Lecture-02: Review of Linear Algebra and Convex Optimization

1 Review of Linear Algebra

1.1 Vector Space

Definition 1.1 (Vector addition). A set V is set to be equipped with vector addition mapping $+ : V \times V \to V$ defined by $+(v, w) = v + w$ for any two elements $v, w \in V$, if this mapping satisfies the following four axioms.

1. **Associativity:** For all vectors $u, v, w \in V$, we have $u + (v + w) = (u + v) + w$.
2. **Commutativity:** For all vectors $u, v \in V$, we have $u + v = v + u$.
3. **Additive identity:** There exists a zero vector $0 \in V$, such that $u + 0 = u$ for all $u \in V$.
4. **Additive inverse:** For each vector $u \in V$, there exists an additive inverse $-u \in V$ such that $u + (-u) = 0$.

Definition 1.2 (Scalar multiplication). A set V equipped with vector addition $+ : V \times V \to V$ is also equipped with field scalar multiplication mapping $\cdot : \mathbb{F} \times V \to V$ defined by $\cdot(\alpha, v) = \alpha v \in V$, if this mapping satisfies the following four axioms.

1. **Field compatibility:** For all scalars $\alpha, \beta \in \mathbb{F}$ and vector $u \in V$, we have $\alpha(\beta u) = (\alpha\beta)u$.
2. **Multiplicative identity:** There exists a multiplicative identity element $1 \in \mathbb{F}$, such that $1u = u$ for all $u \in V$.
3. **Distributivity over vector addition:** For each scalar $\alpha \in \mathbb{F}$ and vectors $u, v \in V$, we have $\alpha(v + u) = \alpha u + \alpha v$.
4. **Distributivity over field addition:** For all scalars $\alpha, \beta \in \mathbb{F}$ and vector $u \in V$, we have $(\alpha + \beta)u = \alpha u + \beta u$.

Definition 1.3. A vector space over the field $\mathbb{F}$ is a set V equipped with vector addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{F} \times V \to V$.

Definition 1.4. A set of vectors $W \subseteq V$ are called linearly independent, if for any nonzero vector $\alpha \in \mathbb{F}^W$ with finite $\sum_w \alpha_w$, we have $\sum_{w \in W} \alpha_w w \neq 0 \in V$.

Definition 1.5. The span of a set of vectors $W \subseteq V$ is defined by

$$\mathrm{Span}(W) \triangleq \left\{ \sum_{w \in W} \alpha_w w : \alpha \in \mathbb{R}^W, \sum_{w \in W} \alpha_w \text{ finite} \right\}.$$

Definition 1.6. A basis of any vector space V, is a spanning set of linearly independent vectors.

Theorem 1.7. *All bases of a vector space V have identical cardinality, and defined to be its dimension.*

Example 1.8 (Vector space). Following are some common examples of vector spaces.

1. Euclidean space of N-dimensions, denoted by $\mathbb{R}^N$.
2. Space of continuous functions over a compact subset $[a, b]$ denoted by $C([a, b])$.
3. Space of random variables defined over probability space $(\Omega, \mathcal{F}, P)$ with finite pth moment denoted by L^p.

1.2 Inner Product Space

A *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:** For all vectors $x,y \in V$, we have $\langle x,y \rangle = \langle y,x \rangle$.
2. **Linearity:** For all scalars $\alpha,\beta \in \mathbb{F}$ and vectors $x,y,z \in V$, we have $\langle \alpha x + \beta y, z \rangle = \alpha \langle x,z \rangle + \beta \langle y,z \rangle$.
3. **Definiteness:** For all vectors $x \in V$, we have $\langle x,x \rangle \geqslant 0$, and $\langle x,x \rangle = 0$ iff $x = 0$.

Example 1.9 (inner product spaces). Following are some common examples of inner product spaces.

1. For the vector space $V = \mathbb{R}^N$ of N-dimensional vectors, the inner product is defined as $\langle x,y \rangle \triangleq x^T y = \sum_i^N x_i y_i$.
2. For vector space $V = C(\mathbb{R}^N)$ of continuous functions, the inner product is defined as $\langle f,g \rangle \triangleq \int_{\mathbb{R}^N} (f,g)(t)dt$.
3. For the vector space of random variables, the inner product $\langle \cdot,\cdot \rangle : L^p \times L^q \to \mathbb{R}$ is defined as $\langle X,Y \rangle \triangleq \mathbb{E}XY$ for conjugate pairs $p,q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

1.3 Norms

Definition 1.10. Norm is a mapping $\|\cdot\| : V \to \mathbb{R}_+$ that satisfy the following axioms.

