1 Review of Linear Algebra

1.1 Vector Space

A vector space over the field $\mathbb{R}$ is a set $V$ equipped with following two operations, each satisfying four axioms.

1.1.1 Vector addition

Vector addition is a mapping $+: V \times V \rightarrow V$ defined by $+(v, w) = v + w$ for any two elements $v, w \in V$, that satisfies the following four axioms.

1. Associativity of addition: $u + (v + w) = (u + v) + w$; for all $u, v, w \in V$

2. Commutativity of addition: $u + v = v + u$; for all $u, v \in V$

3. Existence of Identity: There exists a zero vector ($0 \in V$) s.t, $u + 0 = u$; for all $u \in V$

4. Existence of Inverse: For every $u \in V$, there exists an element $-u \in V$; s.t, $u + (-u) = 0$

1.1.2 Scalar Multiplication

Scalar multiplication is a mapping $\cdot: \mathbb{R} \times V \rightarrow V$ defined by $\cdot(\alpha, v) = \alpha v \in V$, that satisfies the following four axioms.

1. Compatibility with the field: $a(bu) = (ab)u$; for all $a, b \in \mathbb{R}$ and $u \in V$

2. Existence of Identity: For multiplicative identity element $1 \in \mathbb{R}$, $1u = u$; for all $u \in V$
3. Distributivity over vector addition: \( \alpha (vu) = \alpha u + \alpha v \); for all \( \alpha \in \mathbb{R} \) and \( u, v \in V \)

4. Distributivity over field addition: \( (\alpha + \beta)u = \alpha u + \beta u \); for all \( \alpha, \beta \in \mathbb{R} \) and \( u \in V \)

**Example 1.1 (Vector space).** Following are some common examples of vector spaces.

1. Space of all real numbers \( \mathbb{R} \).
2. Euclidean space of \( N \)-dimensions, denoted by \( \mathbb{R}^N \).
3. Space of continuous functions over a compact subset \([a, b]\) denoted by \( C([a, b]) \).

### 1.2 Inner Product Space

An *inner product space* is a vector space equipped with an inner product denoted by \( \langle \cdot, \cdot \rangle : V \times V \to \mathbb{R} \) that satisfies the following axioms.

1. **Symmetry:** \( \langle x, y \rangle = \langle y, x \rangle \)
2. **Linearity:** \( \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \)
3. **Definiteness:** \( \langle x, x \rangle \geq 0; \langle x, x \rangle = 0 \text{ iff } x = 0 \)

**Example 1.2 (inner product spaces).** Following are some common examples of inner product spaces.

1. For the vector space \( V = \mathbb{R}^N \), we can define the inner product between two \( N \)-dimensional vectors as
   \[
   \langle x, y \rangle = \langle [x_1, \ldots, x_N]^T, [y_1, \ldots, y_N]^T \rangle = x^T y = \sum_{i=1}^{N} x_i y_i.
   \]

2. For vector space \( V = C(\mathbb{R}^N) \), we can define the inner product of two continuous functions over \( \mathbb{R}^N \) as
   \[
   \langle f, g \rangle = \int_{\mathbb{R}^N} (f, g)(t) dt.
   \]

3. For the vector space of random variables, we can define the inner product of two random variables as
   \[
   \langle X, Y \rangle = \mathbb{E}(XY).
   \]
1.3 Norms

Norm is a mapping $\|\cdot\| : V \to \mathbb{R}_+$ that satisfy the following axioms.

1. **Definiteness**: $\|v\| = 0$ iff $v = 0$

2. **Homogeneity**: $\|\alpha v\| = |\alpha| \|v\|$

3. **Triangle inequality**: $\|v + w\| \leq \|v\| + \|w\|$

**Example 1.3 (Norms).** Following are examples of commonly defined norms on some example vector spaces.

1. $V = \mathbb{R}; \|X\| = |X|$

2. $V = \mathbb{R}^N; \|X\|_p = \left(\sum_{i=1}^{N} |X_i|^p\right)^\frac{1}{p}$

3. $V = \mathbb{R}^N; \|X\|_2 = \left(\sum_{i=1}^{N} |X_i|^2\right)^\frac{1}{2}$

**Proposition 1.4 (Holder’s Inequality).** Let $p,q \geq 1$ be conjugate, i.e. $\frac{1}{p} + \frac{1}{q} = 1$ Then,

$$|\langle x,y \rangle| \leq \|x\|_p \|y\|_q \text{ for all } x,y \in \mathbb{R}^N.$$ 

**Proof.** For any positive $a,b \in \mathbb{R}$ and conjugate pair $p,q \geq 1$ such that $1/p + 1/q = 1$, we have from the concavity of log

$$\ln\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right) \geq \frac{1}{p} \ln a^p + \frac{1}{q} \ln b^q = \ln ab.$$ 

Since $\ln(\cdot)$ is an increasing function, the above inequality implies the Young’s inequality

