Chapter 3

Differentiable Nonlinear Programming

3.1 Unconstrained Problems

The first class of problems we consider in this chapter has the following form

$$\begin{aligned}&\textbf{minimize } f(x)\\ &\textbf{subject to } x \in X\end{aligned} \tag{3.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and X is an open nonempty subset of $\mathbb{R}^n$.

Theorem 3.1.1 ***Let X be an open set in $\mathbb{R}^n$, and let $f : X \longrightarrow \mathbb{R}$. Suppose x_0 is a local minimizer of f. If f has first order partial derivatives at x_0, then***

$$\frac{\partial f}{\partial x_i}(x_0) = 0, \qquad i = 1, 2, \cdots, n.$$

If f is of class $C^{(2)}$ on some open ball $B(x_0, \delta)$ centered at x_0, then the Hessian of f at x_0

$$H(x_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)\right)$$

is positive semidefinite.

Proof **:** Since X is open there exists a $\delta > 0$ such that

$$V = \{x \in \mathbb{R}^n : |x_i - x_{0i}| < \delta, i = 1, \cdots, n\} \subset X.$$

Here $x = (x_1, \cdots, x_n)$, $x_0 = (x_{01}, \cdots, x_{0n})$. Set $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ unit vector in the x_i coordinate direction. Next, set $\varphi_i(t) = f(x_0 + te_i)$, $|t| < \delta$.
Then, $t = 0$ is a local minimizer for $\varphi_i(\cdot)$. Thus, $\varphi_i^{'}(0) = 0$. However, $\varphi_i^{'}(0) = \frac{\partial f}{\partial x_i}(x_0)$.

Next, suppose f is $C^{(2)}$ on V. Let u be a unit vector in $\mathbb{R}^n$. Set

$$\varphi(t; u) = f(x_0 + tu), \qquad |t| < \delta.$$

The function $\varphi(\cdot\,; u)$ is well defined and is of class $C^{(2)}$, and has a local minimum at $t = 0$. Thus,

$$\frac{d}{dt}\varphi(t; u)\,|_{t=0}\, = 0.$$

and

$$\frac{d^2}{dt^2}\varphi(t; u)\,|_{t=0}\, \geq 0.$$

However,

$$\frac{d^2}{dt^2}\varphi(t; u) = \langle u, H(x_0 + tu)u\rangle$$

Thus,

$$\langle u, H(x_0 + tu)u\rangle \geq 0$$

since u is an arbitrary unit vector it follows that H is positive semidefinite. □

In $\mathbb{R}^n$ the points c such that $\nabla f(c) = 0$ are called critical points of f.

Let A be an $n \times n$ matrix with entries a_{ij}. Let $\Delta_1 = a_{11}$ and

$$\Delta_k = \mathbf{det} \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}, \qquad k = 2, 3, 4, \cdots, n$$

The determiants Δ_k are called the principal minors of A. The following lemma gives a criterion for positive definiteness of a symmetric matrix A with entries a_{ij} and is useful in dealing with unconstrained problems where the positive definiteness of the Hessian is an issue.

Lemma 3.1.1 *The symmetric matrix A is positive definite if and only if $\Delta_k > 0$ for $k = 1, 2, \cdots, n$ and is negative definite if and only if $(-1)^k \Delta_k > 0$ for $k = 1, 2, \cdots, n$.*

Proof : Let $Q(x) = x^T Ax$. If $Q(x)$ is positive definite then A has to be nonsingular. Otherwise there exists $x_0 \in \mathbb{R}^n$, $x_0 \neq 0$ such that $Ax_0 = 0$. Then, $Q(x_0) = x_0^T Ax_0 = 0$ and Q would not be positive definite.

If Q is positive semidefinite and A is non singular, then Q must be positive definite. If Q were not positive definite then there exists $x_0 \neq 0$ such that $x \longrightarrow Q(x)$ attains its minimum at x_0. Thus $Q^{'}(x_0) = 0$; i.e. $Ax_0 = 0$ contradicting the fact that A is nosingular.

Next, we show that Q is positive definite implies that $\mathbf{det}(A) > 0$. We note that

$$\langle x, [(1-\lambda)I + \lambda A]\, x\rangle = \lambda\langle x, x\rangle + (1-\lambda)\|x\|^2 > 0$$

if $x \neq 0$ and $0 < \lambda < 1$. Now, set

$$\rho(\lambda) = \mathbf{det}\ [(1-\lambda)I + \lambda A]\,.$$

We note that $\rho(0) = 1$ and $\rho(1) = \mathbf{det}\ (A)$. Since A is nonsingular when Q is positive definite we have det $(A) \neq 0$. Thus, suppose $\rho(1) = \mathbf{det}\ A < 0$. Then, by the intermediate value property there exists $0 < \lambda^* < 1$ such that $\rho(\lambda^*) = 0$. That is, there exists $x^* \neq 0$ such that

$$\langle x^*, [(1-\lambda^*)I + \lambda A]\, x^*\rangle = 0,$$

$$\lambda^*\langle x^*, x^*\rangle + (1-\lambda^*)\|x^*\|^2 = 0,$$

which is not possible. Thus, det $A > 0$.

Next, let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by

$$f(x) = x^T Ax + 2\ \langle b, x\rangle + c.$$

$$f^{'}(x) = 2\ A\ x + 2b = 0$$

therefore $x = -A^{-1}b$. Set $x_0 = -A^{-1}b$.

$$\begin{aligned} f(x_0) &= \langle x_0, Ax_0\rangle + 2\langle b, x_0\rangle + c \\ &= \langle A^{-1}b, AA^{-1}b\rangle - 2\langle b, A^{-1}b\rangle + c \\ &= -\ \langle b, A^{-1}b\rangle + c \\ &= \langle b, x_0\rangle + c \end{aligned}$$

Next,

$$\mathbf{det} \begin{pmatrix} A & b \\ b^T & c \end{pmatrix} = \mathbf{det} \begin{pmatrix} A & Ax_0 + b \\ b^T & \langle b, x_0\rangle + c \end{pmatrix} = \mathbf{det} \begin{pmatrix} A & 0 \\ b^T & \langle b, x_0\rangle + c \end{pmatrix} = \mathbf{det}\ A \cdot (\langle b, x_0\rangle + c)$$

Therefore

$$\langle b, x_0\rangle + c = \frac{\mathbf{det}\begin{pmatrix} A & b \\ b^T & c \end{pmatrix}}{\mathbf{det}\, A}$$

Finally, set

$$\begin{aligned} g(x,y) &= \langle x, Ax\rangle + 2y\langle b, x\rangle + cy^2 \\ &= (x,y)^T\ B\ (x,y) \end{aligned}$$

where

$$B = \begin{pmatrix} A & b \\ b^T & c \end{pmatrix}$$

Now suppose g **is positive definite. Then, by what was proved earlier det**$B > 0$**. Setting** $y = 0$**,** $g(x,0) = \langle x, Ax\rangle$ **showing** A **is positive definite. Conversely, suppose** A **is positive definite and det**$B > 0$**. Then** $g(x,0) = \langle x, Ax\rangle > 0$ **for all** $x \neq 0$**. If** $y \neq 0$**,**

$$g(x,y) = y^2 g(\frac{x}{y}, 1) \geq y^2 f(\frac{x}{y}) \geq y^2 f(x_0)$$

Thus,

$$g(x,y) \geq y^2 \frac{\mathbf{det}\begin{pmatrix} A & b \\ b^T & c \end{pmatrix}}{\mathbf{det}\, A}$$

that is,

$$g(x,y) \geq y^2 \frac{\mathbf{det}\, B}{\mathbf{det}\, A}.$$

We know det $A > 0$ **since** A **is positive definite, and det** $B > 0$ **by assumption. Thus** $g(x,y) > 0$ **if det** $B > 0$ **and** A **is positive definite.**

To prove the lemma, let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

Write A **as**

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \cdots & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix} \\ &= \begin{pmatrix} A_{n-1} & b \\ b^T & a_{nn} \end{pmatrix} \end{aligned}$$

Since A **is positive definite we immediately conclude** A_{n-1} **is also positive definite. The proof follows by induction. In the case** A **is negative definite we apply the argument to** $-A$ **which is positive definite.** □

We state the following theorem without proof.

