APM462: Nonlinear Optimization

Tianyu Du

October 12, 2019

Contents

1 Preliminaries

1.1 Mean Value Theorems and Taylor Approximations.

Definition 1.1. Let $f : S \subset \mathbb{R}^n \to \mathbb{R}$, the **gradient** of f at $x \in S$, if exists, is a vector $\nabla f(x) \in \mathbb{R}^n$ characterized by the property

$$\lim_{v \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{||v||} = 0 \tag{1.1}$$

Theorem 1.1 (The First Order of Mean Value Theorem)**.** Let f be a C^1 real-valued function defined on $\mathbb{R}^n$, then for any $x, v \in \mathbb{R}^n$, there exists some $\theta \in (0,1)$ such that

$$f(x+v) = f(x) + \nabla f(x + \theta v) \cdot v \tag{1.2}$$

Proof. Let $x, v \in \mathbb{R}^n$, define $g(t) : \mathbb{R} \to \mathbb{R} := f(x+tv)$, which is C^1. By the mean value theorem on $\mathbb{R}^{\mathbb{R}}$, there exists $\theta \in (0,1)$ such that $g(0+1) = g(0) + g'(\theta)(1-0)$, that is, $f(x+v) = f(x) + g'(\theta)$. Note that $g'(\theta) = \nabla(x + \theta v) \cdot v$, what desired is immediate. ■

Proposition 1.1 (The First Order Taylor Approximation)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + o(||v||) \tag{1.3}$$

that is

$$\lim_{||v|| \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{||v||} = 0 \tag{1.4}$$

Proof. By the mean value theorem, $\exists \theta \in (0,1)$ such that $f(x+v) - f(x) = \nabla f(x+\theta v) \cdot v$. The limit becomes $\lim_{||v|| \to 0} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{||v||} = \lim_{||v|| \to 0; x+\theta v \to x} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{||v||}$. Since $f \in C^1$, $\lim_{x+\theta v \to x} \nabla f(x+\theta v) = \nabla f(x)$. And $\frac{v}{||v||}$ is a unit vector, and every component of it is bounded, as the result, the limit of inner product vanishes instead of explodes. ■

Theorem 1.2 (The Second Order Mean Value Theorem)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, then for any $x, v \in \mathbb{R}^n$, there exists $\theta \in (0,1)$ satisfying

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v' H_f(x+\theta v)\ v \tag{1.5}$$

where H_f is the Hessian matrix of f, may also be written as $\nabla^2 f$.

Proposition 1.2 (The Second Order Taylor Approximation)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, and $x, v \in \mathbb{R}^n$, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v' H_f(x)\ v + o(||v||^2) \tag{1.6}$$

that is

$$\lim_{||v|| \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v - \frac{1}{2} v' H_f(x)\ v}{||v||^2} = 0 \tag{1.7}$$

Proof. By the second mean value theorem, there exists $\theta \in (0,1)$ such that the limit is equivalent to

$$\lim_{||v|| \to 0} \frac{1}{2} \left(\frac{v}{||v||} \right)' [H_f(x+\theta v) - H_f(x)] \frac{v}{||v||} \tag{1.8}$$

Since $f \in C^2$, the limit of $[H_f(x+\theta v) - H_f(x)]$ is in fact $\mathbf{0}_{n \times n}$. And every component of unit vector $\frac{v}{||v||}$ is bounded, the quadratic form converges to zero as an immediate result. ■

It is often noted that the gradient at a particular $x_0 \in dom(f) \subset \mathbb{R}^n$ gives the direction f increases most rapidly. Let $x_0 \in dom(f)$, and v be a unit vector representing a *feasible direction* of change. That is, there exists $\delta > 0$ such that $x_0 + tv \in dom(f)\ \forall t \in [0, \delta)$. Then the rate of change of f along feasible direction v can be written as

$$\left.\frac{d}{dt}\right|_{t=0} f(x_0 + tv) = \nabla f(x_0) \cdot v = ||\nabla f(x_0)||\ ||v|| \cos(\theta) \tag{1.9}$$

where $\theta = \angle(v, \nabla f(x_0)$. And the derivative is maximized when $\theta = 0$, that is, when v and ∇f point the same direction.

1.2 Implicit Function Theorem

Theorem 1.3 (Implicit Function Theorem)**.** Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a C^1 function, let $(a, b) \in \mathbb{R}^n \times \mathbb{R}$ such that $f(a, b) = 0$. If $\nabla f(a, b) \neq 0$, then $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x, y) = 0\}$ is locally a graph of a function $g : \mathbb{R}^n \to \mathbb{R}$.

Remark 1.1. $\nabla f(x_0) \perp$ level set of f near x_0.

2 Convexity

2.1 Terminologies

Definition 2.1. Set $\Omega \subset \mathbb{R}^n$ is **convex** if and only if

$$\forall x_1, x_2 \in \Omega,\ \lambda \in [0, 1],\ \lambda x_1 + (1 - \lambda)x_2 \in \Omega \tag{2.1}$$

Definition 2.2. A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is **convex** if and only if Ω is convex, and

$$\forall x_1, x_2 \in \Omega,\ \lambda \in [0, 1],\ f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{2.2}$$

Definition 2.3. A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if and only if Ω is convex and

$$\forall x_1, x_2 \in \Omega,\ \lambda \in (0, 1),\ f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{2.3}$$

2.2 Basic Properties of Convex Functions

Definition 2.4. A function $f : \Omega \to \mathbb{R}$ is **concave** if and only if $-f$ is **convex**.

Proposition 2.1.
(i) If f_1, f_2 are convex on Ω, so is $f_1 + f_2$;

(ii) If f is convex on Ω, then for any $a > 0$, af is also convex on Ω;

(iii) Any **sub-level/lower contour set** of a convex function f

$$SL(c) := \{x \in \mathbb{R}^n : f(x) \leq c\} \tag{2.4}$$

is convex.

