First order optimality conditions

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology
Spring 2024

Contents

- **Lagrange multiplier rule: Thm 7.1.1/2**
- **History of KKT conditions Thm 7.2.1/2/3**
- **KKT conditions under the regularity condition**

Lagrange multiplier rule

Consider an optimization problem with some equality constriants.

$$\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0, \quad i = 1, \ldots, m, \end{array}$$

where the functions are differentiable. Let $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$ be the Lagrange function. Is it always true that at all local minimizers x^*,

$$0 = \nabla L(x, \lambda)$$

for some Lagrange multiplier λ? The answer is negative.

counter-example

Consider the following simple example.

$$\begin{array}{ll} \min & x \\ s.t. & x^2 \leq 0. \end{array}$$

The only feasible solution is $x = 0$ and hence the optimal solution is $x^* = 0$. However it is easy to see that there is no scalar λ such that

$$0 = f'(x^*) + \lambda h'(x^*).$$

The difficulty of deriving a necessary optimality condition is with the constraints.
Eliminate the constraints, then the problem will become a unconstrained problem.
One way to eliminate the constraints is to use implicit function theorem.

Implicit function theorem

Theorem (Implicit function theorem) Let $F_1, \cdots, F_m : \mathbf{R}^{m+n} \to \mathbf{R}$ be C^1 functions. Consider the system of equations

$$\begin{array}{rcl} F_1(y_1, \cdots, y_m, x) & = & c_1 \\ \vdots & & \vdots \\ F_m(y_1, \cdots, y_m, x) & = & c_m \end{array}$$

as possibly defining $y_1, \cdots, y_m$ as implicit functions of x. Suppose that (y^*, x^*) is a solution of the above system of equations. If the determinant of the Jacobian of F with respect to y, i.e., the $m \times m$ matrix

$$D_y F(y, x) = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}$$

evaluated at (y^*, x^*) is nonzero,

then there exist C^1 functions

$$\begin{array}{rcl} y_1 & = & f_1(x) \\ \vdots & & \vdots \\ y_m & = & f_m(x) \end{array}$$

defined on a ball B about x^* such that

$$\begin{array}{rcl} F_1(f_1(x), \cdots, f_m(x), x) & = & c_1 \\ \vdots & & \vdots \\ F_m(f_1(x), \cdots, f_m(x), x) & = & c_m \end{array}$$

for all x in B and

$$\begin{array}{rcl} y_1^* & = & f_1(x^*) \\ \vdots & & \vdots \\ y_m^* & = & f_m(x^*) \end{array}$$

Corollary: Let $F_1, \cdots, F_m : \mathbf{R}^n \to \mathbf{R}$ be C^1 functions. Consider the system of equations

$$\begin{array}{rcl} F_1(x_1, \cdots, x_n) & = & -t \\ \vdots & & \vdots \\ F_m(x_1, \cdots, x_n) & = & 0 \end{array}$$

as possibly defining $x_1, \cdots, x_n$ as implicit functions of t. Suppose that $(x^* 0)$ is a solution of the above system of equations. If the Jacobian matrix F at x^* has rank m, then there exist C^1 functions $x_i(t)$ defined on an open interval B about 0 such that

$$\begin{array}{rcl} F_1(x_1(t), \cdots, x_n(t)) & = & -t \\ \vdots & & \vdots \\ F_m(x_1(t), \cdots, x_n(t)) & = & 0 \end{array}$$

for all t in B and $x(0) = x^*$.

Illustrate how to implicit function theorem

Consider the following case where there are two variables and one equality constraint in the optimization problem.

$$\begin{aligned} \min \quad & f(x_1, x_2) \\ s.t. \quad & h(x_1, x_2) = 0. \end{aligned}$$

Let $x^* = (x_1^*, x_2^*)$ be a local solution of the above problem. There are two possible cases.

Case 1: If $\nabla h(x_1^*, x_2^*) = 0$, then $0 = 0 \cdot \nabla f(\boldsymbol{x}^*) + \nabla h(\boldsymbol{x}^*)$.