Theorem 3.1.2 ***Let*** X ***be an open set in*** $\mathbb{R}^n$***. Let*** $f : X \longrightarrow \mathbb{R}$ ***be of class*** $C^{(2)}$ ***on*** $B(x_0;\delta)$ ***for some*** $\delta > 0$***. Suppose*** $\nabla\ f(x_0) = 0$ ***and*** $H(x_0)$ ***is positive definite. Then,*** x_0 ***is a strict local minimizer. If*** X ***is convex,*** f ***is of class*** $C^{(2)}$ ***on*** X***, and*** $H(x)$ ***is positive semidefinite for all*** x ***in*** X***, then*** x_0 ***is a minimizer for*** f***. If*** $H(x)$ ***is positive definite for all*** x ***in*** X***, then*** x_0 ***is a strict minimizer.***

Example 3.1.1

Let $f(x,y) = x^2 + y^3$ **and** $g(x,y) = x^2 + y^2$**. Then** $\nabla\ f(0,0) = \nabla\ g(0,0) = 0$**.** f **and** g **have positive semidefinite Hessians at** $(0,0)$**. However,** $(0,0)$ **is a local minimizer for** g **but not for** f**.**

3.2 Constrained Problems

In this section we derive first order necessary conditions for differentiable nonlinear programming problems (i.e. the objective function and the constraint functions are nonlinear). The problem we consider has the form

$$
\begin{array}{ll}
\min\ f(x) & \\
\textbf{subject to} & \\
g_1(x) \leq 0 & \\
\quad\vdots & \\
g_m(x) \leq 0 & \qquad (P) \\
h_1(x) = 0 & \\
\quad\vdots & \\
h_k(x) = 0 & \\
x \in X_0 \subset \mathbb{R}^n &
\end{array}
$$

In this problem $f : X_0 \subset \mathbb{R}^n \longrightarrow \mathbb{R}$**,** $g_i : X_0 \longrightarrow \mathbb{R}$**,** $i = 1, \cdots, m$ **and** $h_i : X_0 \longrightarrow \mathbb{R}$**,** $i = 1, \cdots, k$**. Further** X_0 **is a nonempty open subset of** $\mathbb{R}^n$**. The functions** $f, g_1, \cdots, g_m; h_1, \cdots, h_k$ **are** $C^{(1)}$ **functions on** X_0**.**

Definition 3.2.1 ***The set*** S ***deined by***

$$S = \{x \in X_0 : g_i(x) \leq 0, \quad i = 1, ..., m\ ; \quad h_j(x) = 0,\ j = 1, ..., k\}$$

is called the feasible or admissible set for (P)***. Each point of the feasible set*** S ***is said to be a feasible point or an admissible point.***

Theorem 3.2.1 ***Let*** x^* ***be a solution of problem*** (P)***. Then, there exists a real number*** $\lambda_0 \geq 0$***, a vector*** $\lambda \geq 0$ ***in*** $\mathbb{R}^m$***, and a vector*** μ ***in*** $\mathbb{R}^k$ ***such that***

(i) $(\lambda_0, \lambda, \mu) \neq 0$***;*** $\lambda = (\lambda_1, \cdots, \lambda_m),\ \mu = (\mu_1, \cdots, \mu_k)$

(ii) $\langle \lambda, g(x^*) \rangle = 0$***; i.e.*** $\lambda_i g_i(x^*) = 0,\ i = 1, \cdots, m$***.***

(iii) $\lambda_0 \nabla\ f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla\ g_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla\ h_i(x^*) = 0$

Remark 3.2.1 ***Theorem 3.2.1 with the additional hypothesis guaranteeing*** $\lambda_0 > 0$ ***was proved by Kuhn and Tucker (1951). The theorem with*** $\lambda_0 \geq 0$ ***had been proved by John (1948). Theorem 2.2.1 was also proved by Karush (1939) under conditions guaranteeing*** $\lambda_0 > 0$***. Theorem 2.2.1 with additional hypothesis guaranteeing that*** $\lambda_0 > 0$ ***is referred to as Karush-Kuhn-Tucker Theorem. Theorem 2.2.1 as stated above is referred to as the Fritz John Theorem, and the multipliers*** λ_0, λ, μ ***are called fritz Jihn multipliers.***

Below we will give the proof of Theorem 2.2.1 based on McShane's (1973) penalty method.
Proof **: (Proof of Theorem 3.2.1)**
We may assume that $x^* = 0$ **and that** $f(x^*) = 0$**. Let**

$$E = \{i : g_i(0) < 0\} \tag{3.2}$$
$$I = \{i : g_i(0) = 0\} \tag{3.3}$$

To simplify the notation we assume that $E = \{1, 2, \cdots, r\}$ **and** $I = \{r + 1, \cdots, m\}$**,** $0 \leq r \leq m$**.**

Let ω **be a** $C^{(1)}$ **function from** $\mathbb{R}$ **to** $[0, \infty)$ **such that (i)** ω **is strictly increasing (ii)** $\omega(t) = 0$ **for** $t \leq 0$**, (iii)** $\omega(t) > 0$ **for** $t > 0$**. We note that** $\omega^{'}(t) > 0$ **for** $t > 0$**.**

Since $g_1, \cdots, g_m$ are continuous on the open set X_0, there exists an $\varepsilon_0 > 0$ such that $B(0;\varepsilon_0) \subset X_0$ and $g_i(x) < 0$ for $x \in B(0;\varepsilon_0)$ and $i \in I$.

Define a penalty function F as follows

$$F(x,K) = f(x) + \|x\|^2 + K\left\{\sum_{i=1}^{m} \omega\ (g_i(x)) + \sum_{i=1}^{k} (h_i(x))^2\right\} \tag{3.4}$$

Now we assert that for each $0 < \varepsilon < \varepsilon_0$ there exist a positive integer $K(\varepsilon)$ such that for any x with $\|x\| = \varepsilon$ we have $F(x, K(\varepsilon)) > 0$. If this assertion were false then there would exist $\varepsilon^{'}$, $0 < \varepsilon^{'} < \varepsilon_0$ such that for each positive integer K there exists a point x_K with $\|x_K\| = \varepsilon^{'}$ and $F(x_K, K) \le 0$. Thus,

$$f(x_K) + \|x_K\|^2 \le -K\ \left\{\sum_{i=1}^{m} \omega\ (g_i(x_K)) + \sum_{i=1}^{k} (h_i(x_K))^2\right\} \tag{3.5}$$

There exist positive integers $K_1 < K_2 < \cdots < K_n < \cdots \uparrow \infty$ and $x_{\varepsilon'}$ $\|x_{\varepsilon'}\| = \varepsilon^{'}$ such that $x_{K_n} \longrightarrow x_{\varepsilon'}$. Since $f, g_1, \cdots, g_m, h_1, \cdots, h_k$ are all continuous it follows from (3.5) that

$$\sum_{i=1}^{m} \omega\ (g_i(x_{\varepsilon'})) + \sum_{i=1}^{k} (h_i(x_{\varepsilon'}))^2 = 0 \tag{3.6}$$

Thus,

$$g_i(x_{\varepsilon'}) \le 0, \qquad 1 \le i \le m \tag{3.7}$$
$$h_i(x_{\varepsilon'}) = 0, \qquad 1 \le i \le k \tag{3.8}$$

From (3.5) we also have

$$f(x_{\varepsilon'}) + \varepsilon^{'2} \le 0 \tag{3.9}$$

From (3.7)–(3.9) we see that $x_{\varepsilon'}$ is feasible with $f(x_{\varepsilon'}) \le \varepsilon^{'2} < f(0)$ a contradiction. Thus, there exist a positive integer K_ε such that $F(x, K_\varepsilon) > 0$ on $\{x : \|x\| = \varepsilon\}$.