Proof of (iii). Let $c \in \mathbb{R}$, and $x_1, x_2 \in SL(c)$. Let $s \in [0, 1]$. Since $x_1, x_2 \in SL(c)$, and $f(\cdot)$ is convex, $f(sx_1 + (1 - s)x_2) \leq sf(x_1) + (1 - s)f(x_2) \leq sc + (1 - s)c = c$. Which implies $sx_1 + (1 - s)x_2 \in SL(c)$. ∎

Example 2.1. $f(x) : \mathbb{R}^n \to \mathbb{R} := ||x||$ is convex.

Proof. Note that for any $u, v \in \mathbb{R}^n$, by triangle inequality, $||u - (-v)|| \le ||u - 0|| + ||0 - (-v)|| = ||u|| + ||v||$. Consequently, let $u, v \in \mathbb{R}^n$ and $s \in [0,1]$, then $||su + (1-s)v|| \le ||su|| + ||(1-s)v|| = s||u|| + (1-s)||v||$. Therefore, $||\cdot||$ is convex. ■

2.3 Characteristics of C^1 Convex Functions

Theorem 2.1 (C^1 criterions for convexity)**.** Let $f \in C^1$, then f is convex on a convex set Ω if and only if

$$\forall x, y \in \Omega,\ f(y) \ge f(x) + \nabla f(x) \cdot (y - x) \tag{2.5}$$

that is, *the linear approximation is never an overestimation of value of f.*

Proof. ($\implies$) Suppose f is convex on a convex set Ω. Then $f(sy + (1-s)x) \le sf(y) + (1-s)f(x)$ for every $x, y \in \Omega$ and $s \in [0,1]$, which implies, for every $s \in (0,1]$:

$$\frac{f(sy + (1-s)x) - f(x)}{s} \le f(y) - f(x) \tag{2.6}$$

By taking the limit of $s \to 0$,

$$\lim_{s \to 0} \frac{f(x + s(y-x)) - f(x)}{s} \le f(y) - f(x) \tag{2.7}$$

$$\implies \left.\frac{d}{ds}\right|_{s=0} f(x + s(y-x)) \le f(y) - f(x) \tag{2.8}$$

$$\implies \nabla f(x) \cdot (y - x) \le f(y) - f(x) \tag{2.9}$$

($\impliedby$) Let $x_0, x_1 \in \Omega$, let $s \in [0,1]$. Define $x^* := sx_0 + (1-s)x_1$, then

$$f(x_0) \ge f(x^*) + \nabla f(x^*) \cdot (x_0 - x^*) \tag{2.10}$$

$$\implies f(x_0) \ge f(x^*) + \nabla f(x^*) \cdot [(1-s)(x_0 - x_1)] \tag{2.11}$$

Similarly,

$$f(x_1) \ge f(x^*) + \nabla f(x^*) \cdot (x_1 - x^*) \tag{2.12}$$

$$\implies f(x_1) \ge f(x^*) + \nabla f(x^*) \cdot [s(x_1 - x_0)] \tag{2.13}$$

Therefore, $sf(x_0) + (1-s)f(x_1) \ge f(x^*)$. ■

Theorem 2.2 (C^2 criterion for convexity)**.** $f \in C^2$ is a convex function on a convex set $\Omega \subset \mathbb{R}^n$ if and only if $\nabla^2 f(x) \succcurlyeq 0$ for all $x \in \Omega$.

Remark 2.1. When f is defined on $\mathbb{R}$, the C^2 criterion becomes $f''(x) \ge 0$.

Proof. ($\impliedby$) Suppose $\nabla^2 f(x) \succcurlyeq 0$ for every $x \in \Omega$, let $x, y \in \Omega$. By the second order MVT,

$$f(y) = f(x) + \nabla f(x) \cdot (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x + s(y-x))(y-x) \text{ for some } s \in [0,1] \tag{2.14}$$

$$\implies f(y) \ge f(x) + \nabla f(x) \cdot (y - x) \tag{2.15}$$

So f is convex by the C^1 criterion of convexity.
($\implies$) Let $v \in \mathbb{R}^n$. Suppose, for contradiction, that for some $x \in \Omega$, $\nabla^2 f(x) \not\succcurlyeq 0$. If such $x \in \partial\Omega$, note that $v^T \nabla^2 f(\cdot) v$ is continuous because $f \in C^2$, then there exists $\varepsilon > 0$ such that $\forall x' \in V_\varepsilon(x) \cap \Omega^{int},\ v^T \nabla^2 f(x') v <$

0. Hence, one may assume with loss of generality that such $x \in \Omega^{int}$. Because $x \in \Omega^{int}$, exists $\varepsilon' > 0$, such that $V_{\varepsilon'}(x) \subseteq \Omega^{int}$. Define $\hat{v} := \frac{v}{\sqrt{\varepsilon'}}$, then for every $s \in [0,1]$, $\hat{v}^T \nabla^2 f(x + s\hat{v})\hat{v} < 0$. Let $y = x + \hat{v}$, by the mean value theorem, $f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2}(y-x)^T \nabla^2 f(x + s(y-x))(y-x)$ for some $s \in [0,1]$. This implies $f(y) < f(x) + \nabla f(x) \cdot (y - x)$, which contradicts the C^1 criterion for convexity. ■

2.4 Minimum and Maximum of Convex Functions

Theorem 2.3. Let $\Omega \subset \mathbb{R}^n$ be a convex set, and $f : \Omega \to \mathbb{R}$ is a convex function. Let

$$\Gamma := \left\{ x \in \Omega : f(x) = \min_{x \in \Omega} f(x) \right\} \equiv \operatorname*{argmin}_{x \in \Omega} f(x) \tag{2.16}$$

If $\Gamma \neq \varnothing$, then

(i) Γ is convex;

(ii) any local minimum of f is the global minimum.

