independent)

Definition: Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be subspaces of the vector space $\mathbf{V}$. If $\mathbf{W}_1 \cap \mathbf{W}_2 = \{\mathbf{0}\}$, the space $\mathbf{W}_1 + \mathbf{W}_2$ is called the **direct sum** of $\mathbf{W}_1$ and $\mathbf{W}_2$. The notation $\mathbf{W}_1 \oplus \mathbf{W}_2$ is used for the direct sum of $\mathbf{W}_1$ and $\mathbf{W}_2$.

Definition: If $\mathbf{W}_1$ and $\mathbf{W}_2$ are subspaces of $\mathbf{V}$ such that $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2$, we say that $\mathbf{W}_1$ and $\mathbf{W}_2$ are complements of one another in $\mathbf{V}$.

example: If $\mathbf{V} = \mathbf{R}^2$ then $\mathbf{W}_1 = \mathbf{Sp}(\mathbf{e}_1)$ has an infinite number of complements. i.e. if $\{\mathbf{e}_1, \mathbf{x}\}$ is linearly independent, then $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{Sp}(\mathbf{x})$.

2.6 Bases and Dimension of a Vector Space

Definition: The set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of vectors in a vector space $\mathbf{V}$ is called a (finite) basis of $\mathbf{V}$ if

- the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent
- $\mathbf{V} = \mathbf{Sp}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$

example 1: $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form a basis of $\mathbf{R}^n$.

example 2: $\mathbf{f}_1 = (1, 0), \mathbf{f}_2 = (1, 1)$ form a basis for $\mathbf{R}^2$.

Theorem 2.7 If $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a basis for $\mathbf{V}$, then any vector $\mathbf{x} \in \mathbf{V}$ can be uniquely expressed as a linear combination of the $\mathbf{x}_i$'s.

Definition: If $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a basis for the vector space $\mathbf{V}$, the scalar α_i in the unique representation of $\mathbf{x}$ is called the ith **coordinate** of $\mathbf{x}$ (relative to this basis.)

example: $\mathbf{V} = \mathbf{R}^2$. Let $\mathbf{e}_1, \mathbf{e}_2$ be standard unit vectors in $\mathbf{R}^2$ and $\mathbf{f} = (1, 1)$ Then $\mathbf{V} = \mathbf{Sp}(\mathbf{e}_1, \mathbf{f}_1)$ and $\mathbf{e}_1, \mathbf{f}_1$ are linearly independent so $\{\mathbf{e}_1, \mathbf{f}_1\}$ is a basis. Now if $\mathbf{x} = (-1)\mathbf{e}_1 + (-3)\mathbf{e}_2$ then $(-1, -3)$ are the coordinates of $\mathbf{x}$ relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$. However the coordinates of $\mathbf{x}$ relative to the basis $\{\mathbf{e}_1, \mathbf{f}_1\}$ are $(2, -3)$.

Note that $\mathbf{V} = \mathbf{Sp}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{f}_1)$ but $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{f}_1\}$ are linearly dependent so they do not form a basis (i.e. $1\mathbf{e}_1 + 1\mathbf{e}_2 + (-1)\mathbf{f}_1 = \mathbf{0}$).

Note also that you can express a vector $\mathbf{x}$ in terms of "coordinates" relative to a set of generators that are not a basis, however the "coordinates" are not uniquely determined i.e.

$$
\begin{aligned}
\mathbf{x} &= (-1)\mathbf{e}_1 + (-3)\mathbf{e}_2 + (0)\mathbf{f}_1 \\
\mathbf{x} &= (+2)\mathbf{e}_1 + (-3)\mathbf{e}_2 + (-3)\mathbf{f}_1
\end{aligned}
$$

Theorem 2.8 Any two bases for a vector space $\mathbf{V}$ contain the same number of vectors.

example: Let $\mathbf{x} \in \mathbf{V}$ and let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ and $\mathbf{y}_1, \cdots, \mathbf{y}_n$ be two bases for $\mathbf{V}$. Then

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \ \text{ and } \ \mathbf{x} = \sum_{j=1}^{n} \beta_j \mathbf{y}_j$$

But

$$\mathbf{y}_j = \sum_{i=1}^{n} \gamma_{ji} \mathbf{x}_i$$

so that

$$\begin{aligned} \sum_{i=1}^{n} \alpha_i \mathbf{x}_i &= \sum_{j=1}^{n} \beta_j \left(\sum_{i=1}^{n} \gamma_{ji} \mathbf{x}_i \right) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \beta_j \gamma_{ji} \mathbf{x}_i \right) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} (\beta_j \gamma_{ji} \right) \mathbf{x}_i \end{aligned}$$

Thus $\alpha_i = \sum_{j=1}^{n} \beta_j \gamma_{ji}$ relates the representation in terms of the two bases.