1. **Definiteness:** For all vectors $v \in V$, we have $\|v\| = 0$ iff $v = 0$.
2. **Homogeneity:** For all scalars $\alpha \in \mathbb{R}$ and vectors $v \in V$, we have $\|\alpha v\| = |\alpha| \|v\|$.
3. **Triangle inequality:** For all vectors $v,w \in V$, we have $\|v + w\| \leqslant \|v\| + \|w\|$.

Example 1.11 (Norms). Let $p \geqslant 1$, then following are common examples of p-norms.

1. For a vector space $V = \mathbb{R}^N$, we can define the p-norm as $\|x\|_p \triangleq \left(\sum_{i=1}^N |x_i|^p \right)^{\frac{1}{p}}$ for all $x \in \mathbb{R}^N$.
2. For vector space $V = C(\mathbb{R}^N)$, the p-norm is defined as $\|f\|_p \triangleq \left(\int_{\mathbb{R}^N} |f|^p (t)dt \right)^{\frac{1}{p}}$ for all $f \in C(\mathbb{R}^N)$.
3. For vector space of random variables, the p-norm is defined as $\|X\|_p \triangleq \left(\mathbb{E}|X|^p \right)^{\frac{1}{p}}$ for all $X \in L^p$.

Example 1.12 (Special norms). Let $p \in \{1,2,\infty\}$, then following are common examples of p-norms.

1. Consider the vector space $V = \mathbb{R}^N$ and $x \in V$. For $p = 1$, we have $\|x\|_1 = \sum_{i=1}^N |x_i|$. For $p = 2$, the norm is Euclidean norm such that $\|x\|_2^2 = \langle x,x \rangle$. For $p = \infty$, we have $\|x\|_\infty = \max_i |x_i|$.
2. Consider the vector space $V = C(\mathbb{R}^N)$ and $f \in V$. For $p = 1$, we have $\|f\|_1 = \int_{t \in \mathbb{R}^N} |f|(t)dt$. For $p = 2$, the norm is Euclidean norm such that $\|f\|_2^2 = \langle f,f \rangle = \int_{t \in \mathbb{R}^N} |f|^2 (t)dt$. For $p = \infty$, we have $\|f\|_\infty = \sup_t |f|(t)$.
3. Consider the vector space V of random variables and $X \in V$. For $p = 1$, we have $\|X\|_1 = \mathbb{E}|X|$. For $p = 2$, the norm is Euclidean norm such that $\|X\|_2^2 = \langle X,X \rangle = \mathbb{E}X^2$. For $p = \infty$, we have $\|X\|_\infty = \sup_\omega |X|(\omega)$.

Proposition 1.13 (Holder's Inequality). *Let p,q $\geqslant 1$ be a conjugate pair, i.e.* $\frac{1}{p} + \frac{1}{q} = 1$. *Then,*

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

Proof. The Holder's inequality is trivially true if $x = 0$ or $y = 0$. Hence, we assume that $\|x\|_p \|y\|_q > 0$, and let $a \triangleq \frac{|x_i|}{\|x\|_p}$ and $b \triangleq \frac{|y|_i}{\|y\|_q}$. We will use the Young's inequality $\frac{1}{p}a^p + \frac{1}{q}b^q \geqslant ab$ for all $a,b > 0$, that implies that

$$\frac{|x_i|^p}{p\|x\|_p^p} + \frac{|y_i|^q}{q\|y\|_q^q} \geqslant \frac{|x|_i |y|_i}{\|x\|_p \|y\|_q}, \text{ for all } i \in [N].$$

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

2 Review of Convex Optimization

Let $\mathcal{X} \subseteq \mathbb{R}^N$ for $N \geqslant 1$ and $f : \mathcal{X} \to \mathbb{R}$ be a smooth function.