Now, if we consider the problem of minimizing the function $F(\cdot, K_\varepsilon)$ on the set $\{x : \|x\| \le \varepsilon\}$, then $F(\cdot, K_\varepsilon)$ attains its minimum at a point $x(\varepsilon)$, $\|x(\varepsilon)\| < \varepsilon$. Thus,

$$\frac{d}{d\,\varepsilon}\ F(x, K_\varepsilon)|_{x=x(\varepsilon)} = 0 \tag{3.10}$$

That is,

$$\frac{\partial\, f}{\partial\, x_j}\,(x(\varepsilon)) + 2\,x_j(\varepsilon) + \sum_{i=1}^{m} K_\varepsilon\ \omega^{'}\ (g_i\,(x(\varepsilon)))\,\frac{\partial\, g_i}{\partial\, x_j}\,(x(\varepsilon)) + \sum_{i=1}^{k} 2\ K_\varepsilon\ \ (h_i\,(x(\varepsilon)))\,\frac{\partial\, h_i}{\partial\, x_j}\,(x(\varepsilon)) = 0, \quad j = 1, \cdots, n \tag{3.11}$$

Define

$$M(\varepsilon) = 1 + \sum_{i=1}^{m}\left[K_\varepsilon\ \omega^{'}\ (g_i\,(x(\varepsilon)))\right]^2 + \sum_{i=1}^{k}\left[2\ K_\varepsilon\ \ (h_i\,(x(\varepsilon)))\right]^2 \tag{3.12}$$

Set

$$\begin{aligned} \lambda_0(\varepsilon) &= \frac{1}{\sqrt{M(\varepsilon)}}, \\ \lambda_i(\varepsilon) &= \frac{K_\varepsilon\ \omega^{'}\ (g_i\,(x(\varepsilon)))}{\sqrt{M(\varepsilon)}} \ge 0, \quad i = 1, \cdots, m \end{aligned} \tag{3.13}$$

$$\mu_i(\varepsilon) = \frac{2\ K_\varepsilon\ \ (h_i\,(x(\varepsilon)))}{\sqrt{M(\varepsilon)}} \tag{3.14}$$

Since $0 < \varepsilon < \varepsilon_0$ and $\|x(\varepsilon)\| < \varepsilon < \varepsilon_0$ we note that

$$\lambda_i(\varepsilon) = 0, \quad i = r+1, \cdots, m \tag{3.15}$$

Let

$$\lambda_0 = \frac{1}{\lim_{\varepsilon \to 0^+} \sqrt{M(\varepsilon)}} \tag{3.16}$$

Take an appropriate subsequence $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_n > \cdots$ converging to zero so that

$$\begin{aligned} \lambda_0(\varepsilon_j) &\to \lambda_0 \geq 0 & (3.17)\\ \lambda_i(\varepsilon_j) &\to \lambda_i \geq 0, \quad 1 \leq i \leq m & (3.18)\\ \mu_i(\varepsilon_j) &\to \mu_i, \quad 1 \leq i \leq k \end{aligned}$$

Then, from (3.11) we obtain

$$\lambda_0 \frac{\partial f}{\partial x_j}(0) + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j}(0) + \sum_{i=1}^{k} \mu_i \frac{\partial h_i}{\partial x_j}(0) = 0 \tag{3.19}$$

Since

$$\lambda_0^2(\varepsilon_j) + \sum_{i=1}^{m} \lambda_i^2(\varepsilon_j) + \sum_{i=1}^{k} \mu_i^2(\varepsilon_j) = 1,$$

we have

$$\lambda_0 + \sum_{i=1}^{m} |\lambda_i| + \sum_{i=1}^{k} |\mu_i| \neq 0 \tag{3.20}$$

From (3.14) we have

$$\lambda_i = 0, \qquad i \leq r+1, \cdots, m \tag{3.21}$$

Thus,

$$\lambda_i \; g_i(0) = 0, \qquad i = 1, \cdots, m \tag{3.22}$$

From (3.17) – (3.22) we see that the proof is complete. □

Definition 3.2.2 ***In problem (P) the functions $g_1, \cdots, g_m$; $h_1, \cdots, h_k$ satisfy the constraint qualification at a feasible point x_0 if***

(i) ***the vectors $\nabla h_1(x_0), \cdots, \nabla h_k(x_0)$ are linearly independent and***

(ii) ***the system***

$$\begin{cases} \nabla g_i(x_0) \cdot z < 0, & i \in E = \{\ell : g_\ell(x_0) = 0\} \\ \nabla h_i(x_0) \cdot z = 0, & i = 1, \cdots, k \end{cases}$$

has a solution z in $\mathbb{R}^n$.

If the equality constraints are absent, then the statements involving $\nabla h_i(x_0)$, $1 \leq i \leq m$ are to be deleted.

Lemma 3.2.1 ***Let x^* be a solution of problem (P). Let the constraint qualification hold at x^*. Then, λ_0 in Theorem 2.2.1 is strictly positive, that is $\lambda_0 > 0$.***

Proof **: Suppose $\lambda_0 = 0$. Then, from (iii) of Theorem 2.2.1**

$$\sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla h_i(x^*) = 0 \tag{3.23}$$

If $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$, then from (3.23) and from the fact that $\nabla h_1(x^*), \cdots, \nabla h_k(x^*)$ are linearly independent $\mu_i = 0, \quad i = 1, \cdots, m$. However, from (i) of Theorem 2.2.1 $(\lambda_0, \lambda, \mu) \neq 0$. thus, we can't have $\lambda_1, \cdots, \lambda_m$ all zero. In particular $\lambda_i \neq 0$ for all $i \in E = \{\ell : g_\ell(x^*) = 0\}$. Thus,

$$\sum_{i \in E} \lambda_i \cdot \nabla g_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla h_i(x^*) = 0 \tag{3.24}$$

for any vector $z \in \mathbb{R}^n$ as in (ii) of Definition 2.2.1, we obtain from (3.24)

$$\sum_{i \in E} \lambda_i \nabla g_i(x^*)^T \cdot z = 0 \tag{3.25}$$

since

$$\lambda_i \geq 0, \quad \sum_{i \in E} \lambda_i \neq 0 \text{ and } \nabla g_i(x^*)^T \cdot z < 0, \quad i \in E$$

we have a contradiction. Thus $\lambda_0 > 0$. □

Theorem 3.2.2 ***Let x^* be a solution of problem (P). Let the constraint qualification hold at x^*. Then, there exist scalars $\lambda_1 \geq 0, \lambda_2 \geq 0, \cdots, \lambda_m \geq 0$, $\mu_1, \mu_2, \cdots, \mu_k$ such that***

(i) $\lambda_i g_i(x^*) = 0$, $i = 1, \cdots, m$

(ii) $\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^k \mu_i \nabla h_i(x^*) = 0$

Proof **: By Lemma 2.2.1 in Theorem 2.2.1 $\lambda_0 > 0$. Dividing $(\lambda_0, \lambda, \mu)$ in Theorem 2.2.1 by λ_0 we obtain the KKT conditions. Let**

$$E = \{\ell : g_\ell(x^*) = 0, 1 \leq \ell \leq m\} \tag{3.26}$$

$$I = \{\ell : g_\ell(x^*) < 0, 1 \leq \ell \leq m\} \tag{3.27}$$

□

The necessary conditions (i) and (ii) are referred to as the Karush-Kuhn-Tucker (KKT) conditions, and the multipliers $\lambda_1, \lambda_2, \cdots, \lambda_m$, $\mu_1, \mu_2, \cdots, \mu_k$ are referred to as the Karush-Kuhn-Tucker multipliers.

Theorem 3.2.3 ***Let x^* be a solution of problem (P). Suppose the vectors in the set***

$$\{\nabla h_i(x^*) : i = 1, \cdots, k\} \cup \{\nabla g_i(x^*) : i \in E\}$$

are linearly independent. Then, the conclusions of Theorem 2.2.2 hold.

Proof **: It suffices to prove that λ_0 in Theorem 2.2.1 is strictly positive in this case. in fact, if $\lambda_0 = 0$, then (ii) and (iii) of Theorem 2.2.1 give**

$$\sum_{i \in E} \lambda_i \cdot \nabla g_i(x^*) + \sum_{i=1}^{k} \mu_i \cdot \nabla h_i(x^*) = 0. \tag{3.28}$$

By hypothesis of the theorem equation (3.28) $\mu_1 = \cdots = \mu_k = 0$, $\lambda_i = 0, i \in E$. Then, we contradict (i) of Theorem 2.2.1 and the proof is complete. □

The Karush-Kuhn-tucker necessary coditions are also sufficient for problem (P) where $f, g_1, \cdots, g_m$ are convex and $h_1 = \cdots = h_k = 0$.