Continue the Process

Case 2: If $\nabla h(x_1^*, x_2^*) \neq 0$, without loss of generality we may assume that $\frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) \neq 0$. Then by the implicit function theorem, since $\frac{\partial h}{\partial x_1} dx_1 + \frac{\partial h}{\partial x_2} dx_2 = 0$ and $\frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) \neq 0$, there is a C^1 function $a(x_1)$ such that $x_2 = a(x_1)$ near $\boldsymbol{x}^*$. Then x_1^* is a local minimum of $f(x_1, a(x_1))$. By the chain rule we have

$$0 = \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*) + \frac{\partial f}{\partial x_2}(\boldsymbol{x}^*) a'(x_1^*).$$

But $h(x_1, x_2) = 0$ and $x_2 = a(x_1)$, so applying the chain rule we have

$$0 = \frac{\partial h}{\partial x_1}(\boldsymbol{x}^*) + \frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) a'(x_1^*).$$

Continue the Process

If $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = 0$, then $0 = \nabla f(\boldsymbol{x}^*) + 0 \cdot \nabla h(\boldsymbol{x}^*)$.
It is easy to see that it is not possible to have $\frac{\partial f}{\partial x_1} \neq 0$ and $\frac{\partial f}{\partial x_2} = 0$. If $\frac{\partial f}{\partial x_1} = 0$ and $\frac{\partial f}{\partial x_2} \neq 0$, then $a'(x_1^*) = 0$ and hence $\frac{\partial h}{\partial x_1} = 0$. So for $\lambda = \frac{\frac{\partial f}{\partial x_2}}{\frac{\partial h}{\partial x_2}}$, we have

$$\begin{cases} 0 = \lambda \cdot 0 \\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}^*) = \lambda \frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) \end{cases} \iff \begin{cases} \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*) = \lambda \frac{\partial h}{\partial x_1}(\boldsymbol{x}^*) \\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}^*) = \lambda \frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) \end{cases} \iff \nabla f(\boldsymbol{x}^*) + \lambda \nabla h(\boldsymbol{x}^*) = 0.$$

Finally, if $\frac{\partial f}{\partial x_1} \neq 0$ and $\frac{\partial f}{\partial x_2} \neq 0$, then $\frac{\frac{\partial h}{\partial x_1}}{\frac{\partial f}{\partial x_1}} = \frac{\frac{\partial h}{\partial x_2}}{\frac{\partial f}{\partial x_2}} := -\lambda$.
So

$$\begin{cases} \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*) = \lambda \frac{\partial h}{\partial x_1}(\boldsymbol{x}^*) \\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}^*) = \lambda \frac{\partial h}{\partial x_2}(\boldsymbol{x}^*) \end{cases} \iff \nabla f(\boldsymbol{x}^*) + \lambda \nabla h(\boldsymbol{x}^*) = 0.$$

Fritz John condition

Combining cases 1 and 2, there exists $\lambda_0 \geq 0$ and λ, not all zero, such that $0 = \lambda_0 \nabla f(\boldsymbol{x}^*) + \lambda \nabla h(\boldsymbol{x}^*)$. This is called the Fritz John condition $\boldsymbol{x}^*$. It holds without any further assumptions!

Theorem: Fritz John condition with equality constraints only

Theorem 7.1.1 *Let x^* be a local optimal solution of the problem*

$$\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0, \quad i = 1, \ldots, m. \end{array}$$

Then there exists $\lambda_0 \geq 0, \lambda_i, i = 1, \ldots, m$ not all zero such that

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

Proof of Fritz John condition (equality constraints only)

Proof. Case 1: If $\nabla h_1(\boldsymbol{x}^*), \cdots, \nabla h_m(\boldsymbol{x}^*)$ are linearly dependent, then there exist $\lambda_i, i = 1, \cdots, m$ not all equal to zero such that $0 = \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*)$. Let $\lambda_0 = 0$, then $\lambda_i, i = 0, \cdots, m$ are not all equal to zero and $0 = \sum_{i=0}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*)$. So the Fritz John condition holds.