Definition 2.1 (Gradient). The gradient of function f at point $x \in \mathcal{X}$ is defined as the column vector $\nabla f(x) \in \mathbb{R}^N$, where the entry $i \in [N]$ is defined as $(\nabla f(x))_i \triangleq \frac{\partial f}{\partial x_i}(x)$.

Definition 2.2 (Hessian). The Hessian of function f at point $x \in \mathcal{X}$ is denoted by the matrix $\nabla^2 f(x) \in \mathbb{R}^{N \times N}$, where the entry $(i,j) \in [N] \times [N]$ is defined as $\nabla^2 f_{i,j}(x) \triangleq \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

Remark 1. Let $f : \mathbb{R}^N \to \mathbb{R}$ be a smooth function over N-dimensional reals. Then, we can write its Taylor series expansion around the neighborhood of $x \in \mathbb{R}^N$, in terms of the gradient vector $\nabla f(x) \in \mathbb{R}^N$ and the Hessian matrix $\nabla^2 f(x) \in \mathbb{R}^{N \times N}$, as

$$f(y) = f(x) + \langle \nabla f(x), (y-x) \rangle + \frac{1}{2} \left\langle (y-x), \nabla^2 f(x)(y-x) \right\rangle + o(\|y-x\|_2^2). \tag{1}$$

Definition 2.3 (Stationary Point). A point $x \in \mathcal{X}$ is called a stationary point of $f : \mathcal{X} \to \mathbb{R}$, if f attains a local extremum at x.

Remark 2. If $f : \mathcal{X} \to \mathbb{R}$ is smooth, then $\nabla f(x) = 0$ at a stationary point $x \in \mathcal{X}$.

2.1 Convexity

Definition 2.4 (Convex Set). A set $\mathcal{X}$ is called convex if for all $x, y \in \mathcal{X}$ and $\alpha \in [0,1]$, the convex combination $\alpha x + \bar{\alpha} y \in \mathcal{X}$ where $\bar{\alpha} \triangleq (1-\alpha)$.

Definition 2.5 (Convex Hull). A convex hull of a set A is the smallest convex set including A, i.e. $\text{conv}(A) \triangleq \{\sum_{x \in A} \alpha_x x : 0 \leqslant \alpha_x \leqslant 1, \sum_{x \in A} \alpha_x = 1\}$.

Definition 2.6. Let $\mathcal{X} \subseteq \mathbb{R}^N$. For a function $f : \mathcal{X} \to \mathbb{R}$, we define its epigraph as

$$\text{Epi}(f) \triangleq \{(x,y) \in \mathcal{X} \times \mathbb{R} : y \geqslant f(x)\}.$$

Definition 2.7. A function $f : \mathcal{X} \to \mathbb{R}$ is convex if the associated domain $\mathcal{X}$ and epigraph $\text{Epi}(f)$ are convex sets.

Theorem 2.8. *Let $\mathcal{X} \subset \mathbb{R}^N$ be a convex set. Then the following are equivalent statements.*

1. *$f : \mathcal{X} \to \mathbb{R}$ is a convex function.*
2. *For all $\alpha \in [0,1]$, we have $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y)$.*
3. *For differentiable f, we have $f(y) - f(x) \geqslant \langle \nabla f(x), y-x \rangle$ for all $x, y \in \mathcal{X}$.*
4. *For twice differentiable f, we have $\nabla^2 f \succeq 0$, i.e. $\nabla^2 f$ is a positive semi-definite matrix.*

Proof. For convex set $\mathcal{X} \subseteq \mathbb{R}^N$ and a function $f : \mathcal{X} \to \mathbb{R}$, we will show that statement 1 implies statement 2, which implies statement 3, which implies statement 4, which implies statement 1.

$1 \implies 2$: Let $(x, f(x)), (y, f(y)) \in \text{Epi}(f)$ for $x, y \in \mathcal{X}$. Let $\alpha \in [0,1]$, then from the convexity of $\mathcal{X}$, we have $\alpha x + \bar{\alpha} y \in \mathcal{X}$. Further from the convexity of $\text{Epi}(f)$, we have $(\alpha x + \bar{\alpha} y, \alpha f(x) + \bar{\alpha} f(y)) \in \text{Epi}(f)$. That is, $\alpha f(x) + \bar{\alpha} f(y) \geqslant f(\alpha x + \bar{\alpha} y)$.