Corollary 2.9 Let $\mathbf{V}$ be a vector space with basis $\mathbf{x}_1, \cdots, \mathbf{x}_n$ if $m > n$, then any set of m vectors of $\mathbf{V}$ is linearly dependent.

Definition: A vector space $\mathbf{V}$ with a finite basis is said to have dimension n if the number of vectors in the (any) basis is n. We say $\mathbf{V}$ has dimension n and write $\dim(\mathbf{V}) = n$. The space $\{\mathbf{0}\}$ is said to have dimension zero.

Theorem 2.10 Any linearly independent set of vectors of a finite dimensional vector space can be enlarged to a basis.

Theorem 2.11 Let $\mathbf{V}$ be a (finite dimensional) vector space with subspaces $\mathbf{W}_1$ and $\mathbf{W}_2$. Then

(1) $\dim(\mathbf{W}_i) \leq \dim(\mathbf{V})$ and any basis for $\mathbf{W}_i$ can be enlarged to a basis for $\mathbf{V}$

(2) If $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2$ then a basis for $\mathbf{V}$ is obtained by combining any basis for $\mathbf{W}_1$ with any basis for $\mathbf{W}_2$. Also $\dim(\mathbf{V}) = \dim(\mathbf{W}_1) + \dim(\mathbf{W}_2)$.

2.7 Inner Product Spaces

Definition: An inner product space consists of a vector space $\mathbf{V}$, together with a real valued function $\langle \mathbf{x}, \mathbf{y} \rangle$ called the inner product on $\mathbf{V}$ which is defined for each ordered pair of vectors $\mathbf{x}$,$\mathbf{y}$ in $\mathbf{V}$ and which has the following four properties:

(1) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ (symmetry)

(2) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$ (positive definite)

(3) $\langle \mathbf{x}+\mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ and $\langle \mathbf{x}, \mathbf{y}+\mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$ (bilinearity)

(4) $\langle \lambda\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \lambda\mathbf{y} \rangle = \lambda\langle \mathbf{x}, \mathbf{y} \rangle$ (homogeneity)

Definition: Let $\mathbf{V}$ be an inner product space. Then

(1) The length of a vector $\mathbf{x} \in \mathbf{V}$ is the non-negative real number $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, denoted $\|\mathbf{x}\|$, and called the **norm** of $\mathbf{x}$

(2) Two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{V}$ are said to be **orthogonal** if $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{0}$. We write in this case $\mathbf{x} \perp \mathbf{y}$

(3) The **distance** between two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{V}$ is the real number

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$$

(4) If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbf{V}$ the **projection** of $\mathbf{y}$ on $\mathbf{x}$ is the vector

$$\hat{\mathbf{y}} = \left(\frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \right) \mathbf{x}$$

Definition: Let $\mathbf{V} = \mathbf{R}^n$ with basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. The standard inner product in $\mathbf{R}^n$ is defined by

$$\langle \mathbf{y}, \mathbf{x} \rangle = \alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_n\beta_n$$

where $\alpha_1, \alpha_2, \cdots + \alpha_n, \beta_1, \beta_2, \cdots, \beta_n$ are the coordinates of $\mathbf{x}$ and $\mathbf{y}$, respectively, relative to the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

example 1: Let $\mathbf{V} = \mathbf{R}^2$ and $\langle \cdot, \cdot \rangle$ be the standard inner product. Then the definition of length, distance, orthogonality and projection correspond to the usual definition in Euclidean geometry. If $\mathbf{x} = \mathbf{e}_1 + \mathbf{e}_2$ then

$$\begin{aligned}
\|\mathbf{x}\| &= \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \\
&= \sqrt{1^2 + 1^2} \\
&= \sqrt{2} \\
\langle \mathbf{e}_1, \mathbf{e}_2 \rangle &= 1 \cdot 0 + 0 \cdot 1 \\
&= 0 \ \text{ so that } \ \mathbf{e}_1 \perp \mathbf{e}_2 \\
d(\mathbf{x}, \mathbf{e}_1) &= \|\mathbf{x} - \mathbf{e}_1\| \\
&= \sqrt{\langle \mathbf{x} - \mathbf{e}_1, \mathbf{x} - \mathbf{e}_1 \rangle} \\
&= \sqrt{0^2 + 1^2} \\
&= 1
\end{aligned}$$