Theorem 3.2.4 ***Let*** $f, g_1, \cdots, g_m$ ***be*** $C^{(1)}$ ***functions that are convex and defined on the nonempty convex set*** $X_0 \subset \mathbb{R}^n$***. Consider the problem***

$$\begin{array}{ll} \min\ f(x) & \\ \textbf{\textit{subject to}} & \\ g_1(x) \leq 0 & \\ \quad\vdots & \\ g_m(x) \leq 0 & (P_c) \\ x \in X_0 \subset \mathbb{R}^n & \end{array}$$

Let $x_* \in X_0$***,*** $\lambda_1, \lambda_2, \cdots, \lambda_m$ ***scalars such that***

(i) $g_i(x_*) \leq 0, \qquad i = 1, \cdots, m$

(ii) $\lambda_i \geq 0, \qquad i = 1, \cdots, m$

(iii) $\sum_{i=1}^m \lambda_i g_i(x_*) = 0$

(iv) $\nabla f(x_*) + \sum_{i=1}^m \lambda_i g_i(x_*) = 0$

Then, x_* ***is a solution to problem*** (P_c)***.***

Proof **: Let**

$$X_i = \{x \in X_0 : g_i(x) \leq 0\}, \quad i = 1, \cdots, m.$$

Then, X_i **is convex for** $i = 1, \cdots, m$**. Set**

$$X = \cap_{i=1}^m X_i.$$

Then, X **is convex and nonempty. From (i)** $x_* \in X$**. Since** f **is covex,**

$$f(x) - f(x_*) \geq \langle \nabla f(x_*), x - x_* \rangle, \qquad x \in X$$

Now, by (iv)

$$\begin{aligned} f(x) - f(x_*) &\geq -\sum_{i=1}^m \langle \lambda_i \nabla g_i(x_*), x - x_* \rangle \\ &= -\sum_{i=1}^m \lambda_i \langle \nabla g_i(x_*), x - x_* \rangle \end{aligned} \tag{3.29}$$

Since g_i **is convex for** $i = 1, \cdots, m$ **we have**

$$g_i(x) - g_i(x_*) \geq \langle \nabla g_i(x_*), x - x_* \rangle$$

Thus, from (3.29) we have

$$\begin{aligned} f(x) - f(x_*) &\geq -\sum_{i=1}^m \lambda_i \left(g_i(x) - g_i(x_*) \right) \\ &= \sum_{i=1}^m \lambda_i g_i(x_*) - \sum_{i=1}^m \lambda_i g_i(x) \\ &= 0 - \sum_{i=1}^m \lambda_i g_i(x) \\ &= \sum_{i=1}^m (-\lambda_i) g_i(x_*) \\ &\geq 0 \end{aligned}$$

thus, $f(x) \geq f(x_*)$**.** □

Corollary 3.2.1 ***Let*** $f, g_1, \cdots, g_m$ ***be*** $C^{(1)}$ ***functions defined on the nonempty convex set*** $X_0 \subset \mathbb{R}^n$***. Let*** $h_i, i = 1, \cdots, \ell$ ***be an affine map from*** $\mathbb{R}^n$ ***to*** $\mathbb{R}$***. Consider the problem***

$$\begin{array}{lll} \min\ f(x) & & \\ \textbf{\textit{subject to}} & & \\ g_i(x) \le 0 & i = 1, \cdots, m & \\ h_i(x) = 0, & i = 1, \cdots, \ell & (P_{c'}) \\ x \in X_0 \subset \mathbb{R}^n & & \end{array}$$

Let $x_* \in X_0$***,*** $\lambda_1, \lambda_2, \cdots, \lambda_m$ ***scalars such that***

(i) $g_i(x_*) \le 0, \qquad i = 1, \cdots, m$ ***and*** $h_i(x_*) = 0, \qquad i = 1, \cdots, \ell$

(ii) $\lambda_i \ge 0, \qquad i = 1, \cdots, m$

(iii) $\sum_{i=1}^m \lambda_i g_i(x_*) = 0$

(iv) $\nabla f(x_*) + \sum_{i=1}^m \lambda_i g_i(x_*) = 0$

Then, x_* ***is a solution to problem*** $(P_{c'})$***.***

3.3 Second Order Conditions

In what follows we refer to problem (P) **of Section 2.2 for our discussion.**

Definition 3.3.1 ***Let*** $x_0 \in X$***. The vector*** $z \in \mathbb{R}^n$, $z \ne 0$ ***is said to satisfy the*** *tangential constraints at* x_0 ***if***

$$\nabla g_E(x_0)^T \cdot z \le 0, \qquad \nabla h(x_0)^T \cdot z = 0 \tag{3.30}$$

where

$$E = \{i : g_i(x_0) = 0\}$$

Notation 3.3.1 ***In (3.30)*** $\nabla g_E(x_0)^T \cdot z \le 0$ ***means for every*** i ***such that*** $g_i(x_0) = 0$ ***we have*** $\nabla g_i(x_0)^T \cdot z \le 0$***. Next*** $\nabla h(x_0)^T \cdot z = 0$ ***means*** $\nabla h_i^T(x_0) \cdot z = 0$ ***for*** $i = 1, \cdots, \ell$***.***

For definiteness, in (3.30), we shall henceforth suppose that $E = \{1, \cdots, r\}$**. Let** $I = \{r + 1, \cdots, m\}$**. Since either** E **or** I **can be empty,** $0 \le r \le m$**.**

Let z **be a vector that satisfies the tangential constraints. Let**

$$\begin{aligned} E_1 &= \{i : i \in E, \nabla g_i(x_0)^T \cdot z = 0\} \\ E_2 &= \{i : i \in E, \nabla g_i(x_0)^T \cdot z < 0\} \end{aligned}$$

Then,

$$\begin{aligned} E &= E_1 \cup E_2 \\ \nabla g_{E_1}(x_0)^T \cdot z &= 0, \qquad \nabla h(x_0)^T \cdot z = 0 \end{aligned} \tag{3.31}$$

and

$$\nabla g_{E_2}(x_0)^T \cdot z < 0 \tag{3.32}$$

For the sake of definiteness suppose $E_1 = \{1, \cdots, q\}$**. Since** E_1 **may be empty,** $0 \le q \le r$**. If** $q = 0$**, statements involving** E_1**, in (3.31) in the ensuing discussion should be deleted. We should keep in mind that the sets** E_1 **and** E_2 **depend on** x_0 **and** z**.**

Let z **be a given vector that satisfies the tangential constraints.In the next lemma is presented a condition under which how one can construct a curve** $\xi(\cdot)$ **emanating from** x_0 **and going into the feasible set in the direction given by** z**.**

Lemma 3.3.1 ***Let*** $g_1, \cdots, g_m$***,*** $h_1, \cdots, h_k$ ***and*** X_0 ***be as in problem*** (P)***. Assume*** $g_1, \cdots, g_m$***,*** $h_1, \cdots, h_k$ ***are of class*** $C^(p)$***,*** $p \geq 1$***. Let*** x_0 ***be feasible and let*** $z \in \mathbb{R}^n$ ***satisfy the tangential constraints at*** x_0***. Suppose that the vectors***

$$\nabla g_1(x_0), \cdots, \nabla g_q(x_0), \nabla h_1(x_0), \cdots, \nabla h_k(x_0) \tag{3.33}$$

are linearly independent. Then, there exists $\tau > 0$ ***and a*** $C^{(p)}$ ***mapping*** $\xi(\cdot)$ ***from*** $(-\tau, \tau)$ ***to*** $\mathbb{R}^n$ ***such that***

$$\begin{aligned} \xi(0) &= x_0, \quad \xi'(0) = -z \\ g_{E_1}(\xi(t)) &= 0, \quad h(\xi(t)) = 0, \quad g_I(\xi(t)) < 0 \quad g_{E_2}(\xi(t)) < 0 \ \textit{\textbf{for}}\ 0 < t < \tau \end{aligned}$$

Proof **: Let $\rho = q + k$. In view of (3.33), $\rho \leq n$. If $\rho = n$, then the only solution of (3.31) is $z = 0$. Hence, $E_2 = \emptyset$ and $E_1 = E$. The mapping $\xi(\cdot)$ is then the constant mapping $\xi(k) = x_0$.**

We now suppose that $\rho < n$. Let Γ denote the $\rho \times n$ matrix whose rows are the vectors in (3.33). Then

$$\Gamma = \begin{pmatrix} \nabla g_{E_1}(x_0) \\ \nabla h(x_0) \end{pmatrix}$$

and Γ has rank ρ. By relabeling the coordinates corresponding to ρ linearly independent columns of Γ, we may suppose without any loss of generality that the first ρ columns of the matrix Γ are linearly independent. Thus, the $\rho \times \rho$ matrix

$$\Gamma_\rho = \begin{pmatrix} \frac{\partial g_i(x_0)}{\partial x_j} \\ \frac{h_s(x_0)}{\partial x_j} \end{pmatrix}, \quad i = 1, \cdots, q, \quad s = 1, \cdots, k, \quad j = 1, \cdots, \rho$$

is nonsingular.