Case 2

Case 2: Suppose $\nabla h_1(\boldsymbol{x}^*), \cdots, \nabla h_m(\boldsymbol{x}^*)$ are linearly independent. If $\nabla f(\boldsymbol{x}^*), \nabla h_i(\boldsymbol{x}^*)$ are linearly dependent then we are done. Otherwise suppose that $\nabla f(\boldsymbol{x}^*), \nabla h_i(\boldsymbol{x}^*)$ are linearly independent. Consider the system of $m+1$ equations with $n+1$ variables

$$\begin{aligned} f(x) - f(x^*) &= -\varepsilon \\ h_1(x) &= 0 \\ &\vdots \\ h_m(x) &= 0. \end{aligned}$$

Since the Jacobian at x^* has a maximal rank, by the implicit function theorem (Corollary 1.4.1), for small $\varepsilon > 0$, there exists a solution x_ε^*

$$\begin{aligned} f(x_\varepsilon^*) - f(x^*) &= -\varepsilon \\ h_1(x_\varepsilon^*) &= 0 \\ &\vdots \\ h_m(x_\varepsilon^*) &= 0 \end{aligned}$$

It follows by combining cases 1 and 2 that there exist $\lambda_i, i = 0, 1, \ldots, m$ not all zero such that

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

■

Theorem 7.1.2 *Let x^* be a local optimal solution of the problem*

$$\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0, \quad i = 1, \ldots, m. \end{array}$$

If the gradient vectors $\nabla h_1(x^*), \cdots, \nabla h_m(x^*)$ are linearly independent, then there exists Lagrange multipliers $\lambda_i, i = 1, \ldots, m$ such that

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

Theorem 7.2.1 (Fritz John Necessary Optimality Condition) *Suppose that x^* is a local optimal solution of the following problem:*

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) \leq 0 \quad i = 1, \ldots, m \end{array}$$

where f, g are differentiable functions. Then there exist $\lambda_i, i = 0, 1, \ldots, m$ not all zero such that

$$0 = \lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*)$$
$$0 = \lambda_i g_i(x^*), \quad \lambda_i \geq 0 \quad i = 0, 1, \ldots, m.$$

NNAMOCQ and PLICQ

Let x^* be a feasible solution. We say that the no nonzero abnormal constraint qualification (NNAMCQ) holds at x^* if there is no nonzero abnormal multipliers.
On the other hand, we call the condition (7.1) Positive Linear Independence Constraint Qualification (PLICQ) if

$$\left.\begin{array}{l} 0 = \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) \\ 0 = \lambda_i g_i(x^*) \quad i = m, \ldots, p \\ \lambda_i \geq 0, i = m, \ldots, p \end{array}\right\} \Longrightarrow \lambda_1 = \lambda_2 = \cdots = \lambda_p = 0 \qquad (7.1)$$

NNAMCQ is the same as PLICQ.

Theorem 7.2.2 (KKT condition under the NNAMCQ (PLICQ)) *Suppose that x^* is a local optimal solution of the following problem:*

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) = 0 \quad i = 1, \ldots, m-1 \\ & g_i(x) \leq 0 \quad i = m, \ldots, p \end{array}$$

where f, g are differentiable functions. Suppose that NNAMCQ or equivalently PLICQ holds at x^*. Then there exist $\lambda_i, i = 1, \ldots, p$ not all zero such that

$$\begin{aligned} 0 &= \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) \\ \lambda_i &\geq 0, 0 = \lambda_i g_i(x^*) \quad i = m, \ldots, p. \end{aligned}$$

The following theorem will be useful for interpreting the NNAMCQ in an equivalent way.

Theorem 7.2.3 (Motzkin's Theorem of Alternative) *Given matrices $A_{p\times n}$ which is nonvacuous and $D_{m\times n}$. One and only one of the following systems has a solution.*

(I) $Ax < 0, Dx = 0, x \in R^n$

(II) $A^T y + D^T z = 0, y \geq 0$ *and* $y \neq 0 \quad (y \in R^p, z \in R^m)$.

Let x^* be feasible and denote the set of active constraints (binding constrains) by

$$I(x^*) := \{i = m, \cdots, p : g_i(x^*) = 0\}.$$

$\nabla g_1(x^*), \ldots, \nabla g_{m-1}(x^*)$ are linear independent and

system I $\left\{ \begin{array}{ll} \nabla g_i(x^*)^T d = 0 & i = 1, \ldots, m-1 \\ \nabla g_i(x^*)^T d < 0 & i \in I(x^*) \end{array} \right.$ has a solution.d.

The above condition is called Mangasarian Fromovitz constraint qualification (MFCQ).

KKT conditions under the regularity condition

Theorem 7.3.1 (KKT condition with inequality constraints) *Let x^* be a local optimal solution of the problem*

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) \leq 0, i = 1, \ldots, m. \end{array}$$

Suppose that x^* is a regular point then the KKT condition holds at x^*.