The projection of $\mathbf{x}$ onto $\mathbf{e}_1$ is

$$\hat{\mathbf{x}} = \left(\frac{\langle \mathbf{x}, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle} \right) \mathbf{e_1} = \mathbf{e}_1$$

example 2: Let $\mathbf{V} = \mathbf{R}^n$ with basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and define $\langle \mathbf{x}, \mathbf{y} \rangle$ by

$$\alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_n\beta_n$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n, \beta_1, \beta_2, \cdots, \beta_n$ are the coordinates of $\mathbf{x}$ and $\mathbf{y}$ relative to the basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The standard inner product in $\mathbf{R}^n$ is a special case when

$$\mathbf{x}_i = \mathbf{e}_i \;\; 1 \le i \le n$$

Note, however, that when

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \neq \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$

the resulting definitions corresponding to length, orthogonality, etc. do not correspond to the usual Euclidean geometric definitions. In particular, with the standard inner product, each vector in the basis has unit length and any two basis vectors are orthogonal.

example 3: Let $\mathbf{A}$ be an $n \times n$ positive definite symmetric matrix. Then if α, β are coordinates of $\mathbf{x}, \mathbf{y}$ relative to $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, define

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j A_{ij}$$

where A_{ij} is the (i, j)th element of the matrix $\mathbf{A}$ and note that this satisfies the requirements of an inner product.

Definition: If $\mathbf{x}$, $\mathbf{y}$ are non-zero vectors in $\mathbf{V}$, then the angle θ between $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

This relationship between $\langle \mathbf{x}, \mathbf{y} \rangle$ and θ in $\mathbf{R}^n$ may be represented by geometric considerations in the plane spanned by $\mathbf{x}$ and $\mathbf{y}$.

2.7.1 Fundamental Inequalities in Inner Product Spaces

There are several results useful in analytic geometry that have immediate extensions to inner product spaces. These results are:

Theorem 2.12: (The Pythagorean Theorem). Let $\mathbf{x}$,$\mathbf{y}$ be vectors in $\mathbf{V}$. Then $\langle \mathbf{y}, \mathbf{x} \rangle = 0$ if and only if

$$\|\mathbf{y} + \mathbf{x}\|^2 = \|\mathbf{y}\|^2 + \|\mathbf{x}\|^2$$

Lemma 2.13: (Cauchy-Schwarz Inequality) If $\mathbf{x}$,$\mathbf{y}$ are vectors in $\mathbf{V}$, then

$$|\langle \mathbf{y}, \mathbf{x} \rangle| \leq \|\mathbf{y}\| \|\mathbf{x}\|$$

Moreover,$|\langle \mathbf{y}, \mathbf{x} \rangle| = \|\mathbf{y}\| \|\mathbf{x}\|$ if and only if $\mathbf{y}$ and $\mathbf{x}$ are linearly dependent.

Theorem 2.14: (The Triangle Inequality) If $\mathbf{x}$,$\mathbf{y}$ are vectors in $\mathbf{V}$, then

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

Equality occurs if and only if $\mathbf{x} = \mathbf{0}$ or $\mathbf{x} = \alpha \mathbf{y}$ for some non-negative real number α.

2.7.2 Orthonormal Basis

Definition: A basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ for $\mathbf{V}$ is called an orthonormal basis (ONB) if the vectors $\mathbf{x}_i$ are pairwise orthogonal and have unit norm. i.e.

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

example 1: Let $\mathbf{V} = \mathbf{R}^n$. Then $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, the standard unit vectors in $\mathbf{R}^n$, is an orthonormal basis.

example 2: Let $\mathbf{V} = \mathbf{R}^2$ and

$$\begin{aligned} \mathbf{x}_1 &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \\ \mathbf{x}_2 &= \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \end{aligned}$$

Then $\{\mathbf{x}_1, \mathbf{x}_2\}$ is an ONB.

Note that if a set of vectors (non-zero) are pairwise orthogonal, then they are linearly independent. Thus if dim $(\mathbf{V}) = n$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are pairwise orthogonal, they constitute an orthogonal basis for $\mathbf{V}$. Also, by normalizing the $\mathbf{x}_i$'s, we have that

$$\left\{ \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}, \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|}, \ldots, \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|} \right\}$$

is an ONB for $\mathbf{V}$.

To show that pairwise orthogonality implies linear independence, we suppose that the $\mathbf{x}_i$'s are linearly dependent and arrive at a contradiction. Suppose there exist $\alpha_1, \ldots, \alpha_n$ such that

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = \mathbf{0}$$

Then for all j we have

$$\begin{aligned} 0 &= \langle \mathbf{x}_j, \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \rangle \\ &= \sum_{i=1}^{n} \alpha_i \langle \mathbf{x}_j, \mathbf{x}_i \rangle \\ &= \alpha_j \langle \mathbf{x}_j, \mathbf{x}_j \rangle \end{aligned}$$

But $\langle \mathbf{x}_j, \mathbf{x}_j \rangle = 1$ which implies $\alpha_j = 0$ for all j. Thus the vectors are linearly independent.

The classic result about ONB's is an existence theorem which is proved by construction. The method of construction is called the Gram-Schmidt process.