$2 \implies 3$: Recall that $\alpha x + \bar{\alpha} y = x + \bar{\alpha}(y-x)$. From statement 2, we have $f(y) - f(x) \geqslant \frac{f(\alpha x + \bar{\alpha} y) - f(x)}{\bar{\alpha}}$. Taking $\bar{\alpha} \to 0$, we observe that the right hand side is equal to $\langle \nabla f(x), y-x \rangle$.

$3 \implies 4$: From (1) and statement 3, it follows that for any $x, y \in \mathcal{X}$ $f(y) - f(x) - \langle \nabla f(x), y-x \rangle = \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x) + o(\|y-x\|_2^2) \geqslant 0$.

4 $\Longrightarrow$ 1: Let $\alpha \in [0,1]$. Then, it suffices to show that $\alpha f(x_1) + \bar{\alpha} f(x_2) \geqslant f(\alpha x_1 + \bar{\alpha} x_2)$. From the Taylor expansion of f in the neighborhood of x_2, we get

$$\alpha(f(x_1) - f(x_2)) = \alpha \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\alpha}{2} \Big\langle (x_1 - x_2), \nabla^2 f(x_2)(x_1 - x_2) \Big\rangle + o(\|x_1 - x_2\|_2^2).$$

Similarly, we write the Taylor expansion of f in the neighborhood of x_2, to get

$$f(\alpha x_1 + \bar{\alpha} x_2) - f(x_2) = \alpha \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\alpha^2}{2} \Big\langle (x_1 - x_2), \nabla^2 f(x_2)(x_1 - x_2) \Big\rangle + o(\|x_1 - x_2\|_2^2).$$

Taking the difference, we get $\alpha(f(x_1) - f(x_2)) \geqslant f(\alpha x_1 + \bar{\alpha} x_2) - f(x_2)$.

□

Example 2.9 (Convex Function). Following functions $f : \mathbb{R}^N \to \mathbb{R}$ are convex.

1. Linear Function: $f(x) = \langle w, x \rangle$ for each $w \in \mathbb{R}^N$.
2. Quadratic Function: $f(x) = x^T A x$ for a positive semi definite matrix $A \in \mathbb{R}^{N \times N}$.
3. Abs Maximum: $f(x) = \max\{|x_i| : i \in [N]\} = \|x\|_\infty$.

Lemma 2.10 (Composition of functions). *We define a composition function $f = h \circ g$ for functions $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^N \to \mathbb{R}$, by defining $f(x) \triangleq h(g(x))$ for all $x \in \mathbb{R}^N$. Then, the following statements are true.*

1. *If h is convex and nondecreasing and g is convex, then f is convex.*
2. *If h is convex and nonincreasing and g is concave, then f is convex.*
3. *If h is concave and nondecreasing and g is concave, then f is concave.*
4. *If h is concave and nonincreasing and g is convex, then f is concave.*

Proof. We will use the property that a function f is convex iff $\mathrm{dom}(f)$ is convex and $f(\alpha x + \bar{\alpha} y) \leqslant \alpha f(x) + \bar{\alpha} f(y)$ for all $\alpha \in [0,1]$. Recall that $\mathbb{R}^N$ is convex for all $N \geqslant 1$. We will only show the first statement, and rest follow the same steps. Let $x, y \in \mathbb{R}^N$ and $\alpha \in [0,1]$. From the convexity of g, we get $g(\alpha x + \bar{\alpha} y) \leqslant \alpha g(x) + \bar{\alpha} g(y)$. From the nondecreasing property of h, we get $h(g(\alpha x + \bar{\alpha} y)) \leqslant h(\alpha g(x) + \bar{\alpha} g(y))$. From the convexity of h, we get $h(\alpha g(x) + \bar{\alpha} g(y)) \leqslant \alpha h(g(x)) + \bar{\alpha} h(g(y))$. □

Theorem 2.11 (Jensen's Inequality). *Let $X : \Omega \to \mathcal{X} \subseteq \mathbb{R}^N$ be a random vector with finite marginal means, and $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then the mean $\mathbb{E}[X] \in \mathcal{X}$, the mean $\mathbb{E}[f(X)]$ is fnite, and $f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)]$.*