Proof. Without loss of generality we suppose that only $g_1, \ldots, g_e$ are active at x^*. That is,

$$g_1(x^*) = 0, \ldots, g_e(x^*) = 0$$
$$g_{e+1}(x^*) < 0, \ldots, g_m(x^*) < 0.$$

Then x^* being a regular point means that the $e \times n$ Jacobian matrix

$$Dg_E(x^*) = \begin{pmatrix} \nabla g_1(x^*) \\ \nabla g_2(x^*) \\ \\ \nabla g_e(x^*) \end{pmatrix}$$

has rank e. We want to prove that there exist $\lambda_1, \ldots, \lambda_e$ nonnegative such that

$$0 = \nabla f(x^*) + \sum_{i=1}^{e} \lambda_i \nabla g_i(x^*).$$

Since g_i are continuous functions, there is an open ball $B = B_r(x^*)$ of radius $r > 0$ about x^* such that $g_i(x) < 0$ for all $x \in B, j = e+1, \ldots, m$. So with respect to local solutions, inactive constraints can be deleted and the active inequality constraints become equality constraints. By the KKT condition for the equality case, there exist $\lambda_1, \ldots, \lambda_e$ such that

$$0 = \nabla f(x^*) + \sum_{i=1}^{e} \lambda_i \nabla g_i(x^*). \tag{7.2}$$

Now we only need to prove that $\lambda_i \geq 0, i = 1, \ldots, e$.

Consider the system of e equations in $n+1$ variables:

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= -t \\ g_i(x_1, \ldots, x_n) &= 0 \quad i = 2, \ldots, e. \end{aligned}$$

Since $Dg_E(x^*)$ has rank e by the implicit function theorem Corollary 1.4.1, there $\varepsilon > 0$ and C^1 functions $x_1(t), \ldots, x_n(t)$ defind for $t \in [0, \varepsilon)$ such that $x(0) = x^*$ and for all $t \in [0, \varepsilon)$,

$$\begin{aligned} g_1(x_1(t), \ldots, x_n(t)) &= -t \\ g_i(x_1(t), \ldots, x_n(t)) &= 0 \quad \forall i = 2, \ldots, e. \end{aligned}$$

Let $v = x'(0)$. Applying the chain rule we conclude that

$$\begin{aligned} \nabla g_1(x^*) \cdot v &= -1 \\ \nabla g_i(x^*) \cdot v &= 0 \quad \forall i = 2, \ldots, e. \end{aligned}$$

Since $x(t)$ lies in the constraint region for all t and x^* minimizes f in the constraint region. $t \to f(x(t))$ attains minimum at the left boundary point $t = 0$. Therefore

$$\frac{d}{dt} f(x(t))|_{t=0} = \nabla f(x^*) \cdot v \geq 0.$$

By (7.2), $\nabla f(x^*) \cdot v = -\sum_{i=1}^{e} \lambda_i \nabla g_i(x^*) \cdot v = \lambda_1$ and hence $\lambda_1 \geq 0$. Similarly we can show that $\lambda_i \geq 0 \quad i = 2, \ldots, e$.

Sufficient optimality conditions

Consider the following optimization problem:

$$
\begin{array}{llll}
(P) & \min & f(x) & \\
 & \text{s.t.} & g_i(x) = 0 & i = 1, \cdots, m-1 \\
 & & g_i(x) \le 0 & i = m, \cdots, p \\
 & & x \in C. &
\end{array}
$$

Let x^* be a feasible point of problem (P). Denote by

$$I(x^*) := \{i = m, \cdots, p \,:\, g_i(x^*) = 0\}$$

the index set of all active inequalities and

$$J(x^*) = I(x^*) \cup \{1, \cdots, m-1\}$$

the index set of all active constraints.

Definition (Quasiconvex function)

The function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be quasiconvex if for each $x, y \in \mathbb{R}^n$, the following inequality is true:

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\} \qquad \text{for each } \lambda \in (0, 1).$$

f is quasiconcave if and only $-f$ is quasiconvex.

Proposition 4.4.1 *Let $f : C \to R$ where C is a nonempty convex set in R^n. The function f is quasiconvex if and only if the sublevel set*

$$C_\alpha = \{x \in C : f(x) \leq \alpha\}$$

is convex for each real number α.