Proof (i). Let $x, y \in \Gamma$, $s \in [0,1]$, then $sx + (1-s)y \in \Omega$ because Ω is convex. Since f is convex, $f(sx + (1-s)y) \leq sf(x) + (1-s)f(y) = \min_{x \in \Omega} f(x)$. The inequality must be equality since it would contradicts the fact that $x, y \in \Gamma$. Therefore, $sx + (1-s)y \in \Gamma$. ■

Proof (ii). Let $x \in \Omega$ be a local minimizer for f, but assume, for contradiction, it is not a global minimizer. That is, there exists some other y such that $f(y) < f(x)$. Since f is convex,

$$f(x + t(y - x)) = f((1-t)x + ty) \leq (1-t)f(x) + tf(y) < f(x) \tag{2.17}$$

for every $t \in (0,1]$. Therefore, for every $\varepsilon > 0$, there exists $t^* \in (0,1]$ such that $x + t^*(y - x) \in V_\varepsilon(x)$ and $f(x + t^*(y-x)) < f(x)$, this contradicts the fact that x is a local minimum. ■

Theorem 2.4. Let $\Omega \subset \mathbb{R}^n$ be a convex and compact set, and $f : \Omega \to \mathbb{R}$ is a convex function. Then

$$\max_{x \in \Omega} f(x) = \max_{x \in \partial\Omega} f(x) \tag{2.18}$$

Proof. As we assumed, Ω is closed, therefore $\partial\Omega \subseteq \Omega$. Hence, $\max_{x \in \Omega} f \geq \max_{x \in \partial\Omega} f$. Suppose $\max_{x \in \Omega} f > \max_{x \in \partial\Omega} f$, let $x^* := \operatorname{argmax}_{x \in \Omega} f \in \Omega^{int}$. Then we can construct a straight line through x^* and intersects $\partial\Omega$ at two points, $y_1, y_2 \in \partial\Omega$, such that $x^* = sy_1 + (1-s)y_2$ for some $s \in (0,1)$. Further, since f is convex, $\max_{x \in \Omega} f(x) = f(x^*) \leq sf(y_1) + (1-s)f(y_2) \leq s \max_{\partial\Omega} f + (1-s)\max_{\partial\Omega} f = \max_{\partial\Omega} f$, which leads to a contradiction. Therefore, $\max_{x \in \Omega} f = \max_{x \in \partial\Omega} f$. ■

Proposition 2.2. For $p, g > 1$ and $\frac{1}{p} + \frac{1}{g} = 1$,

$$|ab| \leq \frac{1}{p}|a|^p + \frac{1}{g}|b|^g \tag{2.19}$$

Proof.

$$\begin{aligned}
(-\log)|ab| &= (-\log)|a| + (-\log)|b| && (2.20)\\
&= \frac{1}{p}(-\log)|a|^p + \frac{1}{g}(-\log)|b|^p && (2.21)\\
(\because (-\log) \text{ is convex}) \ &\geq (-\log)\left(\frac{1}{p}|a|^p + \frac{1}{g}|b|^p\right) && (2.22)
\end{aligned}$$

And since $(-\log)$ is monotonically decreasing,

$$|ab| \leq \frac{1}{p}|a|^p + \frac{1}{g}|b|^p \tag{2.23}$$

■

Corollary 2.1.

$$|ab| \leq \frac{|a|^2 + |b|^2}{2} \tag{2.24}$$

3 Finite Dimensional Optimization

3.1 Unconstraint Optimization

Theorem 3.1 (Extreme Value Theorem)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}^n$ be a compact set, then the minimization problem $\min_{x \in K} f(x)$ has a solution.

Remark 3.1. $f : \Omega \to \mathbb{R}$ is convex does not imply f is continuous.

Proposition 3.1. A convex function f defined on a convex open set is continuous.

Proof. Let $f : \Omega \to \mathbb{R}$ be a convex function, where $\Omega \subset \mathbb{R}^n$ is open. TODO ■

Corollary 3.1. A convex function f defined on an open interval in $\mathbb{R}$ is continuous.

Proof. See homework 1, using squeeze theorem. ■

Proof of EVT.. Let $f : K \to \mathbb{R}$ be a continuous function defined on a compact set K.
WLOG, we only prove the existence of $\min f$, since the existence of max can be easily proven by applying the exact same argument on $-f$. Because K is compact, the continuity of f implies $f(K)$ is compact. By the completeness axiom of $\mathbb{R}$, $m := \inf_{x \in K} f(x)$ is well-defined. There exists a sequence $(x_i) \subset K$, such that $(f(x_i)) \to m$. Because K is compact, there exists a subsequence (x_{ik}) of (x_i) converges to some limit $x^* \in K$. Because f is continuous, $(f(x_{ik})) \to f(x^*)$, which is a subsequence of the convergent sequence $(f(x_i))$, and they must converge to the same limit. Hence, $f(x^*) = m$, and the infimum is attained at $x^* \in K$. ■

Theorem 3.2 (Heine–Borel)**.** Let $K \subset \mathbb{R}^n$, then K is compact (every open cover of K has a finite sub-cover) $\iff$ K is closed and bounded.

Proposition 3.2. Let $\{h_i\}$ and $\{g_i\}$ be sets of continuous functions on $\mathbb{R}^n$, the the set of all points in $\mathbb{R}^n$ that satisfy

$$\begin{cases} h_i(x) = 0 \ \forall i \\ g_j(x) \leq 0 \ \forall j \end{cases} \tag{3.1}$$

is a closed set (intersection of finitely many closed sets). Moreover, if the qualified set is also bounded, then it is compact.

Proof. For every equality constraint h_i, it can be represented as the conjunction of two inequality constraint, namely $h_i^\alpha(x) := -h_i(x) \leq 0 \land h_i^\beta(x) := h_i(x) \leq 0$. Then the constraint collection is equivalent to

$$\begin{cases} h_i^\alpha(x) \leq 0 \ \forall i \\ h_i^\beta(x) \leq 0 \ \forall i \\ g_j(x) \leq 0 \ \forall j \end{cases} \tag{3.2}$$

The subset of $\mathbb{R}^n$ qualified by each individual constraint is closed by the property of continuous functions (i.e. the continuous function's pre-image of closed set is closed). And the intersection of arbitrarily many closed sets is closed. ■

Example 3.1. The set $\{(x, y) \in \mathbb{R}^2 : x^2 - y^2 - 1 = 0\}$ is closed and bounded, therefore it is compact.

Remark 3.2. Computer algorithms for solving minimization problems try to construct a sequence of (x_i) such that $f(x_i)$ decreases to $\min f$ rapidly.

The optimization problems investigated in this section can be formulated as

$$\min_{x \in \Omega} f(x) \tag{3.3}$$

where $\Omega \subset \mathbb{R}^n$. Typically, for simplicity, Ω are often $\mathbb{R}^n$, an open subset of R^n, or the closure of some open subset of $\mathbb{R}^n$.