Proof. We will show this for simple random vector $X : \Omega \to \{x_1, \dots, x_m\} \subseteq \mathcal{X}$, such that $\alpha_i \triangleq P\{X = x_i\}$ for all $i \in [m]$. Then, the mean $\mathbb{E}X = \sum_{i=1}^m \alpha_i x_i \in \mathcal{X}$ from the convexity of $\mathcal{X}$, and $\mathbb{E}f(X) = \sum_{i=1}^m \alpha_i f(x_i)$ is finite. Further, from the convexity of f, we get $f\Big(\sum_{i=1}^m \alpha_i x_i\Big) \leqslant \sum_{i=1}^m \alpha_i f(x_i)$. □

Corollary 2.12 (Young's inequality). *Let $p, q \geqslant 1$ be a conjugate pair such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, $ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}$.*

Proof. Take a random variable $X : \Omega \to \{a^p, b^q\}$ with probability mass function $P_X(a^p) = \frac{1}{p}$ and $P_X(b^q) = \frac{1}{q}$. Then, from the concavity of log

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

Since $\ln(\cdot)$ is an increasing function, the above inequality implies the result. □

2.2 Constrained Optimization

Problem 1 (Primal problem). Consider a cost function $f : \mathbb{R}^N \to \mathbb{R}$ and a constraint function $g : \mathbb{R}^N \to \mathbb{R}^m$. The **primal problem** is $p^* \triangleq \inf\{f(x) : x \in \mathcal{X}\}$, where the constraint set is

$$\mathcal{X} \triangleq \cap_{i=1}^m \left\{x \in \mathbb{R}^N : g_i(x) \leqslant 0\right\}. \tag{2}$$

Definition 2.13 (Lagrangian). For the Problem 1, we define an associated Lagrangian function $\mathcal{L} : \mathbb{R}^N \times \mathbb{R}_+^m \to \mathbb{R}$ for Lagrange or dual variables $\alpha \in \mathbb{R}_+^m$ and primal variables $x \in \mathbb{R}^N$, as

$$\mathcal{L}(x,\alpha) \triangleq f(x) + \langle \alpha, g(x) \rangle. \tag{3}$$

Definition 2.14 (Dual function). The dual function $F : \mathbb{R}_+^m \to \mathbb{R}$ associated with the Problem 1 is defined for dual variables $\alpha \in \mathbb{R}_+^m$ as

$$F(\alpha) \triangleq \inf\left\{\mathcal{L}(x,\alpha) : x \in \mathbb{R}^N\right\}. \tag{4}$$

Theorem 2.15. *The following are true for the dual function $F : \mathbb{R}_+^m \to \mathbb{R}$ defined in (4) for the Problem 1.*

1. *F is concave in $\alpha \in \mathbb{R}_+^m$.*
2. *$F(\alpha) \leqslant \mathcal{L}(x,\alpha)$ for all $\alpha \in \mathbb{R}_+^m$ and $x \in \mathbb{R}^N$.*
3. *$F(\alpha) \leqslant p^*$ for all $\alpha \in \mathbb{R}_+^m$.*

Proof. Recall that $\mathcal{L}(\alpha) = f(x) + \langle \alpha, g(x) \rangle$ is a linear function of $\alpha \in \mathbb{R}_+^m$, and $F(\alpha) = \inf_x \mathcal{L}(x,\alpha)$.

1. Let $\beta \in [0,1]$ and $\alpha_1, \alpha_2 \in \mathbb{R}_+^m$ and $x \in \mathcal{X}$. It follows from the linearity of Lagrangian in α that

$$F(\beta\alpha_1 + \bar{\beta}\alpha_2) = \inf_x \left[\beta\mathcal{L}(x,\alpha_1) + \bar{\beta}\mathcal{L}(x,\alpha_2)\right] \geqslant \beta \inf_x \mathcal{L}(x,\alpha_1) + \bar{\beta} \inf_x \mathcal{L}(x,\alpha_2) = \beta F(\alpha_1) + \bar{\beta} F(\alpha_2).$$

2. From the definition of F, it follows that $F(\alpha) \leqslant \mathcal{L}(x,\alpha)$ for all $x \in \mathbb{R}^N$.
3. Recall that $g_i(x) \leqslant 0$ for all $x \in \mathcal{X}$, and hence $\langle \alpha, g(x) \rangle \leqslant 0$ for all $x \in \mathcal{X}$. Therefore, $F(\alpha) \leqslant f(x)$ for all $x \in \mathcal{X}$, and hence the result follows.