The system of equations

$$\begin{aligned} g_{E_1}(x) &= 0 \\ h(x) &= 0 \\ x_{\rho+1} - x_{0,\rho+1} - t\, z_{\rho+1} &= 0 \\ x_n - x_{0,n} - t\, z_n &= 0 \end{aligned} \tag{3.34}$$

where $z_{\rho+1}, \cdots, z_n$ are the last $n-\rho$ components of the vector z, has a solution $(x, t) = (x_0, 0)$. Now, using the implicit function theorem we can assert that there exists a $\tau > 0$ and a C^ρ function $\xi(\cdot)$ defined on $(-\tau, \tau)$ with range in $\mathbb{R}^n$ such that

$$\begin{aligned} \xi(0) &= x_0 \\ g_{E_1}(\xi(t)) &= 0, \qquad h(\xi(t)) = 0 \\ \xi_i(t) &= x_{0,i} + t\, z_i, \quad i = \rho + 1, \cdots, n \end{aligned} \tag{3.35}$$

Differentiating (3.35) with respect to t we have

$$\begin{pmatrix} & \Gamma & \\ O_{n-\rho,\rho} & & I_{n-\rho} \end{pmatrix} \xi'(0) = \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix}$$

where $\hat{z} = (z_{\rho+1}, \cdots, z_n)$. The system

$$\begin{pmatrix} & \Gamma & \\ O_{n-\rho,\rho} & & I_{n-\rho} \end{pmatrix} w = \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix}$$

has a unique solution. Since $\xi'(0)$ and z are both solutions, we get $\xi'(0) = z$.

Since $g_I(x_0) < 0$ and $\xi(0) = x_0$, from the continuity of g and ξ it follows that for τ sufficiently small we have $g_I(\xi(t)) < 0$ for $|t| \leq \tau$.

Since

$$\frac{d\, g_{E_2}(\xi(t))}{dt} = \nabla g_{E_2}(\xi(t))\, \xi^{'}(t)$$

and since $\xi(0) = x_0$ and $\xi^{'}(0) = z$, it follows from (3.32) that $\frac{d\, g_{E_2}(\xi(0))}{dt} < 0$. It then follows from continuity that by taking τ to be smaller, if necessary, there is an interval $[0, \tau)$ on which all of the previous conclusions hold and on which $\frac{d\, g_{E_2}(\xi(0))}{dt} < 0$. Since $g_{E_2}(\xi(0)) = g_{E_2}(\xi(x_0)) = 0$, we have $g_{E_2}(\xi(t)) < 0$ on $[0, \tau)$. $\square$

Corollary 3.3.1 ***Let the constraint qualification hold at x_0 in problem (P). Then for every vector z satisfying (ii) of definition 1.2.2 there exists a $C^{(\rho)}$ function $\xi(\cdot)$ such that $\xi(0) = x_0$, $\xi^{'}(0) = z$ and $\xi(t)$ is feasible for all t in some interval $[0, \tau)$. Moreover, $g(\xi(t)) < 0$ on $(0, \tau)$.***

Proof **: A vector z satisfying (ii) of Definition 1.2.2 satisfies the tangential constraints, and the set of indices E_1 corresponding to z is empty.** $\square$

Definition 3.3.2 ***Let (x^*, λ, μ) be as in Theorem 1.2.2. Let z be a vector that satisfies the tangential constraint (3.30) at x. Let***

$$E_1 = E_1(z) = \{i : i \in E, \quad \nabla\, g_i(x^*)^T \cdot z = 0\},$$

$$E_2 = E_2(z) = \{i : i \in E, \quad \nabla\, g_i(x^*)^T \cdot z < 0\}$$

The vector z will be called a second order test vector if

(i) ***$\lambda_i = 0$ for $i \in E_2$ and***

(ii) ***the rows of the matrix***

$$\begin{pmatrix} \nabla g_{E_1}(x^*) \\ \nabla h(x^*) \end{pmatrix}$$

are linearly independent.

Lemma 3.3.2 ***Let φ be of class $C^{(2)}$ on $(-\tau, \tau)$ such that $\varphi^{'}(0) = 0$ and $\varphi(t) \geq \varphi(0)$ for all t in $[0, \tau)$. Then $\varphi^{''}(0) \geq 0$.***

Proof **: By Taylor's Theorem, for $0 < t < \tau$**

$$\varphi(t) - \varphi(0) = \frac{1}{2}\varphi^{''}(\theta)t^2,$$

where $0 < \theta < 1$. If $\varphi^{''}(0)$ were negative, then by continuity there would exist an interval $[0, \tau^{'})$ on which $\varphi^{''}$ would be negative. Hence $\varphi(t) < \varphi(0)$ on this interval, contradicting the hypothesis. $\square$

Theorem 3.3.1 ***Let f, $g = (g_1, \cdots, g_m)$, $h = (h_1, \cdots, h_k)$ be of class $C^{(2)}$ on X_0. Let x^* be a solution to problem (P). Let λ, μ be KKT multipliers as in Theorem 1.2.2 and let $F(\cdot, \lambda, \mu)$ be defined by***

$$F(x, \lambda, \mu) = f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle \tag{3.36}$$

Then, for every second-order test vector z,

$$\langle z, F_{xx}(x^*, \lambda, \mu) z \rangle \geq 0 \tag{3.37}$$

where

$$F_{xx} = \left(\frac{\partial^2 F}{\partial x_j \partial x_i} \right), \quad i = 1, \cdots, n; \quad j = 1, \cdots, n$$

Proof **: Let z be a second-order test vector. Since $\lambda_i = 0$ for $i \in E_2$ and**

$$\lambda = (\lambda_E, \lambda_I) = (\lambda_{E_1}, \lambda_{E_2}, \lambda_I) = (\lambda_{E_1}, 0, 0)$$

we have

$$F(x, \lambda, \mu) = f(x) + \langle \lambda_{E_1}, g_{E_1}(x) \rangle + \langle \mu, h(x) \rangle$$

Let $\xi(\cdot)$ be the function corresponding to z as in Lemma 2.3.1, with $x_0 = x^*$. Since $g_{E_1}(\xi(t)) = 0$ and $h(\xi(t)) = 0$ for $|t| < \tau$, we have

$$f(\xi(\tau)) = F(\xi(t), \lambda, \mu) \text{ for } |t| < \tau.$$

Since for $0 \leq t < \tau$, all points $\xi(t)$ are feasible, and since $\xi(0) = x^*$ the mapping $t \longrightarrow f(\xi(t))$, where $0 \leq t < \tau$ has a minimum at $t = 0$. Hence, so does the mapping $\varphi(\cdot)$ defined on $[0, \tau)$ by

$$\varphi(t) = F(\xi(t), \lambda, \mu).$$

Now,

$$\varphi^{'}(t) = \langle \nabla F(\xi(t), \lambda, \mu), \xi^{'}(t) \rangle$$

$$\varphi^{''}(t) = \langle \xi^{'}(t), F_{xx}(\xi(t), \lambda, \mu), \xi^{''}(t) \rangle + \langle \nabla F(\xi(t), \lambda, \mu), \xi^{''}(t) \rangle$$