Everything above minimization discussed in this section is applicable to maximization as well using the proposition below.

Proposition 3.3. When $\Omega = \mathbb{R}^n$, the unconstrained minimization has the following properties

(i) $\text{argmax}\, f = \text{argmin}(-f)$;

(ii) $\max f = -\min(-f)$

Proof. Omitted. ■

Definition 3.1. A function $f : \Omega \to \mathbb{R}$ has **local minimum** at $x_0 \in \Omega$ if

$$\exists \varepsilon > 0 \ s.t. \ \forall x \in V_\varepsilon(x_0) \cap \Omega \ f(x_0) \leq f(x) \tag{3.4}$$

f attains **strictly local minimum** at x_0 if

$$\exists \varepsilon > 0 \ s.t. \ \forall x \in V_\varepsilon(x_0) \cap \Omega \backslash \{x_0\} \ f(x_0) < f(x) \tag{3.5}$$

f attains **global minimum** at x_0 if

$$\forall x \in \Omega \ f(x_0) \leq f(x) \tag{3.6}$$

f attains **strict global minimum** at x_0 if

$$\forall x \in \Omega \backslash \{x_0\} \ f(x_0) < f(x) \tag{3.7}$$

Note that strict global minimum is always unique.

Theorem 3.3 (Necessary Condition for Local Minimum)**.** Let $C^1 \ni f : \Omega \to \mathbb{R}$, let $x_0 \in \Omega$ be a local minimum of f, then for every *feasible direction* v at x_0,

$$\nabla f(x_0) \cdot v \geq 0 \tag{3.8}$$

Definition 3.2. For $x_0 \in \Omega \subset \mathbb{R}^n$, $v \in \mathbb{R}^n$ is a **feasible direction**at x_0 if

$$\exists \overline{s} > 0 \ s.t. \ \forall s \in [0, \overline{s}], x_0 + sv \in \Omega \tag{3.9}$$

Proof of Necessary Condition. Let $x_0 \in \Omega$ be a local minimum, and let v be a Define auxiliary function $g(s) := f(x + sv)$. And since g attains minimum at $s = 0$, there exists some $\overline{s} > 0$ such that

$$g(s) - g(0) \geq 0 \ \forall s \in [0, \overline{s}] \tag{3.10}$$

Therefore

$$g'(0) := \lim_{s \to 0} \frac{g(s) - g(0)}{s - 0} \geq 0 \tag{3.11}$$

The alternative form of derivative can be derived using chain rule as

$$g'(0) = \nabla f(x + sv) \cdot v \mid_{s=0} = \nabla f(x) \cdot v \tag{3.12}$$

By combing the two identities above, $\nabla f(x) \cdot v \geq 0$. ■

Alternative Proof of Necessary Condition (not that rigorous). The prove is almost immediate, if there exists a feasible direction v^* such that $\nabla f(x_0) \cdot v^* < 0$, for every $\varepsilon > 0$, one can construct $x' := x^* + sv^*$ with sufficiently small s so that $x' \in V_\varepsilon(x^*) \cap \Omega$ and $f(x') < f(x^*)$. ■

Corollary 3.2. When Ω is open, then x_0 is a local minimum $\implies \nabla f(x_0) = 0$.

Proof. Since Ω is open, any sufficiently small $v \neq 0$ such that both v and $-v$ are feasible directions at x_0, applying the necessary condition on both v and $-v$ provides the equality. ■

Example 3.2. Minimize $f(x, y) = x^2 - xy + y^2 - 3y$ over $\Omega = \mathbb{R}^2$.

Example 3.3. Minimize $f(x, y) = x^2 - x + y + xy$ over $\Omega = \max\{(x, y) \in \mathbb{R}^2 : x, y \geq 0\}$.

Theorem 3.4 (Second Order Necessary Condition for Local Minimum)**.** Let $C^2 \ni f : \Omega \to \mathbb{R}$, let $x_0 \in \Omega$ be a local minimum of f, then for every non-zero feasible direction v at x_0,

(i) $\nabla f(x_0) \cdot v \geq 0$;

(ii) $\nabla f(x_0) \cdot v = 0 \implies v^T \nabla^2 f(x_0) v \geq 0$.

Proof. Let x_0 be a local minimum and v be a feasible direction at Ω, and $s \in (0, \overline{s}]$. The first statement is the immediate result of the first order necessary condition. Now suppose $\nabla f(x_0) = 0$, by the Taylor's theorem,

$$\begin{aligned} 0 \leq f(x_0 + sv) - f(x_0) &= s \nabla f(x_0) \cdot v + \frac{s^2}{2} v^T \nabla^2 f(x_0) v + o(s^2) && (3.13) \\ &= \frac{s^2}{2} v^T \nabla^2 f(x_0) v + o(s^2) && (3.14) \end{aligned}$$

Since $s^2 > 0$, divide both sides by s^2 and take limit,

$$\begin{aligned} \lim_{s \to 0} \frac{f(x_0 + sv) - f(x_0)}{s^2} &= \lim_{s \to 0} \left\{ \frac{1}{2} v^T \nabla^2 f(x_0) v + \frac{o(s^2)}{s^2} \right\} && (3.15) \\ &= \frac{1}{2} v^T \nabla^2 f(x_0) v + \lim_{s \to 0} \frac{o(s^2)}{s^2} && (3.16) \\ &= \frac{1}{2} v^T \nabla^2 f(x_0) v \geq 0 && (3.17) \end{aligned}$$

■

Example 3.4. $f(x,y) = x^2 - xy + y^2 - 3y : \Omega = \mathbb{R}^2 \to \mathbb{R}$. Then at $(x_0, y_0) = (1, 2)$,

$$\nabla f(x_0, y_0) = (2x_0 - y, -x_0 + 2y_0 - 3) = (0, 0) \tag{3.18}$$

$$\nabla^2 f(x_0, y_0) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \succcurlyeq 0 \tag{3.19}$$

Definition 3.3. Let $A \in \mathbb{R}^{n \times n}$, A is

(i) **Positive definite** ($A \succ 0$) if $x^T Ax > 0 \ \forall x \neq 0$, if and only if all eigenvalues $\lambda_i > 0$;

(ii) **Positive Semi-definite** ($A \succcurlyeq 0$) if $x^T Ax \geq \ \forall x \in \mathbb{R}^n$, if and only if all eigenvalues $\lambda_i \geq 0$.