Proof. Suppose that f is quasiconvex, and let $x_1, x_2 \in C_\alpha$. Therefore $x_1, x_2 \in C$ and $\max\{f(x_1), f(x_2)\} \leq \alpha$. Let $\lambda \in (0,1)$ and let $x = \lambda x_1 + (1-\lambda)x_2$. By the convexity of C, $x \in C$. Furthermore by the quasiconvexity of f, $f(x) \leq \max\{f(x_1), f(x_2)\} \leq \alpha$. Hence $x \in C_\alpha$ and thus C_α is convex. Conversely, suppose that C_α is convex for some real number α. $x_1, x_2 \in C$. Let $\lambda \in (0,1)$ and let $x = \lambda x_1 + (1-\lambda)x_2$. Observe that $x_1, x_2 \in C_\alpha$ with $\alpha = \max\{f(x_1), f(x_2)\}$. By assumption, C_α is convex, so that $x \in C_\alpha$. Therefore, $f(x) \leq \alpha = \max\{f(x_1), f(x_2)\}$. Hence f is quasiconvex and the proof is complete.

Definition 4.4.2 (Strict quasiconvex function) *Let $C \subset R^n$ be a convex set and $f : C \to R$. The function f is said to be strictly quasiconvex if, for each $x_1, x_2 \in C$ with $f(x_1) \neq f(x_2)$, the following inequality is true:*

$$f(\lambda x_1 + (1 - \lambda)x_2) < \max\{f(x_1), f(x_2)\} \quad \textit{for each } \lambda \in (0, 1).$$

f is strictly quasiconcave if and only if $-f$ is strictly quasiconvex.

Theorem 4.4.1 *(a) If f and g are quasiconvex functions on a convex set C in R^n, then the sum $f(x) + g(x)$ may not be a quasconvex function.*

(b) If $f(x)$ is a quasiconvex (resp. strictly quasiconvex) function defined on a convex set C in R^n and if α is a positive number, then $\alpha f(x)$ is quasiconvex (resp. strictly quasiconvex) on C.

(c) If $f(x)$ is a quasiconvex (resp. strictly quasiconvex) function defined on the convex set C in R^n, and if $g(y)$ is an increasing (resp. strictly increasing) function defined on the range of $f(x)$ in R, then the composite function $g(f(x))$ is quasiconvex (resp. strictly quasiconvex) on C.

(d) Let $g : R^m \to R$ be a quasiconvex function, and let $h : R^n \to R^m$ be an affine function of the form $h(x) = Ax + b$, where A is an $m \times n$ matrix and b is an $m \times 1$ vector. Then the composite function $f : R^n \to R$ defined as $f(x) = g[h(x)]$ is a quasiconvex function.

Proposition 4.4.2 *Let $g : D \to R$ and $h : D \to R$ where D is a nonempty convex subset of R^n. If*

(a) *g is convex on D and $g(x) \geq 0 \forall x \in D$*

(b) *h is concave on D and $h(x) > 0 \forall x \in D$,*

then $f(x) = g(x)/h(x)$ is quasiconvex. If

(a) *g is convex on D and $g(x) \leq 0 \forall x \in D$*

(b) *h is concave on D and $h(x) > 0 \forall x \in D$,*

then $f(x) = g(x)h(x)$ is quasiconvex.

Theorem 4.4.2 *Let $f : R^n \to R$ be strictly quasiconvex. Consider the problem to minimize $f(x)$ subject to $x \in C$, where C is a nonempty convex set in R^n. If $\bar{x}$ is a local optimal solution, then $\bar{x}$ is also a global optimal solution.*

Proof. Assume, on the contrary, that there exists an $\hat{x} \in C$ with $f(\hat{x}) < f(\bar{x})$. By the convexity of C, $\lambda\hat{x} + (1-\lambda)\bar{x} \in C$ for each $\lambda \in (0,1)$. Since $\bar{x}$ is a local minimum by assumption, then $f(\bar{x}) \leq f(\lambda\hat{x} + (1-\lambda)\bar{x})$ for all small enough $\lambda \in (0,1)$. But because f is strictly quasiconvex, and $f(\hat{x}) < f(\bar{x})$, $f(\lambda\hat{x} + (1-\lambda)\bar{x}) < f(\bar{x})$ for each $\lambda \in (0,1)$. This contradicts the local optimality of $\bar{x}$, and the proof is complete.