Setting $t = 0$ in the first equation and using $\xi(0) = x^*$ and $\nabla F(x^*, \lambda, \mu) = 0$ we have $\varphi^{'}(0) = 0$. Now, by Lemma 2.3.2 we have $\varphi^{''}(0) \geq 0$. Setting $t = 0$ in $\varphi^{''}(t)$ and using $\xi(0) = x^*$, $\nabla F(x^*, \lambda, \mu) = 0$, and $\xi^{'}(0) = z$ give the conclusion of the theorem. □

Corollary 3.3.2 ***Let the function f, g, h be as in Theorem 2.3.1 and let x^* be a solution to problem (P). Suppose the vectors $\nabla g_1(x^*), \cdots, \nabla g_r(x^*)$, $\nabla h_1(x^*), \cdots, \nabla h_k(x^*)$ are linearly independent. Then, there exist KKT multipliers (λ, μ) as in Theorem 1.2.2, and (3.37) holds for every vector z satisfying***

$$\nabla g_E(x^*)^T \cdot z = 0, \quad \nabla h(x^*)^T \cdot z = 0 \tag{3.38}$$

Corollary 3.3.3 ***Let f and $h = (h_1, \cdots, h_k)$ be as in Theorem 2.3.1, let x^* be a solution of problem (P) and let the constraint qualification (Definition 1.2.2) hold at x^*. Then, there exists a unique μ in $\mathbb{R}^k$, $\mu = (\mu_1, \cdots, \mu_k)$ such that the function $H(\cdot, \mu)$ defined by***

$$H(x, \mu) = f(x) + \langle \mu, h(x) \rangle$$

satisfies

(i) $\nabla f(x^*) + \mu^T \cdot \nabla h(x^*) = 0$

(ii) $\langle z, H_{xx}(x^*, \mu)\, z \rangle \geq 0$

for all z in $\mathbb{R}^n$ satisfying $\nabla h(x^*)^T \cdot z = 0$

3.4 Sufficient Conditions

In this section we give some necessary theorems for the existence of a solution of problem (P).

Theorem 3.4.1 ***Let $(x^*, \lambda_0, \lambda, \mu)$ satisfy the Fritz John necessary conditions of Theorem 2.2.1. Suppose that every $z \neq 0$ that satisfies the tangential constraints (3.30) at $x_0 = x^*$ and the inequality $\nabla f(x^*)^T \cdot z \leq 0$ also satisfies***

$$\langle z, F^0_{xx}(x^*, \lambda, \mu)\, z \rangle > 0 \tag{3.39}$$

where

$$F^0_{xx}(x^*, \lambda, \mu) = \lambda_0 f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle.$$

Then, f in problem (P) attains a strict local minimum from problem at x^*.

Proof **: To simplify notation we assume that $x^* = 0$ and that $f(0) = 0$.**

If f did not attain a strict local minimum at $x^* = 0$, then for every positive integer q there would exist a feasible point $v_q \neq 0$ such that $v_q \in B(0; \frac{1}{q})$ and $f(v_q) \leq 0$. Thus there would exist a sequence of points $\{v_q\}$ in X such that

$$v_q \neq 0, \quad \lim_{q\to\infty} v_q = 0, \quad f(v_q) \leq 0, \quad g(v_q) \leq 0, \quad h(v_q) = 0$$

Using the mean value theorem there exist v_0, v_i, v_j in $(0,1)$ such that

$$f(v_q) = f(0) + \langle \nabla f(v_0, v_q), v_q \rangle = \langle \nabla f(v_0 v_q), v_q \rangle$$

Since $f(0) = 0$ and $f(v_q) \leq 0$ we have

$$f(v_q) = \langle \nabla f(v_0 v_q), v_q \rangle \leq 0 \tag{3.40}$$

Recalling $E = \{1, \cdots, v\}$ is such that $g_i(x^*) = 0$, $i \in E$, we have by the mean value theorem again

$$g_i(v_q) = \langle \nabla g_i(v_i v_q), v_q \rangle \leq 0 \tag{3.41}$$

we also have

$$h_j(v_q) = \langle \nabla h_j(v_j v_q), v_q \rangle \leq 0 \tag{3.42}$$

it also follows from Taylor's Theorem that

$$f(v_q) = \langle \nabla f(0), v_q \rangle + \frac{1}{2}\langle v_q, f_{xx}(\theta_0 v_q) v_q \rangle \leq 0 \tag{3.43}$$

and that for $i = 1, \cdots, r$ and $r = 1, \cdots, k$

$$g_i(v_q) = \langle \nabla g_i(0), v_q \rangle + \frac{1}{2}\langle v_q, g_{i\,xx}(\theta_i v_q) v_q \rangle \leq 0 \tag{3.44}$$

$$h_j(v_q) = \langle \nabla h_j(0), v_q \rangle + \frac{1}{2}\langle v_q, h_{j\,xx}(\theta_j v_q) v_q \rangle \leq 0 \tag{3.45}$$

where $\theta_0, \theta_i, \theta_j$ in (3.43), (3.44) and (3.45) are in $(0,1)$.

Since $v_q \neq 0$, $\frac{v_q}{\|v_q\|}$ is a unit vector. Hence there exists a unit vector v and a subsequence that we relabel as $\{v_q\}$ such that

$$\frac{v_q}{\|v_q\|} \to v. \tag{3.46}$$

Now from (3.40)−(3.42) we have

$$\begin{aligned} \langle \nabla f(0), v \rangle &\leq 0 \\ \nabla g_E(0) \cdot v &\leq 0 \\ \nabla h(0) \cdot v &= 0. \end{aligned}$$

Thus, v satisfies the tangential constraints (see Definition 2.3.1).

Now, multiplying (3.43) by λ_0, (3.44) by λ_i, and (3.45) by μ_j and recalling that $\lambda_i = 0, \quad i = r+1, \cdots, m$ we obtain

$$\langle \lambda_0 \nabla f(0) + \lambda^T \nabla g(0) + \mu^T \nabla h(0), v_q \rangle + \frac{1}{2}\langle v_q, F^0_{xx}(\theta v_q, v_q) \rangle \leq 0 \tag{3.47}$$

We devide the inequality in (3.47) by $\|v_q\|$ and letting $q \to \infty$ through an appropriate subsequence as in (3.46) we have

$$\langle \lambda_0 \nabla f(0) + \lambda^T \nabla g(0) + \mu^T \nabla h(0), v \rangle + \frac{1}{2}\langle v, F^0_{xx}(0, v) \rangle \leq 0 \tag{3.48}$$

Since $\langle \nabla f(0), v \rangle \leq 0$, $\langle \nabla g_E(0), v \rangle \leq 0$, and $\langle \nabla h(0), v \rangle \leq 0$ and $\lambda_i = 0, \quad i = r+1, \cdots, m$ we have from (3.48) the inequality

$$\langle v, F^0_{xx}(0, v) \rangle \leq 0, \tag{3.49}$$

contradicting the assumption of the theorem. □

Corollary 3.4.1 ***Let (x^*, λ, μ) satisfy the KKT necessary conditions. Suppose that every $z \neq 0$ in $\mathbb{R}^n$ that satisfies the tangential constraints at $x_0 = x^*$ and the inequality $\langle \nabla f(x^*), z \rangle = 0$ also satisfies***

$$\langle z, F_{xx}(x^*, \lambda, \mu) z \rangle > 0$$

where

$$F(x, \lambda, \mu) = f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle.$$

Then, f attains a strict local minimum for problem (P) at x^*.