Theorem 3.5 (Sylvester's Criterion)**.** Let $A \in \mathbb{R}^{n \times n}$ be a Hermitian matrix (i.e. $A = \overline{A^T}$), then

1. $A \succ 0 \iff$ all *leading principal minors* have positive determinants;
2. $A \succcurlyeq 0 \iff$ all leading principal minors have non-negative determinants.

Theorem 3.6 (Second Order Sufficient Condition for Interior Local Minima)**.** Let $C^2 \ni f : \Omega \to \mathbb{R}$, for some $x_0 \in \Omega$, if

(i) $\nabla f(x_0) = 0$,

(ii) *(and)* $\nabla^2 f(x_0) \succcurlyeq 0$,

then x_0 is a <u>strictly local minimizer</u>.

Lemma 3.1. Suppose $\nabla^2 f(x_0)$ is positive definite, then

$$\exists a > 0 \ s.t. \ v^T \nabla^2 f(x_0) v \geq a||v||^2 \quad \forall v \tag{3.20}$$

Proof of the Lemma. Recall that a squared matrix Q is called **orthogonal** when every column and row of it is an orthogonal unit vector. So that for every orthogonal matrix Q, $Q^T Q = I$, which implies $Q^T = Q^{-1}$. Further, note that

$$||Qv||^2 = (Qv)^T (Qv) = v^T Q^T Qv = ||v||^2 \tag{3.21}$$

$$\implies ||Qv|| = ||v|| \ \forall v \in \mathbb{R}^n \tag{3.22}$$

Let $v \in \mathbb{R}^n$, consider the eigenvector decomposition of $\nabla^2 f(x_0)$, let w satisfy $v = Qw$:

$$Q^T \nabla^2 f(x_0) Q = \operatorname{diag}(\lambda_1, \cdots, \lambda_n) \tag{3.23}$$

$$\implies v^T \nabla^2 f(x_0) v = (Qw)^T \nabla^2 f(x_0)(Qw) \tag{3.24}$$

$$= w^T Q^T \nabla^2 f(x_0) Qw \tag{3.25}$$

$$= w^T \operatorname{diag}(\lambda_1, \cdots, \lambda_n) w \tag{3.26}$$

$$= \lambda_1 w_1^2 + \cdots + \lambda_n w_n^2 \tag{3.27}$$

Let $a := \min\{\lambda_1, \cdots, \lambda_n\}$,

$$... \geq a||w||^2 = a||Q^T v||^2 = a||v||^2 \tag{3.28}$$

∎

Proof of the Theorem. Let $x \in \Omega$, suppose $\nabla f(x_0) = 0$ and $\nabla^2 f(x_0) \succcurlyeq 0$. By the second order Taylor approximation,

$$
\begin{aligned}
f(x_0 + v) - f(x_0) &= \nabla f(x_0)^T v + \frac{1}{2} v^T \nabla^2 f(x_0) v + o(||v||^2) && (3.29)\\
&= \frac{1}{2} v^T \nabla^2 f(x_0) v + o(||v||^2) && (3.30)\\
&\geq \frac{a}{2} ||v||^2 + o(||v||^2) \text{ for some } a > 0 && (3.31)\\
&= ||v||^2 \left(\frac{a}{2} + \frac{o(||v||^2)}{||v||} \right) && (3.32)\\
&> 0 \ \text{ for sufficiently small } v && (3.33)
\end{aligned}
$$

Therefore, $f(x_0) < f(x)\ \forall x \in V_\varepsilon(x_0)$. ■

3.2 Equality Constraints: Lagrangian Multiplier

3.2.1 Tangent Space to a (Hyper) Surface at a Point

Definition 3.4. A **surface** $\mathcal{M} \subset \mathbb{R}^n$ is defined as

$$\mathcal{M} := \{x \in \mathbb{R}^n : h_i(x) = 0\ \forall i\} \tag{3.34}$$

where h_i are all C^1 functions.

Definition 3.5. A **differentiable curve** on a surface $\mathcal{M}$ is a C^1 function mapping from $(-\varepsilon, \varepsilon)$ to $\mathcal{M}$. *Remark: in previous calculus courses, differentiable curves are often referred to as parameterizations.*

Let $x(s)$ be a differentiable curve on $\mathcal{M}$ passes through $x_0 \in \mathcal{M}$, WLOG, $x(0) = x_0$. Then vector

$$v := \frac{d}{ds}\bigg|_{s=0} x(s) \tag{3.35}$$

touches $\mathcal{M}$ *tangentially.*

Definition 3.6. Any vector v generated by some differentiable curve on $\mathcal{M}$ and takes above form is a **tangent vector** on $\mathcal{M}$ through x_0.

Definition 3.7. The set of all tangent vectors is defined to be the **tangent space** to $\mathcal{M}$ at x_0:

$$T_{x_0}\mathcal{M} := \left\{ v \in \mathbb{R}^n : v := \frac{d}{ds}\bigg|_{s=0} x(s) \text{ for some } x(\cdot) \in \mathcal{M}^{(-\varepsilon,\varepsilon)}\ s.t.\ x(0) = x_0 \right\} \tag{3.36}$$

Example 3.5. Define

$$\mathcal{M} := \left\{ x \in \mathbb{R}^2 : ||x||_2 = 1 \right\} \tag{3.37}$$

By defining C^1 functions $g(x) := ||x||_2^2 - 1$, $\mathcal{M}$ is a surface. The tangent space of $\mathcal{M}$ at x_0 is

$$T_{x_0}\mathcal{M} = \{v \in \mathbb{R}^n : \langle v, x_0 \rangle = 0\} \tag{3.38}$$

Definition 3.8. Let $\mathcal{M}$ be a surface defined using C^1 functions, a point $x_0 \in \mathcal{M}$ is a **regular point** of the constraints if

$$\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\} \tag{3.39}$$

are linearly independent.