Theorem 2.15: Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an arbitrary basis for $\mathbf{V}$. Then there exists an orthonormal basis for $\mathbf{V}$, $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ such that each $\mathbf{y}_i$ is a linear combination of the $\mathbf{x}_i$'s.

Using the Gram-Schmidt process, we note that it is always possible to extend an ONB for a subspace $\mathbf{W}$ to an ONB for $\mathbf{V}$. If $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$ is an ONB for $\mathbf{W}$, then append $n - r$ vectors $\{\mathbf{x}_{r+1}, \ldots, \mathbf{x}_n\}$ so that $\{\mathbf{y}_1, \ldots, \mathbf{y}_r, \mathbf{x}_{r+1}, \ldots, \mathbf{x}_n\}$ is a basis for $\mathbf{V}$. Then start the Gram-Schmidt process at the rth step i.e.

$$\mathbf{y}_{r+1} = \frac{\mathbf{y}^*_{r+1}}{\|\mathbf{y}^*_{r+1}\|} \quad \text{where} \quad \mathbf{y}^*_{r+1} = \mathbf{x}_{r+1} - \sum_{i=1}^{r} \langle \mathbf{x}_{r+1}, \mathbf{y}_i \rangle \mathbf{y}_i$$

2.7.3 Orthogonal Complements

The notion of orthogonality of two vectors can also be extended to sets of vectors and subspaces.

Definition: If $\mathbf{W}$ is a set of vectors in $\mathbf{V}$ (but not necessarily a subspace of $\mathbf{V}$), then the set $\mathbf{W}^\perp$ is called the orthogonal complement of $\mathbf{W}$ in $\mathbf{V}$ and is defined as

$$\mathbf{W}^\perp = \{\mathbf{x} : \langle \mathbf{x}, \mathbf{y} \rangle = 0 \; ; \; \mathbf{y} \in \mathbf{W}\}$$

Theorem 2.16: If $\mathbf{W}$ is a subset of $\mathbf{V}$ then $\mathbf{W}^\perp$ is a subspace of $\mathbf{V}$.

Theorem 2.17: If $\mathbf{W}$ is a subspace of $\mathbf{V}$ with $\dim(\mathbf{W}) = r$ and $\dim(\mathbf{V}) = n$, then $\dim(\mathbf{W}^\perp) = n - r$. Also, $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$.

Using the notion of an orthocomplement of a subspace, we say that a vector $\mathbf{x}$ is orthogonal to a subspace $\mathbf{W}$ in $\mathbf{V}$ if $\mathbf{x}$ is orthogonal to every vector in $\mathbf{W}$ or, equivalently, if $\mathbf{x}$ is in $\mathbf{W}^\perp$. This in turn motivates the extension of the notion of the projection of one vector onto another vector to the projection of a vector onto a subspace.

2.7.4 Projections

We have defined projection of one vector in $\mathbf{V}$ onto another vector. The projection of a vector onto a subspace of $\mathbf{V}$ is defined by the following result.

Theorem 2.18: Let $\mathbf{y}$ be a non-zero vector in $\mathbf{V}$ and $\mathbf{W}$ a subspace of $\mathbf{V}$. Then there exist two vectors $\mathbf{y}_1$, $\mathbf{y}_2$ in $\mathbf{V}$ such that

(1) $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$

(2) $\mathbf{y}_1 \in \mathbf{W},\ \mathbf{y}_2 \in \mathbf{W}^{\perp}$

(3) $\mathbf{y}_1$ and $\mathbf{y}_2$ are unique

(4) $d(\mathbf{y}, \mathbf{y}_1) \leq d(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x} \in \mathbf{W}$

We define the projection of $\mathbf{y}$ onto $\mathbf{W}$ as the vector $\mathbf{y}_1$ in the above result. Also, if we can find vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ satisfying properties (1) and (2), then properties (3) and (4) will follow automatically and $\mathbf{y}_1$ is the projection of $\mathbf{y}$ onto $\mathbf{W}$. Property(4) is the least squares property. It states that $\mathbf{y}_1$ is the vector in $\mathbf{W}$ closest to $\mathbf{y}$.

Given an orthonormal basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ for a subspace $\mathbf{W}$ in $\mathbf{V}$, a simple form for the projection of a vector $\mathbf{y}$ onto $\mathbf{W}$ ($\mathbf{y}_1$) may be obtained. Intuition suggests that $\mathbf{y}_1$ should be the sum of the projections of $\mathbf{y}$ onto the orthonormal basis for $\mathbf{W}$

$$\mathbf{y}_1 = \sum_{i=1}^{r} \langle \mathbf{y}, \mathbf{x}_i \rangle \mathbf{x}_i$$

This formula is verified by the following result:

Theorem 2.19 Let $\mathbf{W}$ be a subspace and $\mathbf{y}$ a vector in $\mathbf{V}$. Assume $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is an ONB for $\mathbf{W}$. Then the vector

$$\mathbf{y}_1 = \sum_{i=1}^{r} \langle \mathbf{y}, \mathbf{x}_i \rangle \mathbf{x}_i$$

in $\mathbf{W}$ is the projection of $\mathbf{y}$ onto $\mathbf{W}$.