□

Problem 2. Dual problem The dual problem associated with primal problem defined in Problem 1 is

$$d^* \triangleq \max\{F(\alpha) : \alpha \in \mathbb{R}_+^m\}.$$

Remark 3. From the properties of dual function $F : \mathbb{R}_+^m \to \mathbb{R}$ in Theorem 2.15, we obtain that F is concave in $\alpha \in \mathbb{R}_+^m$. Since $\mathbb{R}_+^m$ is a convex set, it follows that the dual problem is convex. We further observe that the optimal value of dual problem $d^* \leqslant p^*$. The difference of optimal values $(p^* - d^*)$ is called the **duality gap**. For a primal problem, the **strong duality** holds if the duality gap is zero, or $d^* = p^*$.

2.3 Convex constrained optimization

Definition 2.16 (Saddle point). For a Lagrangian $\mathcal{L} : \mathbb{R}^N \times \mathbb{R}_+^m \to \mathbb{R}$, a saddle point (x^*, α^*) sastifies

$$\sup_{\alpha \in \mathbb{R}_+^m} \mathcal{L}(x^*,\alpha) \leqslant \mathcal{L}(x^*,\alpha^*) \leqslant \inf_{x \in \mathbb{R}^N} \mathcal{L}(x,\alpha^*).$$

Theorem 2.17 (Sufficient condition). *For the primal problem defined in Problem 1, if (x^*, α^*) is a saddle point of the associated Lagrangian $\mathcal{L}$, then $x^* \in \mathcal{X}$ and $p^* = f(x^*) = F(\alpha^*)$.*

Proof. Let (x^*, α^*) be the saddle point of the Lagrangian $\mathcal{L}$ associated with the Problem 1. From the definition of dual function F, we get that $\mathcal{L}(x^*,\alpha^*) \leqslant F(\alpha^*) \leqslant \mathcal{L}(x^*,\alpha^*)$. It follows that $F(\alpha^*) = \mathcal{L}(x^*,\alpha^*)$.

Recall that $\mathcal{L}(x^*,\alpha) = f(x^*) + \langle \alpha, g(x^*) \rangle$. We assume that there exists an $i \in [m]$ such that $g_i(x) > 0$, then we can take α_i large enough so that $\mathcal{L}(x^*,\alpha) \geqslant \mathcal{L}(x^*,\alpha^*)$. This contradicts the saddle point condition, and hence $x^* \in \mathcal{X}$. Therefore $\langle \alpha, g(x^*) \rangle \leqslant 0$ for all $\alpha \in \mathbb{R}_+^m$. This implies that $\langle \alpha^*, g(x^*) \rangle = 0$ and hence $p^* = f(x^*) = F(\alpha^*)$. □

Definition 2.18 (Strong constraint qualification). The strong constraint qualification or **Slater's condition** is defined as the existence of a point $x \in \mathcal{X}^o$ such that $g_i(x) < 0$ for all $i \in [m]$.

Theorem 2.19 (Strong necessary condition). *Let the cost function f and constraints g_i for $i \in [m]$ be convex functions, such that the Slater's condition holds, and x^* be the solution of the Problem 1 . Then, there exists $\alpha^* \in \mathbb{R}_+^m$ such that (x^*, α^*) is a saddle point of the associated Lagrangian $\mathcal{L}$.*

Definition 2.20. A function $h : \mathcal{X} \to \mathbb{R}$ is said to be affine if it can be defined as $x \mapsto h(x) \triangleq \langle w, x \rangle + b$ for all $x \in \mathcal{X} \subseteq \mathbb{R}^N$ and some $w \in \mathbb{R}^N$ and $b \in \mathbb{R}$.

Definition 2.21 (Weak constraint qualification). The weak constraint qualification or **weak Slater's condition** is defined as the existence of a point $x \in \mathcal{X}^o$ such that for each $i \in [m]$ either $g_i(x) < 0$ or $g_i(x) = 0$ and g_i affine.