Notation 3.1. Define

$$T_{x_0} := \{x \in \mathbb{R}^n : \langle x_0, \nabla h_i(x_0)\rangle \ \forall i \in [k]\} \tag{3.40}$$

Example 3.6 (Counter example)**.** Define

$$\mathcal{M} := \big\{(x, y) \in \mathbb{R}^2 : h(x, y) = xy = 0\big\} \tag{3.41}$$

Then it is easy to verify that $(0, 0)$ is not a regular point. And

$$T_{0,0} = \{(x, y) \in \mathbb{R}^2 : (x, y) \cdot (0, 0) = 0\} = \mathbb{R}^2 \tag{3.42}$$

$$\neq T_{0,0}\mathcal{M} = \{(x, y) \in \mathbb{R}^2 : x = 0 \vee y = 0\} \tag{3.43}$$

Theorem 3.7. Suppose x_0 is a *regular point* of $\mathcal{M} := \{h_i(x) = 0, i = 1, \cdots, k\}$, then $T_{x_0} = T_{x_0}\mathcal{M}$.

Proof. Show $T_{x_0}\mathcal{M} \subset T_{x_0}$.
Suppose x_0 is a regular point of $\mathcal{M}$. Let $v \in T_{x_0}\mathcal{M}$, then there exists some differentiable curve $x(\cdot) : V_\varepsilon(0) \to \mathcal{M}$ such that $x(0) = x_0$, such that

$$v = \frac{d}{ds}\bigg|_{s=0} x(s) \tag{3.44}$$

Note that $h_i(x(s)) = 0$ is constant for every $i \in [k]$, therefore

$$\frac{d}{ds}\bigg|_{s=0} h_i(x(s)) \tag{3.45}$$

By the chain rule,

$$\nabla h_i(x_0) \cdot v = 0 \ \forall i \tag{3.46}$$

Therefore $v \in T_{x_0}$.
Show $T_{x_0} \subset T_{x_0}\mathcal{M}$.

(i) x_0 is regular $\implies T_{x_0}\mathcal{M}$ is a vector space;

(ii) $T_{x_0} = \text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\}^\perp$.

Show $T_{x_0} \subset \text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\}^\perp$:
Let $v \in T_{x_0}$, then $v \perp \nabla h_i(x_0)$ for every i. Therefore v is orthogonal to every linear combination of $\nabla h_i(x_0)$, and therefore orthogonal to the span.
Show $\text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\}^\perp \subset T_{x_0}$:
Let v in the perp of the span, then v is orthogonal to every basis of the span, so $v \in T_{x_0}$. ■

Lemma 3.2. Let $f, h_1, \cdots, h_k \in C^1$ defined on <u>open</u> subset $\Omega \subset \mathbb{R}^n$. Define $\mathcal{M} := \{x \in \mathbb{R}^n : h_i(x) = 0 \ \forall i\}$. Suppose $x_0 \in \mathcal{M}$ is a local minimum of f on $\mathcal{M}$, then

$$\nabla f(x_0) \perp T_{x_0}\mathcal{M} \tag{3.47}$$

Proof. WLOG $\Omega = \mathbb{R}^n$, take $v \in T_{x_0}\mathcal{M}$. Then there exists some differentiable curve x on $\mathcal{M}$ satisfying $v = x'(0)$. Because x_0 is a local minimum of f on Ω, $s = 0$ is a local minimum of $f(x(s))$, moreover, it is an interior minimum. By chain rule and the necessary condition of local minimum,

$$Df(x(0)) = \nabla f(x(0)) \cdot x'(0) = 0 \tag{3.48}$$

$$\implies \nabla f(x_0) \cdot v = 0 \tag{3.49}$$

Therefore $\nabla f(x_0) \perp T_{x_0}\mathcal{M}$. ■

Theorem 3.8 (Lagrange Multipliers: First Order Necessary Condition)**.** Let $f, h_1, \cdots, h_k \in C^1$ defined on open subset $\Omega \subset \mathbb{R}^n$. Let x_0 be a regular point of the constraint set $\mathcal{M} := \bigcap_{i=1}^k h_i^{-1}(0)$. Suppose x_0 is a local minimum of $\mathcal{M}$, then there exists $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ such that

$$\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) = 0 \tag{3.50}$$

Remark: if we define Lagrangian $\mathcal{L}(x, \lambda_i) := f(x) + \sum_{i=1}^k h_i(x)$, then the theorem says the local minimum is a critical point of $\mathcal{L}$.

Proof. Because x_0 is a regular point, then by previous lemma, $\nabla f(x_0) \perp T_{x_0}\mathcal{M}$. Moreover,

$$T_{x_0}\mathcal{M} = T_{x_0} = (\text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\})^\perp \tag{3.51}$$

Also, because x_0 is a local minimum,

$$\nabla f(x_0) \perp T_{x_0}\mathcal{M} \tag{3.52}$$

Therefore, $\nabla f(x_0) \in (T_{x_0}\mathcal{M})^\perp = (\text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\})^{\perp\perp} = \text{span}\{\nabla h_1(x_0), \cdots, \nabla h_k(x_0)\}$, where the last equality holds in finite dimensional cases. Hence, it is obvious that we can write $\nabla f(x_0)$ as a linear combination of $\{\nabla h_i(x_0)\}$. ■

Theorem 3.9 (Second Order Necessary Condition)**.** Let $f, h_i \in C^2$, if x_0 is a local minimum on previously defined surface $\mathcal{M}$, then there exists Lagrangian multipliers $\{\lambda_i\}$ such that

(i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) = 0$ ($\nabla_x \mathcal{L} = 0$);

(ii) And $\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \succcurlyeq 0$ on $T_{x_0}\mathcal{M}$ ($\nabla_x^2 \mathcal{L} \succcurlyeq 0$).

Remark: whenever x_0 is a local minimum, it must be a critical point of $\mathcal{L}$, and $\mathcal{L}$ is positive semidefinite on the tangent space at x_0.