2.8 Matrix Definitions and Operations

- There is a natural correspondence between inner product spaces of dimension n and the set of all $n \times 1$ matrices.
- If a basis is chosen for the vector space $\mathbf{V}$, then with every vector is associated a unique ordered collection of n numbers corresponding to the coordinates of the vector relative to the chosen basis.
- Arranging these ordered collections of n numbers into $n \times 1$ matrices, we have a one to one correspondence between elements in the vector space $\mathbf{V}$ and $n \times 1$ matrices.

- The notion of linear dependence associated with vectors in $\mathbf{V}$ corresponds to linear dependence of associated $n \times 1$ matrices in R^n.
- Also, multiplication of matrices associated with vectors in $\mathbf{V}$ corresponds to calculation of standard inner products in $\mathbf{V}$.
- Operations on n vectors may be done simultaneously by adjoining the corresponding $n \times 1$ matrices together into an $n \times n$ matrix and manipulating this matrix.
- In the previous sections, vectors in inner product spaces were denoted by bold faced letters (i.e., $\mathbf{x}$).
- The $n \times 1$ matrices corresponding to these vectors will be similarly denoted with the understanding that the vector is a geometric object while the $n \times 1$ matrix is an algebraic object with coordinates relative to a given basis.
- Matrices formed by adjoining several $n \times 1$ matrices will be denoted by bold faced capital letters (i.e. $\mathbf{X}$).

Definition: Let $\mathbf{X}$ be a matrix of dimension $n \times p$. Then the column rank of $\mathbf{X}$ is defined as the number of linearly independent columns of $\mathbf{X}$ and the row rank of $\mathbf{X}$ is defined as the number of linearly independent rows of $\mathbf{X}$.

- The following result establishes that the column and row ranks are equal.
- Thus the rank of $\mathbf{X}$, denoted $\text{rank}(\mathbf{X})$, may unambiguously be defined as the column rank of $\mathbf{X}$.

Theorem 2.20: Let $\mathbf{X}$ be an $n \times p$ matrix. Then the column rank of $\mathbf{X}$ equals the row rank of $\mathbf{X}$.

Given the equality of the row and column ranks of a matrix, a variety of other useful results about the ranks of matrices follow.

For $n \times p$ matrices, $\mathbf{A}$ and $\mathbf{B}$, we have

- $0 \leq \text{rank}\,(\mathbf{A}) \leq \min\,(n, p)$
- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$
- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$
- $\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$
- $\text{rank}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^T) = \text{rank}(\mathbf{A})$
- If $\mathbf{X}$ is an $n \times n$ matrix with $\text{rank}(\mathbf{X}) = n$ and $\mathbf{Y}$ is a $p \times p$ matrix with $\text{rank}(\mathbf{Y}) = p$ then $\text{rank}(\mathbf{XAY}) = \text{rank}(\mathbf{A})$.

Definition: If an $n \times n$ matrix $\mathbf{A}$ has full rank (i.e. $\text{rank}(\mathbf{A}) = n$) then $\mathbf{A}$ is called non-singular.

Theorem 2.21: Let $\mathbf{A}$ be a square matrix of size $n \times n$ and $\text{rank}(\mathbf{A}) = n$. Then there exists a matrix denoted $\mathbf{A}^{-1}$ called the inverse of $\mathbf{A}$ with the property that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ where $\mathbf{I}$ is the $n \times n$ identity matrix given by

$$\mathbf{I} = \{\delta_{ij}\} \quad \text{where} \quad \delta_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

Moreover, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1}$ and $\mathbf{A}^{-1}$ is unique.

There are several particular types of matrices that are useful in many statistical applications. The linear algebra and matrices supplement contains a list of some of these types of matrices and relevant results about them.

Definition: A square $n \times n$ matrix $\mathbf{A}$ is called orthonormal if $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ The transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ associated with the orthonormal matrix $\mathbf{A}$ is called an orthonormal transformation.

- It is easily seen from the definition of an orthonormal matrix, $\mathbf{A}$, that
$$\mathbf{A}^{-1} = \mathbf{A}^T, \mathbf{A}^T\mathbf{A} = \mathbf{I}$$
and that the columns of $\mathbf{A}$ as well as the rows of $\mathbf{A}$ correspond to ONB's for $\mathbf{R}^n$.

- We also note that (standard) inner products are invariant under orthonormal transformations (i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{Ax}, \mathbf{Ay} \rangle$).

- Since an orthogonal transformation preserves the distances between all pairs of points in any configuration, these transformations can be conveniently thought of as rigid rotations about the origin (except for some possible reflections of planes).