Proof **: If (x^*, λ, μ) satisfy the KKT conditions, then**

$$\nabla f(x^*) + \lambda^T \nabla g(x^*) + \mu^T \nabla h(x^*) = 0$$

Since $\lambda_I = 0$,

$$\nabla f(x^*) = -\lambda_E^T \nabla g(x^*) - \mu^T \nabla h(x^*)$$

For every $z \in \mathbb{R}^n$ we have

$$\langle \nabla f(x^*), z \rangle = -\langle \lambda_E, \nabla g(x^*) z \rangle - \langle \mu, \nabla h(x^*) z \rangle \tag{3.50}$$

Since $\lambda_E \geq 0$ for every z that satisfies the tangential constraints we have from (3.50)

$$\langle \nabla f(x^*), z \rangle \geq 0.$$

Hence, the condition $\nabla f(x^*)^T \cdot z \leq 0$ in Theorem 2.4.1 reduces to

$$\langle \nabla f(x^*), z \rangle = 0$$

□

Remark 3.4.1 ***If the inequality constraits in problem (P) are absent, then a vector z satisfies the tangential constraints at a point x_0 if and only if $\nabla h(x_0)T \cdot z = 0$. Now, if the necessary conditions of Theorem 1.2.2 hold at x^* then, from (ii) of Theorem 1.2.2 we have***

$$\nabla f(x^*) = -\sum_{i=1}^{k} \mu_i \nabla h_i(x^*)$$

and thus,

$$\nabla f(x^*)^T \cdot z = -\sum_{i=1}^{k} \mu_i \nabla h_i^T(x^*) \cdot z = 0,$$

and we have the following result.

Corollary 3.4.2 ***Let (x^*, μ) satisfy the necessary conditions of Theorem 2.2.2 when the inequalities are absent. Suppose that for every $z \in \mathbb{R}^n$ such that $\nabla h(x^*)^T \cdot z = 0$ we have***

$$\langle z, H_{xx}(x, \mu) z \rangle > 0$$

where

$$H(x, \mu) = f(x) + \langle \mu, h(x) \rangle.$$

Then, f attains a strict local minimum for prblem (P) where the inequality constraints are absent.

Corollary 3.4.3 ***(Pennisi 1953): Let (x^*, λ, μ) satisfy the KKT conditions. Let $E^* = \{i : i \in E \text{ and } \lambda_i > 0\}$. suppose that for every $z \neq 0$ in $\mathbb{R}^n$ that satisfies***

$$\langle \nabla g_i(x^*), z \rangle = 0, \quad i \in E, \quad \langle \nabla g_i(x^*), z \rangle \leq 0, \quad i \in E, \quad \langle \nabla h(x^*), z \rangle = 0 \tag{3.51}$$

the inequality

$$\langle z, f_{xx}(x^*, \lambda, \mu)\, z \rangle > 0$$

holds, where F is as in Corollary 2.4.1. Then, f attains a strict local minimum for problem (P) at x^*.

Proof **: To prove the Corollary it suffices to show that the set V_1 of vectors that satisfies (3.50) is the same as the set of vectors that satisfy the tangential constraints and the equality $\langle \nabla f(x^*), z\rangle = 0$, and then invoke Corollary 2.4.1.**

We first show that $V_1 \subset V_2$. If $V_1 = \emptyset$ there is nothing to prove. Thus, assume $V_1 \neq \emptyset$. From (ii) of Theorem 2.2.2 we have

$$\langle \nabla f(x^*), z\rangle + \langle \lambda, \nabla g(x^*)z\rangle + \langle \mu, \nabla h(x^*)z\rangle = 0. \tag{3.52}$$

Now, let $z \in V_1$. Then, $\nabla g(x^*)^T \cdot z \leq 0$, and $\nabla h(x^*)^T \cdot z \leq 0$. thus, z satisfies the tangential constraints. Moreover, since $\lambda_i = 0$ for $i \in I$, we have that $\lambda_i = 0$ for $i \notin E^*$. Thus, (3.52) gives

$$\langle \nabla f(x^*), z\rangle + \sum_{i \in E^*} \lambda_i \langle \nabla g_i(x^*), z\rangle = 0$$

for all $z \in V_1$. for such z and $i \in E^*$ we have $\langle \nabla g_i(x^*), z\rangle = 0$, and then, $\langle \nabla f(x^*), z\rangle = 0$, ad so $V_1 \subseteq V_2$.

Next we show that $V_2 \subseteq V_1$. Assume that $V_2 \neq \emptyset$. Since $\lambda_i = 0$ for $i \notin E^*$ and $\langle \nabla f(x^*), z\rangle = 0$, it follows that for $z \in V_2$, (3.52) becomes

$$\sum_{i \in E^*} \lambda_i \langle \nabla g_i(x^*), z\rangle = 0.$$

Since for any vector z that satisfies the tangential constraints we have $\langle \nabla g_i(x^*), z\rangle \leq 0$, and since $\lambda_i > 0$ for $i \in E^*$, the last inequality implies that $\langle \nabla g_i(x^*), z\rangle = 0$ for $i \in E^*$. Since $\nabla g_E(x^*)^T z \leq 0$ for any vector satisfyng the tangential constraints, we get that $V_2 \subseteq V_1$. $\square$

Example 3.4.1

$$\begin{array}{l} \min \dfrac{1}{2}\left(x_1^2 + x_2^2 + x_3^3\right) \\ \textbf{subject to} \\ x_1 + x_2 + x_3 = 3 \end{array}$$

$f(x) = \frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2\right)$ **and** $h(x) = x_1 + x_2 + x_3 - 3$. **Fritz John Theorem gives**

$$\lambda_0 \nabla f(x_0) + \mu \nabla h(x_0) = 0, \quad \lambda_0 \geq 0, \quad , \lambda_0 + |\mu| \neq 0.$$

at an optimal point x_0.

$$\begin{aligned} \lambda_0 (x_1, x_2, x_3) + \mu (1,1,1) &= (0,0,0) \\ \lambda_0 x_1 + \mu &= 0 \\ \lambda_0 x_2 + \mu &= 0 \\ \lambda_0 x_3 + \mu &= 0 \end{aligned}$$

If $\lambda_0 = 0$, then $\mu = 0$. Therefore $\lambda_0 \neq 0$. We may assume $\lambda_0 = 1$. Then,

$$\begin{aligned} x_1 + \mu &= 0 \\ x_2 + \mu &= 0 \\ x_3 + \mu &= 0 \end{aligned}$$

Thus,

$$x_1 = x_2 = x_3 = -\mu.$$

From $x_1 + x_2 + x_3 = 3$ we have $-3\mu = 3$. Thus, $\mu = -1$. Thus,

$$x_0 = (1, 1, 1).$$

Now, set

$$F(x, \mu) = f(x) + \mu h(x)$$

For every $z \neq 0$ satisfying $\langle \nabla h(x_0), z \rangle \neq 0$ we must have

$$\langle z, F_{xx}(x_0, \mu)\ z \rangle \geq 0.$$

In the present case

$$\begin{aligned} \langle \nabla h(x_0), z \rangle &= z_1 + z_2 + z_3, \text{ and} \\ F_{xx}(x_0, \mu) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

we must have

$$\langle z, F_{xx}(x_0, \mu)\ z \rangle = \langle z, z \rangle = |z|^2 \geq 0$$

for any z satisfying $z_1 + z_2 + z_3 = 0$. This is, in fact, true since $|z|^2 \geq 0$ for any $z \neq 0$.

For sufficiency, for any $z \neq 0$ satisfying $\langle \nabla h(x_0), z \rangle = 0$ and $\langle \nabla f(x_0), z \rangle = 0$ the inequality

$$\langle z, F_{xx}(x_0, \mu)\ z \rangle > 0$$

for any $z \neq 0$, although we only need the inequality to hold only for $z \neq 0$ such that $z_1 + z_2 + z_3 = 0$. Note that $\langle \nabla h(x_0), z \rangle = z_1 + z_2 + z_3$ and $\langle \nabla f(x_0), z \rangle = z_1 + z_2 + z_3$.

Example 3.4.2

Let $f(x) = x_1 + \cdots, x_n$

$$\begin{aligned} &\min f(x) \\ &\textbf{subject to} \\ &x_1^2 + \cdots + x_n^2 = 1. \end{aligned}$$

$\nabla f(x) = (1, 1, 1)^T$.
$h(x) = x_1^2 + \cdots + x_n^2 - 1$.
$\nabla h(x) = (2x_1, 2x_2, \cdots, 2x_n)^T$
From Fritz John Theorem we have

$$\lambda_0(1, 1, \cdots, 1) + \mu(2x_1, 2x_2, \cdots, 2x_n) = (0, 0, \cdots, 0)$$

Since $(0, 0, \cdots, 0)^T$ is not feasible, if $\lambda_0 = 0$ then $\mu = 0$. Thus, $\lambda_0 \neq 0$. Thus, we may assume that $\lambda_0 = 1$. Then, $\mu \neq 0$.