Definition 4.4.3 (Pseudoconvex function) *Let $C \subset R^n$ and $\bar{x} \in C$. A differentiable function $f : C \to R$ is pseudoconvex at $\bar{x}$ provided that for any $x \in C$,*

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \Longrightarrow f(x) \geq f(\bar{x}),$$

or equivalently

$$f(x) < f(\bar{x}) \Longrightarrow \nabla f(\bar{x})^T (x - \bar{x}) < 0.$$

The function f is said to be pseudoconvex on C if it is pseudoconvex at each point of C. We say that f is pseudoconcave if and only if $-f$ is pseudoconvex.

Proposition 4.4.3 *Let $c_1, c_2 \in R^n$ be nonzero vectors. $\alpha_1, \alpha_2 \in R$ and $D = \{x \in R^n : c_2^t x + \alpha_2 > 0\}$. Then $f(x) : D \to R$ defined by $f(x) = \frac{c_1^t x + \alpha_1}{c_2^t x + \alpha_2}$ is both pseudoconvex and pseudoconcave.*

Theorem 4.4.3 *Let C is a convex set and $f : C \to R$ be pseudoconvex. Then $\bar{x}$ is a global minimum of f if and only if*

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C.$$

Theorem 4.4.4 *Let $f : C \to R$ be a differentiable convex function defined on an open convex set $C \subset R^n$. Then f is both strictly quasiconvex and quasiconvex.*

Proof. Since a differentiable function must be continuous and a continuous strictly quasiconvex function is quasiconvex. We only need to show that a pseudoconvex function f must be strictly quasiconvex. By contradiction, suppose that there exist $x_1, x_2 \in C$ such that $f(x_1) \neq f(x_2)$ and $f(x') \geq \max\{f(x_1), f(x_2)\}$ where $x' = \lambda x_1 + (1-\lambda)x_2$, for some $\lambda \in (0,1)$. Without loss of generality, assume that $f(x_1) < f(x_2)$, so that

$$f(x') \geq f(x_2) > f(x_1) \tag{4.10}$$

which implies, by the pseudoconvexity of f, that

$$\nabla f(x')^T(x_1 - x') < 0.$$

Now by $\nabla f(x')^T(x_1 - x') < 0$ and $x_1 - x' = -(1-\lambda)(x_2 - x')/\lambda$ we have

$$\nabla f(x')^T(x_2 - x') > 0;$$

and, hence, by the pseudoconvexity of f, we must have $f(x_2) \geq f(x')$. Therefore by (4.10), we get $f(x_2) = f(x')$. Also since the directional derivative $f'(x'; x_2 - x') = \nabla f(x')^T(x_2 - x') > 0$, there exists a point $\hat{x} = \mu x' + (1-\mu)x_2$ with $\mu \in (0,1)$ such that

$$f(\hat{x}) > f(x') = f(x_2)$$

Again by the pseudoconvexity of f, we have $\nabla f(\hat{x})^T(x_2 - \hat{x}) < 0$. Similarly, $\nabla f(\hat{x})^T(x' - \hat{x}) < 0$. Summarizing, we must have

$$\nabla f(\hat{x})^T(x_2 - \hat{x}) < 0$$
$$\nabla f(\hat{x})^T(x' - \hat{x}) < 0$$

Note that $x_2 - \hat{x} = \mu(\hat{x} - x')(1-\mu)$ and, hence, by the above two inequalities are not compatible. This contradiction shows that f is strictly quasiconvex.

Theorem 8.0.1 (First order KKT sufficient condition) *Let $x^* \in C$ be a feasible point of problem (P) where f, g_i are C^1. Suppose that KKT condition holds at x^*, i.e., there exists $\lambda_i, i = 1, \ldots, p$ such that*

$$\langle \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C$$
$$\lambda_i \geq 0, 0 = \lambda_i g_i(x^*) \quad i = m, \ldots, p.$$

Suppose that C is a convex set, f is pseudoconvex at x^* on C, g_i is quasiconvex for all $i \in I(x^*)$, $g_i (i = 1, \ldots, m-1)$ is both quasiconvex and quasiconcave. Then x^* is a global optimal solution to (P).

Proof. Let x^* be any feasible solution to problem (P).