Orthonormal transformations arise in relating two ONB's to each other in the following way:

- If $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ are two ONB's for $\mathbf{R}^n$, and $\mathbf{a}, \mathbf{b}$ represent coordinate vectors with respect to these ONB's for a vector $\mathbf{z}$, i.e.

$$\mathbf{z} = \sum_{i=1}^{n} a_i \mathbf{x}_i = \sum_{i=1}^{n} b_i \mathbf{y}_i$$

- Then there exists an orthonormal matrix $\mathbf{P}$ such that $\mathbf{a} = \mathbf{P}\mathbf{b}$.
  - This is seen by letting $\mathbf{A}$ and $\mathbf{B}$ denote the orthonormal matrices with columns being $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ respectively.
  - Then there exist $\mathbf{C}$ and $\mathbf{D}$ so that $\mathbf{CA} = \mathbf{DB}$. Then $\mathbf{P} = \mathbf{C}^{-1}\mathbf{D}$ which is easily shown to be orthonormal since

$$\mathbf{I} = \mathbf{A}\mathbf{A}^T = \mathbf{P}\mathbf{B}\mathbf{B}^T\mathbf{P}^T = \mathbf{P}\mathbf{P}^T$$

Orthonormal matrices also arise in the following decomposition result for symmetric matrices which will be used repeatedly in this course.

Theorem 2.22: (Spectral Decomposition) Any symmetric matrix $\mathbf{A}$ ($n \times n$) can be written as

$$\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{P}^T$$

where $\mathbf{L}$ is a diagonal matrix and $\mathbf{P}$ is an orthonormal matrix.

- The diagonal elements of $\mathbf{L}$, $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, are the the eigenvalues of $\mathbf{A}$ and $\text{rank}(\mathbf{A}) = p$, if and only if there are exactly p non-zero eigenvalues.

- The columns of $\mathbf{P}$ are called (standardized) eigenvectors of $\mathbf{A}$ and these vectors form an ONB for $\mathbf{R}^n$.

- To indicate the correspondence between eigenvectors and values, we say the ith column of $\mathbf{P}$ is the eigenvector associated with the eigenvalue λ_i (or the ith diagonal element of $\mathbf{L}$ is the eigenvalue associated with the ith eigenvector).

- Since the ordering of the columns of $\mathbf{P}$ and elements of $\mathbf{L}$ is arbitrary (as long as the correspondence is maintained between eigenvector and eigenvalues), the usual convention is to arrange the eigenvalues so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Another decomposition result for non-square matrices follows from the Spectral Decomposition result for square symmetric matrices, and is useful in defining a generalized inverse of a matrix.

Theorem 2.23: (Singular Value Decomposition) Let $\mathbf{A}$ $(n \times p)$ be a matrix of rank r. Then $\mathbf{A}$ can be written

$$\mathbf{A} = \mathbf{U}\mathbf{L}\mathbf{V}^T$$

where $\mathbf{U}$ $(n \times r)$ and $\mathbf{V}$ $(p \times r)$ are such that $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$ (i.e. the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and have unit length) and $\mathbf{L}$ $(r \times r)$ is a diagonal matrix with positive elements.

There are other forms of the SVD that are sometimes useful. In one version, $\mathbf{A} = \mathbf{U}\mathbf{L}\mathbf{V}^T$ where $\mathbf{U}$, $\mathbf{V}$ are $n \times n$ orthonormal matrices and $\mathbf{L}$ is an $n \times n$ matrix of the form

$$\mathbf{L} = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{\Lambda}$ is defined above.

A second form of the SVD sometimes called the rank factorization, writes $\mathbf{A} = \mathbf{B}\mathbf{C}$ where $\mathbf{B}$ is an $n \times r$ matrix of rank r and $\mathbf{C}$ is an $r \times p$ matrix of rank r. Expressing $\mathbf{B}$ and $\mathbf{C}$ in terms of the above SVD we see that $\mathbf{B} = \mathbf{U}$ and $\mathbf{C} = \mathbf{L}\mathbf{V}^T$.

Using the singular value decomposition result, we now consider a method of defining an "inverse" for any matrix.