Now,

$$\begin{gathered} (1, 1, \cdots, 1) + 2\mu(x_1, \cdots, x_n) = (0, 0, \cdots, 0) \\ -2\mu x_i = 1, \quad i = 1, \cdots, n \end{gathered}$$

Therefore

$$x_i = -\frac{1}{2\mu}, \quad i = 1, \cdots, n.$$

Since $x_1^2 + \cdots + x_n^2 = 1$, we have

$$\begin{aligned} \frac{1}{4\mu^2} \cdot n &= 1 \\ \mu^2 &= \frac{n}{4} \\ \mu &= \pm\frac{\sqrt{n}}{2} \end{aligned}$$

Therefore

$$x_i = \mp\frac{1}{2\cdot\frac{\sqrt{n}}{2}} = \mp\frac{1}{\sqrt{n}}$$

For second order conditions, let $z \neq 0$ such that

$$\begin{aligned}
\langle \nabla h(x_0), z\rangle &= 0\\
2x_{01}z_1 + \cdots + 2x_{0n}z_n &= 0\\
x_{01}z_1 + \cdots + x_{0n}z_n &= 0\\
\mp\frac{1}{\sqrt{n}}(z_1 + \cdots + z_n) &= 0
\end{aligned}$$

Therefore

$$z_1 + \cdots + z_n = 0$$

Let

$$\begin{aligned}
F(x,\mu) &= f(x) + \mu h(x)\\
H_{xx}(x,\mu) &= 2\mu I\\
\langle z, H_{xx}z\rangle &= 2\mu|z|^2 \geq 0
\end{aligned}$$

Therefore

$$\begin{aligned}
\mu &\geq 0\\
\mu &= \frac{\sqrt{n}}{2}
\end{aligned}$$

From the sufficiency theory we see we have a local minimum at

$$\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)^T$$

Example 3.4.3

$$\begin{aligned}
&\min f(x) = x_1\\
&\textbf{subject to}\\
&(x_1+1)^2 + x_2^2 \geq 1\\
&x_1^2 + x_2^2 \leq 2
\end{aligned}$$

$$\begin{aligned}
g_1(x) &= 1 - (x_1+1)^2 - x_2^2\\
g_2(x) &= x_1^2 + x_2^2 - 2\\
\nabla g_1 &= (-2(x_1+1), -2x_2)\\
\nabla g_2 &= (2x_1, 2x_2)
\end{aligned}$$

$$\nabla f = (1,0).$$

Check Fritz-John condition at $(-1,\pm 1)$.

$$\begin{aligned}
\nabla g_1(-1,\pm 1) &= (0,\mp 2)\\
\nabla g_2(-1,\pm 1) &= (-2,\pm 2)
\end{aligned}$$

$$\lambda_0(1,0) + \lambda_1(0,\mp 2) + \lambda_2(-2,\pm 2) = (0,0)$$

$$(\lambda_0, 0) + \lambda_1(0, \mp 2\lambda_1) + (-2\lambda_2, \pm 2\lambda_2) = (0, 0)$$
$$(\lambda_0 - 2\lambda_2, \mp 2(\lambda_1 - \lambda_2)) = (0, 0)$$

Therefore $\lambda_1 - \lambda_2 = 0$**. Then**

$$(\lambda_0 - 2\lambda_2, 0) = (0, 0)$$

We can't have $\lambda_0 = 0$**. Therefore** $\lambda_0 = 1$ **and** $\lambda_2 = \frac{1}{2}$**. Since** $\lambda_1 = \lambda_2$**,** $\lambda_1 = \frac{1}{2}$**. Now, set**

$$\begin{aligned} F(x, \lambda_0, \lambda) &= x_1 + \frac{1}{2}\left(1 - (x_1+1)^2 - x_2^2\right) + \frac{1}{2}(x_1^2 + x_2^2 - 2) \\ &= x_1 + \frac{1}{2} - \frac{1}{2}(x_1+1)^2 + \frac{1}{2}x_1^2 - 1 \end{aligned}$$

$$\begin{aligned} \nabla F &= (1 - (x_1 + 1) + x_1, 0) = (0, 0) \\ F_{xx} &= 0. \end{aligned}$$

The point $(0, 0)$ **is feasible. The second constraint is not active. Therefore** $\lambda_2 = 0$**.**

$$\begin{aligned} \lambda_0(1, 0) + \lambda_1(-2, 0) &= (0, 0) \\ (\lambda_0 - 2\lambda_1, 0) &= (0, 0). \end{aligned}$$

λ_0 **can't be zero. Therefore** $\lambda_0 = 1$ **and** $\lambda_1 = \frac{1}{2}$**.**

$$\begin{aligned} F(x, \lambda_0, \lambda) &= x_1 + \frac{1}{2}\left(1 - (x_1+1)^2 - x_2^2\right) \\ F_x &= (1 - (x_1 + 1), -x_2) = (-x_1, x_2) \\ \nabla g_1(x) &= (-(x_1 + 1), -x_2) \\ \nabla g_1(0, 0) &= (-1, 0) \end{aligned}$$

The vector $(0, z_2)$**,** $z_2 \neq 0$ **is a test vector. For** $z = (0, z_2)$**,** $z_2 \neq 1$

$$\langle z, -I \cdot z\rangle = -z_2^2 < 0$$

The second order condition fails to hold. Therefore $(0, 0)$ **can't be a local minimum.**

3.5 Summary

Constraint Qualification Conditions

The constraints satisfy the Constraint Qualification conditions at x^* **if**

(i) $\nabla h_1(x^*), \cdots, \nabla h_k(x^*)$ **are linearly independet.**

(ii) Letting $E = \{i : g_i(x^*) = 0\}$ **the system**

$$\begin{cases} \nabla h(x^*)^T \cdot z = 0 \\ \nabla g_E(x^*)^T \cdot z < 0 \end{cases}$$

has asolution.

Tangential Constraints at x^*

The vector $z \neq 0$ **satisfies the tangential constraints at** x^* **if**

(i) $\nabla h(x^*)^T \cdot z = 0$

(ii) $\nabla g_E(x^*)^T \cdot z \leq 0$

Second Order Conditions

A. (1) The vector $z \neq 0$ **is a second order test vector if** $\lambda_i = 0$ **whenever** $\nabla g_i(x^*) \cdot z < 0$ **and** $g_i(x^*) = 0$

(2) Letting

$$E_2 = \{i \in E : \nabla g_i(x^*)^T \cdot z < 0\}$$
$$E_1 = \{i \in E : \nabla g_i(x^*)^T \cdot z = 0\}$$
$$\begin{pmatrix} \nabla g_{E_1}(x^*) \\ \nabla h(x^*) \end{pmatrix} \textbf{ have linearly independent rows.}$$

For every second order test vector and (2) in force we must have

$$\langle z, F_{xx}(x^*, \lambda, \mu) z \rangle \geq 0$$

B. No inequality Constraint For every $z \in \mathbb{R}^n$**,** $z \neq 0$ **and** $\nabla h(x^*)^T \cdot z = 0$ **we must have**

$$\langle z, F_{xx}(x^*, \mu) z \rangle \geq 0.$$

Sufficiency

A. Let the Fritz John second order condition hold at x^* **with multipliers** λ_0, λ, μ**. If for all** $z \neq 0$ **that satisfy the tangential constraints at** x^* **and** $\nabla f(x^*)^T \cdot z \leq 0$ **we also have**

$$\langle z, F_{xx}(x^*, \lambda_0, \lambda, \mu) z \rangle > 0$$

then x^* **is a strict local minimum.**

B. No Inequality Constraints: For every $z \in \mathbb{R}^n$**,** $z \neq 0$ **such that** $\nabla h(x^*)^T \cdot z = 0$ **we also have**

$$\langle z, F_{xx}(x^*, \lambda_0, \mu) z \rangle > 0$$

then x^* **is a strict local minimum.**