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab.$$ 

The Holder’s inequality is trivially true if $x = 0$ or $y = 0$. Hence, we assume that $\|x\| \|y\| > 0$, and let $a = \frac{|x|}{\|x\|_p}$ and $b = \frac{|y|}{\|y\|_p}$. From Young’s inequality, we have

$$\frac{|x|}{p \|x\|_p} + \frac{|y|}{q \|y\|_q} \geq \frac{|x||y|}{\|x\|_p \|y\|_q}, \text{ for all } i \in [N].$$ 

Since $|\langle x,y \rangle| \leq \sum_{i=1}^{N} |x_i||y_i|$, we get the result by summing both sides over $i \in [N]$ in the above inequality.
2 Review of Convex Optimization

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a function over $N$-dimensional reals. Then, we can write its Taylor series expansion around the neighborhood of $x \in \mathbb{R}^N$ as

$$f(y) = f(x) + \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} (y_i - x_i) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 f}{\partial x_i \partial x_j} (y_i - x_i) (y_j - x_j) + o(\|y - x\|_2^2).$$

We can define the gradient vector $\nabla f = \left[ \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N} \right]^T$, and the Hessian $\nabla^2 f \in \mathbb{R}^{N \times N}$ such that $[\nabla^2 f]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$, to observe

$$f(y) = f(x) + \nabla f^T (y - x) + (y - x)^T \nabla^2 f (y - x) + o(\|y - x\|_2^2).$$

2.1 Convex Function

Let $X \subseteq \mathbb{R}^N$. For a function $f : X \to \mathbb{R}$, we define its epigraph as

$$Epi(f) \triangleq \{(x, y) \in \mathbb{R}^N \times \mathbb{R} : y \geq f(x)\}.$$

**Definition 2.1.** A function $f : \mathbb{R}^N \to \mathbb{R}$ is convex if it’s dom$(f)$ is convex and Epi$(f)$ is convex

**Note 2.2.** For a convex function $f(.)$; $f(\alpha x + \bar{\alpha} y) \leq \alpha f(x) + \bar{\alpha} f(y)$ where $\alpha + \bar{\alpha} = 1$.

- If $f$ is differentiable then $f$ is convex iff
  1. dom$(f)$ is convex
  2. $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$; for all $x, y \in$ dom$(f)$

**Proof:** $f(y) - f(x) = \langle \nabla f(x), y - x \rangle + \frac{1}{2} (y - x)^T \nabla^2 f(x) (y - x) \geq \langle \nabla f(x), y - x \rangle$.

- If $f$ is twice differentiable then $f$ is convex iff dom$(f)$ is convex and it’s Hessian is positive semi definite : $\nabla^2 f(x) \succeq 0$; for all $x \in$ dom$(f)$

**Example 2.3.** Convex Function

1. Linear Function: $f(x) = \langle w, x \rangle$; where $f : \mathbb{R}^N \to \mathbb{R}$
2. Quadratic Function: $f(x) = x^T Ax$; where $A$ is positive semi definite
3. Abs Maximum $f(x) = \max_{i \in N} |X|_{i \in N} = \|X\|_\infty$
Lemma 2.4. Composition of Functions
Let, \( h(.) : \mathbb{R} \to \mathbb{R}; g(.) : \mathbb{R}^N \to \mathbb{R} \) and \( f : \mathbb{R}^N \to \mathbb{R} \); for all \( x \in \mathbb{R}^N \) where \( f(x) \) is defined by \( f(x) = h(g(x)) \), then following inequalities are valid

1. If \( h \) is a convex and non decreasing and \( g \) is convex, \( \implies f(.) \) is convex  
   **Proof:** As \( g(.) \) is convex : \( g(\alpha x + \bar{\alpha} y) \leq \alpha g(x) + \bar{\alpha} g(y) \)  
   Now, \( h(g(\alpha x + \bar{\alpha} y)) \leq h(\alpha g(x) + \bar{\alpha} g(y)) \leq \alpha h(g(x)) + \bar{\alpha} h(g(y)) \)(Proved.)

2. If \( h \) is a convex and non increasing and \( g \) is concave, \( \implies f(.) \) is convex

3. If \( h \) is a concave and non decreasing and \( g \) is concave, \( \implies f(.) \) is concave

4. If \( h \) is a concave and non increasing and \( g \) is convex, \( \implies f(.) \) is concave

Theorem 2.5. Jensen’s Inequality
Let \( X \in C \subset \mathbb{R}^N \) be a r.v with finite mean and \( f : C \to \mathbb{R} \) is convex,  
Then \( E[X] \in C, E[f(X)] \leq \infty \) and \( f(E[X]) \leq E[f(X)] \)

**Proof:** \( f(\sum_{i=1}^{m} \alpha_i x_i) \leq \sum_{i=1}^{m} \alpha_i f(x_i) \); where \( \alpha_i \)s could be interpreted as probabilities as \( \alpha_i \geq 0 \) and \( \sum_{i=1}^{m} \alpha_i = 1 \)

2.2 Constrained Optimization

Let \( f : \mathbb{R}^N \to \mathbb{R} \) and \( g_i : \mathbb{R}^N \to \mathbb{R}, i \in [m] \)  
**Principle Optimization Problem:** \( \min f(x) \) s.t. \( g_i(x) \leq 0 \); for all \( i \in [m] \)

**Note 2.6.** Let \( p^* \) be the optimum value for the above problem.

**Definition 2.7. Lagrangian**
If \( x \in \mathbb{R}^N \) and \( \alpha \in \mathbb{R}_+^M \), then Lagrangian \( \mathcal{L}(x, \alpha) : \mathbb{R}^N \times \mathbb{R}_+^M \to \mathbb{R} \) associated with the principal problem is defined as, \( \mathcal{L}(x, \alpha) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) \); The variables \( \alpha \in \mathbb{R}_+^M \) are called Lagrange or Dual Variables.