Proof. The first result is exactly the same as the first order condition proven above.
To show the second result, let $x(s) \in \mathcal{M}$ be an arbitrary differentiable curve on $\mathcal{M}$ such that $x(0) = x_0$. Then,

$$\frac{d}{ds} f(x(s)) = \nabla f(x(s)) \cdot x'(s) \tag{3.53}$$

$$\frac{d^2}{ds^2} f(x(s)) = x'(s)^T \nabla^2 f(x(s)) x'(s) + \nabla f(x(s)) x''(s) \tag{3.54}$$

By the second order Taylor theorem, for every s such that $x(s) \in \mathcal{M}$,

$$f(x(s)) - f(x_0) = s\nabla f(x_0) \cdot x'(0) + \frac{s^2}{2}\left[x'(0)^T \nabla^2 f(x(0)) x'(s) + \nabla f(x(0)) x''(0)\right] + o(s^2) \tag{3.55}$$

Note that by definition, $x'(0)$ is in the tangent space at x_0. Also, we've shown previously that $\nabla f(x_0)$ is orthogonal to the tangent space at x_0, therefore,

$$f(x(s)) - f(x_0) = \frac{s^2}{2}\left[x'(0)^T\nabla^2 f(x(0))x'(s) + \nabla f(x(0))x''(0)\right] + o(s^2) \tag{3.56}$$

Also, by the definition of $\mathcal{M}$, all constraints hold with equality:

$$f(x_0) = f(x_0) + \sum_{i=1}^{k}\lambda_i h_i(x_0) \tag{3.57}$$

where λ_i's are from the first result. Hence,

$$f(x(s)) - f(x_0) = \frac{s^2}{2}\left[x'(0)^T\left(\nabla^2 f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla^2 h_i(x_0)\right)x'(0) + \left(\nabla f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla h_i(x_0)\right)x''(0)\right] + o(s^2) \tag{3.58}$$

$$= \frac{s^2}{2}x'(0)^T\left(\nabla^2 f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla^2 h_i(x_0)\right)x'(0) + o(s^2) \tag{3.59}$$

And above expression is greater or equal to zero because x_0 is a local minimum,

$$\frac{s^2}{2}x'(0)^T\left(\nabla^2 f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla^2 h_i(x_0)\right)x'(0) + o(s^2) \geq 0 \tag{3.60}$$

$$\implies x'(0)^T\left(\nabla^2 f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla^2 h_i(x_0)\right)x'(0) + \frac{o(s^2)}{s^2} \geq 0 \tag{3.61}$$

$$\overset{s\to 0}{\implies} x'(0)^T\left(\nabla^2 f(x_0) + \sum_{i=1}^{k}\lambda_i\nabla^2 h_i(x_0)\right)x'(0) \geq 0 \tag{3.62}$$

Where $x'(0)$ is a vector in the tangent space at x_0 by definition. Moreover, the curve $x(s)$ was chosen arbitrarily, so the argument works for every curve and therefore every tangent vector, and what's desired is shown. ■

Example 3.7.

$$\min f(x,y) = x^2 - y^2 \tag{3.63}$$

$$s.t.\ h(x,y) = y = 0 \tag{3.64}$$

First order condition suggests $(x_0, y_0) = (0, 0)$ Note that the tangent space at (x_0, y_0) is $\text{span}\{\nabla h_i\}^\perp$:

$$T_{x_0}\mathcal{M} = \{(u, 0) : u \in \mathbb{R}\} \tag{3.65}$$

and

$$\nabla_x^2\mathcal{L} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \tag{3.66}$$

is obviously positive semidefinite (actually positive definition) on the tangent space.

Theorem 3.10 (Second Order Sufficient Conditions)**.** Let $f, h_i \in C^2$ on open $\Omega \subset \mathbb{R}^n$, and $x_0 \in \mathcal{M}$ is a regular point, if there exists $\lambda_i \in \mathbb{R}$ such that

(i) $\nabla_x \mathcal{L}(x_0, \lambda_i) = 0$;

(ii) $\nabla_x^2 \mathcal{L}(x_0, \lambda_i) \succ 0$ on $T_{x_0}\mathcal{M}$,

then x_0 is a *strict* local minimum.

Proof. Recall that $\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)$ positive definite on $T_{x_0}\mathcal{M}$ implies there exists $a > 0$ (a is taken to be equal to the least eigenvalue of $\nabla_x^2 \mathcal{L}$) such that

$$v^T[\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)]v \geq a||v||^2 \quad \forall v \in T_{x_0}\mathcal{M} \tag{3.67}$$

Let $x(s) \in \mathcal{M}$ be a curve such that $x(0) = x_0$ and $v = x'(0)$. WLOG, $||x'(0)|| = 1$. By the second order Taylor expansion,

$$\begin{aligned}
f(x(s)) - f(x(0)) &= s \left.\frac{d}{ds}\right|_{s=0} f(x(s)) + \frac{s^2}{2} \left.\frac{d^2}{ds^2}\right|_{s=0} f(x(s)) + o(s^2) && (3.68)\\
&= s \left.\frac{d}{ds}\right|_{s=0} \left[f(x(s)) + \sum \lambda_i h_i(x(s))\right] + \frac{s^2}{2} \left.\frac{d^2}{ds^2}\right|_{s=0} \left[f(x(s)) + \sum \lambda_i h_i(x(s))\right] + o(s^2) && (3.69)\\
&= s\nabla_x \mathcal{L}(x_0, \lambda_i) \cdot x'(0) + \frac{s^2}{2}\left[x'(0)^T \nabla_x^2 \mathcal{L}(x_0, \lambda_i) x'(0) + \nabla_x \mathcal{L}(x_0, \lambda_i) x''(0)\right] + o(s^2) && (3.70)\\
&= \frac{s^2}{2} x'(0)^T \nabla_x^2 \mathcal{L}(x_0, \lambda_i) x'(0) + o(s^2) && (3.71)\\
&\geq \frac{s^2}{2} a||x'(0)||^2 + o(s^2) \quad \text{where } a > 0 && (3.72)\\
&= s^2 \left(\frac{a}{2} + \frac{o(s^2)}{s^2}\right) && (3.73)\\
&\overset{s\to 0}{>} 0 && (3.74)
\end{aligned}$$