Theorem 2.22 (Weak necessary condition). *Let the cost function f and constraints g_i for $i \in [m]$ be convex differentiable functions, such that the weak Slater's condition holds, and x^* be the solution of the Problem 1 . Then, there exists $\alpha^* \in \mathbb{R}_+^m$ such that (x^*, α^*) is a saddle point of the associated Lagrangian $\mathcal{L}$.*

Remark 4. The strong duality holds when the primal problem is convex with qualifying constraints.

Theorem 2.23 (Karush-Kuhn-Tucker (KKT)). *Let the cost function f and constraint functions g_i for all $i \in [m]$ be convex and differentiable functions, such that the constraints are qualified. Then $x^* \in \mathbb{R}^N$ is a solution of the constrained problem iff there exists $\alpha^* \in \mathbb{R}_+^m$ such that*

$$\nabla_x \mathcal{L}(x^*, \alpha^*) = \nabla_x f(x^*) + \sum_{i=1}^m \alpha_i^* \nabla_x g_i(x^*) = 0, \quad \nabla_{\alpha_i} \mathcal{L}(x^*, \alpha^*) = g_i(x^*) \leqslant 0, i \in [m], \quad \sum_{i=1}^m \alpha_i^* g_i(x^*) = 0.$$

Proof. We will first assume that $x^* \in \mathcal{X}$ is solution of the constrained problem and show that the there exists $\alpha^* \in \mathbb{R}_+^m$ such that the three conditions are met. From the definition of Lagrangian $\mathcal{L}$, we have

$$\nabla_x \mathcal{L}(x^*, \alpha^*) = \nabla_x f(x^*) + \sum_{i=1}^m \alpha_i^* \nabla_x g_i(x^*).$$

From the necessary condition theorem, it follows that if $x^* \in \mathcal{X}$ is a solution to the primal problem, then there exists dual variables $\alpha^* \in \mathbb{R}_+^m$ such that (x^*, α^*) is a saddle point of the Lagrangian. It follows that $\nabla_x \mathcal{L}(x^*, \alpha^*) = 0$ and the first condition holds. Since $x^* \in \mathcal{X}$, it follows that the second condition $g_i(x^*) \leqslant 0$ holds for all $i \in [m]$. Since (x^*, α^*) is a saddle point of the Lagrangian, we have for any $\alpha \in \mathbb{R}_+^m$

$$\mathcal{L}(x^*, \alpha^*) = f(x^*) + \sum_{i=1}^m \alpha_i^* g_i(x^*) \geqslant f(x^*) + \sum_{i=1}^m \alpha_i g_i(x^*).$$

This implies that $-\sum_{i=1}^m \alpha_i^* g_i(x^*) \leqslant \inf_{\alpha \in \mathbb{R}_+^m} -\sum_{i=1}^m \alpha_i g_i(x^*) = 0$.

Conversely, let $(x^*, \alpha^*) \in \mathbb{R}^N \times \mathbb{R}_+^m$ be such that all three KKT conditions are met. We will show that $f(x) \geqslant f(x^*)$ for all $x \in \mathcal{X}$. We first observe from the convexity of f, that for any $x \in \mathbb{R}^{\mathbb{N}}$, we have

$$f(x) - f(x^*) \geqslant \langle \nabla_x f(x^*), x - x^* \rangle.$$

From the first condition, we get $\langle \nabla_x f(x^*), x - x^* \rangle = -\sum_{i=1}^m \alpha_i^* \langle \nabla_x g_i(x^*), x - x^* \rangle$. From the convexity of g_i for all $i \in [m]$ and the third condition, we have

$$-\sum_{i=1}^m \alpha_i^* \langle \nabla_x g_i(x^*), x - x^* \rangle \geqslant -\sum_{i=1}^m \alpha_i^* (g_i(x) - g_i(x^*)) = -\sum_{i=1}^m \alpha_i^* g_i(x).$$

Recall that $g_i(x) \leqslant 0$ for all $x \in \mathcal{X}$ and $i \in [m]$. Thus, combining all these inequalities, we get $f(x) - f(x^*) \geqslant 0$ for all $x \in \mathcal{X}$. □