Then for $i \in I(x^*)$ and any feasible solution x, $g_i(x) \leq g_i(x^*)$. By the quasiconvexity of g_i at x^*, it follows that

$$g_i(x^*+\lambda(x-x^*)) = g_i(\lambda x+(1-\lambda)x^*) \leq \max\{g_i(x), g_i(x^*)\} = g_i(x^*) \quad \forall \lambda \in (0,1).$$

Dividing the above inequality by λ and take $\lambda \downarrow 0$, one has

$$\langle \nabla g_i(x^*), x - x^* \rangle \leq 0 \quad \forall i \in I(x^*). \tag{8.1}$$

Similarly since $g_i (i = 1, \ldots, m-1)$ are both quasiconvex and quasiconcave, we have

$$\langle \nabla g_i(x^*), x - x^* \rangle = 0 \quad \forall i = 1, \ldots, m-1. \tag{8.2}$$

Multiplying (8.1) and (8.2) by $\lambda_i \geq 0 (i \in I(x^*))$ and $\lambda_i (i = 1, \ldots, m-1)$ respectively and adding, we get

$$\langle \sum_{i \in J(x^*)} \lambda_i \nabla g_i(x^*), x - x^* \rangle \leq 0 \qquad x \in C.$$

By assumption the KKT condition holds at x^*. Therefore

$$\langle \nabla f(x^*), x - x^* \rangle \geq -\langle \sum_{i \in J(x^*)} \lambda_i \nabla g_i(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C.$$

By the pseudoconvexity of f at x^*, we must have $f(x) \geq f(x^*)$, and the proof is complete.

$\blacksquare$

Proposition 8.0.1 *Let f be a C^1 function on R^n and C is a convex subset of R^n. Let $\bar{x} \in C$. Suppose that f is a quasiconcave at on C and $\nabla f(\bar{x}) \neq 0$. Then f is pseudoconcave at $\bar{x}$ on C.*

Proof. By the definition of pseudoconcavity at $\bar{x}$, it suffices to show that for all $x \in C$

$$f(x) > f(\bar{x}) \Longrightarrow \nabla f(\bar{x}) \cdot (x - \bar{x}) > 0.$$

Since $f(x) > f(\bar{x})$, one can find $\alpha > 0$ very small such that

$$f(x - \alpha \nabla f(\bar{x})) > f(\bar{x}).$$

Since by assumption f is a quasiconcave at on C, for all $\lambda \in (0, 1)$,

$$f(\lambda(x - \alpha \nabla f(\bar{x})) + (1 - \lambda)\bar{x}) \geq \min\{f(x - \alpha \nabla f(\bar{x})), f(\bar{x})\} = f(\bar{x}).$$

That is,

$$f(\bar{x} + \lambda(x - \alpha\nabla f(\bar{x}) - \bar{x})) - f(\bar{x}) \geq 0.$$

Dividing the above inequality by λ and let λ approach zero, we have

$$\nabla f(\bar{x}) \cdot (x - \alpha\nabla f(\bar{x}) - \bar{x}) \geq 0.$$

Since $\nabla f(\bar{x}) \neq 0$, the above inequality implies that

$$\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq \alpha\|\nabla f(\bar{x})\|^2 > 0.$$

Hence the proof is complete.

Theorem 8.0.1 *Let $x^* \in C$ be a feasible point of the maximization problem*

$$
\begin{array}{lll}
(P_{max}) & \max & f(x) \\
& s.t. & g_i(x) = 0 \quad i = 1, \ldots, m-1 \\
& & g_i(x) \leq 0 \quad i = m, \ldots, p \\
& & x \in C.
\end{array}
$$

where f, g_i are C^1. Suppose that KKT condition holds at x^*, i.e., there exists $\lambda_i, i = 1, \ldots, p$ such that

$$
\begin{aligned}
&\langle \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*), x - x^* \rangle \leq 0 \quad \forall x \in C \\
&\lambda_i \geq 0, 0 = \lambda_i g_i(x^*) \quad i = m, \ldots, p.
\end{aligned}
$$

Suppose that C is a convex set, f is quasiconcave on C and $\nabla f(x^*) \neq 0$, g_i is quasiconvex for all $i \in I(x^*)$, $g_i (i = 1, \ldots, m-1)$ is both quasiconvex and quasiconcave. Then x^* is a global optimal solution to (P_{max}).