Therefore, for sufficiently small s, $f(x(s)) - f(x(0)) > 0$. And this is true for every curve x on $\mathcal{M}$. So $x(0)$ is a strict local minimum. ∎

3.3 Remark on the Connection Between Constrained and Unconstrained Optimizations

Example 3.8. Consider

$$\begin{aligned}
&\min f(x, y, z) && (3.75)\\
&s.t. g(x, y, z) = z - h(x, y) = 0 && (3.76)
\end{aligned}$$

where $\mathcal{M}$ is the graph of h. Using Lagrangian multiplier provides necessary condition: $\nabla f + \lambda \nabla g = 0$,

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} + \lambda \begin{pmatrix} -h_x \\ -h_y \\ 1 \end{pmatrix} = 0 \tag{3.77}$$

Convert the constrained optimization into an unconstrained optimization as

$$\min_{(x,y)\in\mathbb{R}^2} F(x, y) = f(x, y, h(x, y)) \tag{3.78}$$

The necessary condition for unconstrained optimization is

$$\nabla F(x,y) = \begin{pmatrix} f_x + f_z h_x \\ f_y + f_z h_y \end{pmatrix} \tag{3.79}$$

$$= \begin{pmatrix} f_x \\ f_y \end{pmatrix} - f_z \begin{pmatrix} -h_x \\ -h_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.80}$$

Define $\lambda := -f_z$.

$$\nabla F(x,y) = \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} + \lambda \begin{pmatrix} -h_x \\ -h_y \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.81}$$

3.4 Inequality Constraints

Definition 3.9. Let x_0 satisfy the set of constraints

$$h_i(x) = 0 \quad i \in \{1, \cdots, k\} \tag{3.82}$$

$$g_j(x) \leq 0 \quad j \in \{1, \cdots, \ell\} \tag{3.83}$$

we say that the constraint g_i is **active** at x_0 if $g_i(x_0) = 0$, and is **inactive** at x_0 if $g_i(x_0) < 0$.

Definition 3.10. Split the collection of inequality constraints into active and inactive constraints, let $\Theta(x_0)$ denote the collection of active indices, that's:

$$g_j(x_0) = 0 \ \forall j \in \Theta(x_0) \tag{3.84}$$

$$g_j(x_0) < 0 \ \forall j \notin \Theta(x_0) \tag{3.85}$$

Then x_0 is said to be a **regular point** of the constraint if

$$\{\nabla h_i(x_0) \ \forall i \in \{1, \cdots, k\}; \underbrace{\nabla g_j(x_0) \ \forall j \in \Theta(x_0)}_{\text{Active Constraints}}\} \tag{3.86}$$

is linearly independent.

Theorem 3.11 (The First Order Necessary Condition for Local Minimum: Kuhn-Tucker Conditions)**.** Let Ω be an open subset of $\mathbb{R}^n$ with constraints h_i and g_i to be C^1 on Ω. Suppose $x_0 \in \Omega$ is a regular point with respect to constraints, further suppose x_0 is a local minimum, then there exists some $\lambda_i \in \mathbb{R}$ and $\mu_j \in \mathbb{R}_+$ such that

(i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j=1}^\ell \mu_j \nabla g_j(x_0) = 0$;

(ii) $\mu_j g_j(x_0) = 0$ (*Complementary slackness*).

Remark: by complementary slackness, all μ_j corresponding to inactive inequality constraints are zero.

Proof. Let x_0 be a local minimum for f satisfying constraints, equivalently, it is a local minimum for equality constraints and active inequality constraints.
By the first order necessary condition for local minimum with equality constraints, there exists $\lambda_i, \mu_j \in \mathbb{R}$

such that

$$\nabla f(x_0) + \sum_{i=1}^{k} \lambda_i \nabla h_i(x_0) + \sum_{j\in\Theta(x_0)} \mu_j \nabla g_j(x_0) = 0 \tag{3.87}$$

Then by setting $\mu_j = 0$ for all $j \notin \Theta(x_0)$ one have

$$\nabla f(x_0) + \sum_{i=1}^{k} \lambda_i \nabla h_i(x_0) + \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) = 0 \tag{3.88}$$

By construction, the complementary slackness is satisfied. At this stage, we have construct $\lambda_i \in \mathbb{R}$ and $\mu_j \in \mathbb{R}$ satisfying both conditions, we still need to argue that $\mu_j \geq 0$ for every j. ■

Theorem 3.12 (The Second Order Necessary Conditions)**.** Let Ω be an open subset of $\mathbb{R}^n$, and $f, h_1, \cdots, h_k, g_1, \cdots, g_\ell \in C^2$. Let x_0 be a regular point of the constraints:

$$(\dagger) \begin{cases} h_i(x) = 0 & \forall i \\ g_j(x) \leq 0 & \forall j \end{cases} \tag{3.89}$$

Suppose x_0 is a local minimum of f subject to constraint $(\dagger)$, then there exists $\lambda_i \in \mathbb{R}$ and $\mu_j \geq 0$ such that

(i) $\nabla f(x_0) + \sum_{i=1}^{k} \lambda_i \nabla h_i(x_0) + \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) = 0$;

(ii) $\mu_j g_j(x_0) = 0$;

(iii) $\nabla^2 f(x_0) + \sum_{i=1}^{k} \lambda_i \nabla^2 h_i(x_0) + \sum_{j=1}^{\ell} \mu_j \nabla^2 g_j(x_0)$ is positive definite on the tangent space to <u>activate constraints</u> at x_0.

Proof. (i) and (ii) are immediate result from the first order necessary condition.
Suppose x_0 is a local minimum for $(\dagger)$, then x_0 is a local minimum for active constraints at x_0.
Therefore, $\nabla^2 \hat{\mathcal{L}} = \nabla^2 f(x_0) + \sum_{i=1}^{k} \lambda_i \nabla^2 h_i(x_0) + \sum_{j\in I(x_0)} \mu_j \nabla^2 g_j(x_0)$ is positive semidefinite on the tangent space to active constraints. Note that because $\mu_j = 0$ for inactive constraints, therefore $\nabla^2 \hat{\mathcal{L}} = \nabla^2 \mathcal{L}$ at x_0, and both of them are positive semidefinite on the tangent space corresponding to active constraints. ■