Definition: For a matrix $\mathbf{A}$ $(n \times p)$ of rank r, $\mathbf{A}^-$ is called a g-inverse (generalized inverse) of $\mathbf{A}$ denoted by $\mathbf{A}^-$ if

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

- A generalized inverse always exists although in general it is not unique.
- The existence of a generalized inverse for any matrix $\mathbf{A}$ is easy to verify using Theorem 2.23 to write $\mathbf{A} = \mathbf{U}\mathbf{L}\mathbf{V}^T$ and checking that $\mathbf{V}\mathbf{L}^{-1}\mathbf{U}^T$ defines a g-inverse:

$$\begin{aligned}
\mathbf{A}\mathbf{A}^-\mathbf{A} &= \mathbf{U}\mathbf{L}\mathbf{V}^T(\mathbf{V}\mathbf{L}^{-1}\mathbf{U}^T)\mathbf{U}\mathbf{L}\mathbf{V}^T \\
&= \mathbf{U}\mathbf{L}\mathbf{I}\mathbf{L}^{-1}\mathbf{I}\mathbf{L}\mathbf{V}^T \\
&= \mathbf{U}\mathbf{L}\mathbf{L}^{-1}\mathbf{L}\mathbf{V}` \\
&= \mathbf{U}\mathbf{L}\mathbf{V}^T \\
&= \mathbf{A}
\end{aligned}$$

A g-inverse of a matrix is most useful in solving systems of linear equations.

- Suppose that a vector $\mathbf{y}$ $(n \times 1)$ is in the space spanned by the columns of a matrix $\mathbf{A}$ and we wish to obtain a solution to the equations $\mathbf{Ax} = \mathbf{y}$.
- If the columns of $\mathbf{A}$ are linearly independent, then we know that the columns $\{\mathbf{A}_1^c, \ldots, \mathbf{A}_p^c\}$ form a basis for $\mathbf{Sp}(\mathbf{A}_1^c, \ldots, \mathbf{A}_p^c)$ and the elements of $\mathbf{x}$ are the uniquely defined coordinates of $\mathbf{y} \in \mathbf{Sp}(\mathbf{A}_1^c, \ldots, \mathbf{A}_p^c)$ relative to this basis.
- The solution to the equations will then be unique.
- If $\{\mathbf{A}_1^c, \ldots, \mathbf{A}_p^c\}$ are linearly dependent, there are an infinite number of solutions to these equations.

To show $\mathbf{A}^{-}\mathbf{y}$ is a solution to the equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ for any choice of $\mathbf{A}^{-}$, we need only note that $\mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}$ by the defining property of a g-inverse. This implies that

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^{-}\mathbf{y} = \mathbf{A}(\mathbf{A}^{-}\mathbf{y})$$

(i.e., $\mathbf{A}^{-}\mathbf{y}$ is a candidate for the vector $\mathbf{x}$ which satisfies $\mathbf{Y} = \mathbf{A}\mathbf{x}$).

As noted above, when $\text{rank}(\mathbf{A}) = p$ then $\mathbf{A}^{-}\mathbf{y}$ is the unique solution, if $\text{rank}(\mathbf{A}) < p$ then $\mathbf{A}^{-}\mathbf{y}$ for a given choice of $\mathbf{A}^{-}$ is only one of an infinite number of solutions.

We also note that a general solution to the consistent system of equations $\mathbf{y} = \mathbf{A}\mathbf{x}$ is

$$\{\mathbf{x} : \mathbf{x} = \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z};\ \mathbf{z} \in \mathbf{R}^{n}\}$$

since $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^{-}\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^{-})$ and $(\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z}$ is in $\mathbf{A}^{\perp}$.

Definition: A square matrix $\mathbf{A}$ $(n \times n)$ is said to be idempotent if $\mathbf{AA} = \mathbf{A}$.

Idempotent matrices are intimately related to projections of vectors onto subspaces. The following propositions concerning idempotent matrices are of interest.

(1) If $\mathbf{A}$ is idempotent then

   (a) $\mathbf{I} - \mathbf{A}$ is idempotent

   (b) $\text{rank}\,(\mathbf{A}) = \text{tr}\,(\mathbf{A})$

   (c) $n = \text{rank}\,(\mathbf{A}) + \text{rank}\,(\mathbf{I} - \mathbf{A})$

(2) Suppose $\mathbf{A}_1, \ldots, \mathbf{A}$ are square $(n \times n)$ matrices. If

$$\mathbf{A}_i\mathbf{A}_j = \begin{cases} \mathbf{A}_i & i = j \\ \mathbf{0} & i \neq j \end{cases}$$

then $\sum_{i=1}^{k} \mathbf{A}_i$ is idempotent and $\text{rank}\,(\sum_{i=1}^{k} \mathbf{A}_i) = \sum_{i=1}^{k} \text{rank}\,(\mathbf{A}_i)$.

The connection between idempotent matrices and projections is described in the following proposition:

Theorem 2.24: Let $\mathbf{W}$ be a subspace in $\mathbf{V}$ and $\mathbf{y}$ an arbitrary vector in $\mathbf{V}$.