**Definition 2.8. Dual Function**
Dual function associated with the Principal Optimization Problem is defined as \( F : \mathbb{R}_+^M \to \mathbb{R} \) defined as \( F(\alpha) = \inf \mathcal{L}(x, \alpha) \) where \( x \in \mathbb{R}^N \)

**Remark 2.9.** Important Properties of Dual Function

1. \( F \) is concave in \( \alpha \)
2. \( F(\alpha) \leq \mathcal{L}(x, \alpha) \leq f(x) \)

3. \( F(\alpha) \leq \inf_{x \in \mathbb{R}^N} f(x) = p^* \)

4. \( F(\alpha) \leq p^* \) such that \( g_i(x) \leq 0 \)

### 2.2.1 Dual Problem:

Dual Problem associated with Principal Optimization Problem is as follows

\[ \text{Max } F(\alpha); \text{ such that } \alpha \in \mathbb{R}^M \]

**Note 2.10.** Let \( d^* \) be the optimal value of this dual problem.

**Remark 2.11.**

- Dual Function
  
  1. Dual problem is always convex.
  
  2. \( d^* \leq p^* \)
  
  3. \( (p^* - d^*) \) is called duality gap. When \( d^* = p^* \), it is known as strong duality. It holds for convex optimization problems where constraints are qualifying.

**Definition 2.12. Strong Constraint Qualification:**

Assume that \( \text{int}(\mathcal{X}) \neq \phi \), then the strong constraint qualification or **Slater’s Condition** is defined as, there exists \( \bar{x} \in \text{int}(\mathcal{X}) \), such that \( g(\bar{x}) < 0 \)

**Definition 2.13. Weak Constraint Qualification:** Assume that \( \text{int}(\mathcal{X}) \neq \phi \), then the strong constraint qualification or **weak Slater’s Condition** is defined as there exists \( \bar{x} \in \text{int}(\mathcal{X}) : \) for all \( i \in [1, m], (g_i(\bar{x}) < 0) \lor (g_i(\bar{x}) = 0 \land g_i \text{ affine}) \)

**Theorem 2.14. Saddle Point: Sufficient Condition**

Let \( P \) be a constrained optimum problem over \( \mathcal{X} = \mathbb{R}^N \) \( \text{If } (x^*, \alpha^*) \text{ is a saddle point of the associated Lagrangian, i.e. for all } x \in \mathbb{R}^N, \text{ for all } \alpha \geq 0, \mathcal{L}(x^*, \alpha^*) \leq \mathcal{L}(x, \alpha) \leq \mathcal{L}(x, \alpha^*) \leq \mathcal{L}(x^*, \alpha^*) \text{ Then, } (x^*, \alpha^*) \text{ is a saddle point of } P. \)

**Theorem 2.15. Saddle point-Necessary Condition**

- Assume that \( f \) and \( g_i, i \in [1, m] \text{ are convex functions and Slater’s condition holds, then if } x \text{ is a solution of the constrained optimization problem, then there exists } \alpha \geq 0 \text{ s.t } (x, \alpha) \text{ is a saddle point of the Lagrangian.} \)

- Assume that \( f \) and \( g_i, i \in [1, m] \text{ are convex differentiable functions and weak } \text{Slater’s condition holds, then if } x \text{ is a solution of the constrained optimization problem, then there exists } \alpha \geq 0 \text{ s.t } (x, \alpha) \text{ is a saddle point of the Lagrangian.} \)
Theorem 2.16. Karush-Kuhn-Tucker’s Theorem
Let \( f, g_i : \mathcal{X} \rightarrow \mathbb{R} \), for all \( i \in [1, m] \) are convex and differentiable function and that the constrains are qualified. Then \( \bar{x} \) is a solution of the constrained problem iff there exists \( \bar{\alpha} \geq 0 \) s.t.

\[
\begin{align*}
\nabla_x \mathcal{L}(\bar{x}, \bar{\alpha}) &= \nabla_x f(\bar{x}) + \langle \bar{\alpha}, \nabla_x g(\bar{x}) \rangle = 0 \\
\nabla_{\alpha} \mathcal{L}(\bar{x}, \bar{\alpha}) &= g(\bar{x}) \leq 0 \\
\langle \bar{\alpha}, g(\bar{x}) \rangle &= 0
\end{align*}
\]