Theorem (second order KKT sufficient condition) Let $x^* \in \mathrm{int}(C)$ be feasible to (P), where f, g_i are of class C^2. Suppose that KKT condition holds at x^*, then for any direction $v \neq 0$ with

$$\nabla f_i(x^*)^T v \leq 0; \nabla g_i\left(x^*\right)^T v = 0, \forall i = 1, 2, \cdots, m-1; \nabla g_i(x^*)^T v \leq 0, \forall i \in I(x^*) \quad (**)$$

there exists λ of KKT multipliers such that

$$v^T \nabla_x^2 L\left(x^*, \lambda\right) v > 0.$$

Then x^* is a strict local optimal solution to (P).

Proof: Suppose the conclusion does not hold. Then $\exists \{x^k\} \to x^*$ of feasible points such that $f(x^k) \leq f(x^*)$. Denote $v^k = (x^k - x^*) / \|x^k - x^*\|$, then extracting a subsequence if necessary, we can assume that $v^k \to \bar{v}$ with $\|\bar{v}\| = 1$. Then

(1.) $\bar{v}$ satisfies (**).

Actually, by the Taylor expansion, there exists $\tilde{x}^k \in [x^*, x^k]$ such that

$$f\left(x^k\right) = f(x^*) + \left(x^k - x^*\right)^T \nabla f\left(\tilde{x}^k\right)$$

$$g_i\left(x^k\right) = g\left(x^*\right) + \left(x^k - x^*\right)^T \nabla g_i\left(\tilde{x}^k\right) \forall i = 1, 2, \cdots, m-1$$

$$g_i\left(x^k\right) = g\left(x^*\right) + \left(x^k - x^*\right)^T \nabla g_i\left(\tilde{x}^k\right) \forall i \in I(x^*)$$

Recall that $f(x^k) \le f(x^*)$, $g_i(x^k) = 0$ for $i = 1, 2, \cdots, m-1$ and $g_i(x^k) \le 0 = g(x^*)$ for $i \in I(x^*)$. We have

$$
\begin{aligned}
0 &\ge \frac{f(x^k) - f(x^*)}{\|x^k - x^*\|} = \left(v^k\right)^T \nabla f\left(\tilde{x}^k\right) \\
0 &= \left(v^k\right)^T \nabla g_i\left(\tilde{x}^k\right), \quad \forall i = 1, 2, \cdots, m-1 \\
0 &\ge \frac{g(x^k)}{\|x^k - x^*\|} = \left(v^k\right)^T \nabla g_i\left(\tilde{x}^k\right), \quad \forall i \in I(x^*).
\end{aligned}
$$

Taking the limit,

$$0 \ge \bar{v}^T \nabla f(x^*); \quad 0 = \bar{v}^T \nabla g_i(x^*), \forall i = 1, 2, \cdots, m-1;\ 0 \ge \bar{v}^T \nabla g_i(x^*), \forall i \in I(x^*).$$

(2.) However for all λ of KKT multipliers, we have $\bar{v}^T \nabla_x^2 L(x^*, \lambda) \bar{v} \le 0$. Notice that

$$
\begin{aligned}
f\left(x^k\right) \ge & f(x^k) + \sum_{i=1}^{p} \lambda_i g_i(x^k) \\
= & f(x^*) + \sum_{i=1}^{p} \lambda_i g_i(x^*) + \left(x^k - x^*\right)^T \left(\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*)\right) \\
& + \frac{1}{2}\left(x^k - x^*\right)^T \left(\nabla^2 f(y^k) + \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(y^k)\right)\left(x^k - x^*\right) \\
= & f(x^*) + \frac{1}{2}\left(x^k - x^*\right)^T \left(\nabla^2 f(y^k) + \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(y^k)\right)\left(x^k - x^*\right),
\end{aligned}
$$

where $y^k \in [x^*, x^k]$ and the last equality holds since λ is a KKT multiplier.

Recall $f(x^k) \leq f(x^*)$, we have

$$0 \geq \frac{f(x^k) - f(x^*)}{\left\|x^k - x^*\right\|^2} = \frac{1}{2}\left(v^k\right)^T \nabla_x^2 L\left(y^k, \lambda\right) v^k.$$

Taking the limit,

$$0 \geq \bar{v}^T \nabla_x^2 L\left(x^*, \lambda\right) \bar{v},$$

which gives a contradiction.