- Then by Theorem 2.18, $\mathbf{y}$ has the unique decomposition $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$ with $\mathbf{y}_1 \in \mathbf{W}$ and $\mathbf{y}_2 \in \mathbf{W}^{\perp}$.
- Let $\mathbf{Y}$, $\mathbf{Y}_1$ denote $n \times 1$ matrices of coordinates of $\mathbf{y}$, $\mathbf{y}_1$ respectively related to any basis of $\mathbf{V}$.
- The mapping (function) $\mathbf{P}$ that takes $\mathbf{Y}$ into $\mathbf{Y}_1$ (called the projection of $\mathbf{Y}$ onto $\mathbf{W}$) is called the projection operator of $\mathbf{Y}$ onto $\mathbf{W}$.

$\mathbf{P}$ has the following properties:

(1) $\mathbf{P}$ is a linear transformation and so may be represented by a (unique) matrix

(2) $\mathbf{P}$ is an idempotent matrix

(3) $\mathbf{I} - \mathbf{P}$ is the projection operator on $\mathbf{W}^{\perp}$

(4) A matrix $\mathbf{P}$ is a projection operator (into the space spanned by its column vectors) if and only if $\mathbf{P}$ is idempotent and symmetric.

(5) Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of vectors in $\mathbf{V}$ not necessarily linearly independent. Then the projection operator onto $\mathbf{Sp}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ has an explicit form given by

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$$

which is unique for any choice of g-inverse, where $\mathbf{X}$ is the matrix with column vectors $(\mathbf{x}_1, \ldots, \mathbf{x})$.

Proof of (5)

- Since

$$\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^-]^T\mathbf{X}^T\mathbf{X}$$

  we have $[(\mathbf{X}^T\mathbf{X})^-]^T$ as a g- inverse of $\mathbf{X}^T\mathbf{X}$.

  - This may then be used to show $\mathbf{D}\mathbf{D}^T = \mathbf{0}$ where

$$\mathbf{D} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} - \mathbf{X}$$

  - and hence that

$$\mathbf{D} = \mathbf{0} \text{ i.e. } \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}$$

- This in turn may be used to show that $\mathbf{E}\mathbf{E}^T = \mathbf{0}$ where

$$\mathbf{E} = \mathbf{X}(\mathbf{X}^T\mathbf{X})_1^-\mathbf{X}^T - \mathbf{X}(\mathbf{X}^T\mathbf{X})_2^-\mathbf{X}^T$$

  and $(\mathbf{X}^T\mathbf{X})_1^-$, $(\mathbf{X}^T\mathbf{X})_2^-$ are any two generalized inverses of $\mathbf{X}^T\mathbf{X}$ which implies $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is unique.

- Thus it is also symmetric.

- Idempotency follows from the fact that $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-$ is a g-inverse of $\mathbf{X}^T\mathbf{X}$ and so

$$\begin{aligned}\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T &= \mathbf{X}\left[(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\right]\mathbf{X}^T \\ &= \left[\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\right]\left[\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\right]\end{aligned}$$

- Thus $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is unique, symmetric and idempotent and hence is a projection operator.

Definition: Let $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be a set of n vectors in $\mathbf{R}^n$. A mapping from the n-fold Cartesian product of $\mathbf{R}^n$ denoted by $\det(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ or $\det(\mathbf{X})$ (where $\mathbf{X}$ is the $n \times n$ matrix with $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as its columns) with the following four properties exists and is unique.

(1) $\det(\mathbf{x}_1, \ldots, \mathbf{x}_n) = 0$ if $\mathbf{x}_i = \mathbf{x}_j$ for any $i \neq j$.

(2) $\det(\mathbf{x}_1, \ldots, \alpha_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n) = \alpha \det(\mathbf{x}_1, \ldots, \mathbf{x}_n)$

(3)

$$\begin{aligned}\det(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i + \mathbf{y}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n) &= \det(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n) \\ &\quad + \det(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{y}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)\end{aligned}$$

(4) $\det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$ where $\mathbf{e}_i$ is the ith standard unit vector for $\mathbf{R}^n$.

Algorithms for computing the determinants of square matrices will not be presented here. The following results about determinants are useful:

Let $\mathbf{X}$ be an $n \times n$ matrix with column vectors $\mathbf{X}_1^c, \ldots, \mathbf{X}_n^c$. Then

- $\det(\mathbf{X}) = 0$ if and only if $\mathbf{X}_1^c, \ldots, \mathbf{X}_n^c$ are linearly dependent.
- the parallelepiped determined by $\mathbf{X}_1^c, \ldots, \mathbf{X}_n^c$,

$$\left\{ \sum_{i=1}^{n} \alpha_i \mathbf{X}_i^c : \; 0 \leq \alpha_i \leq 1 \right\}$$

  has volume equal to $\det(\mathbf{X})$
- $\det(\mathbf{X}) = \prod_{i=1}^{n} \lambda_i$ where the λ_i's are the eigenvalues of $\mathbf{X}$. The

An excellent reference on determinants is Rao, Advanced Statistical Methods in Biometric Research, John Wiley.