Chapter 4

Constrained Optimality Criteria

4.1 Equality Constrained Problems

NLP Problem 1

$$\min f(\mathbf{x})$$

subject to

$$h_k(\mathbf{x}) = 0 \qquad k = 1, \ldots, K$$

$$\mathbf{x} = (x_1, x_2, \ldots, x_N)^T \in R^N$$

Example 4.1

$$\min f(\mathbf{x}) = x_1 x_2 x_3$$

subject to

$$h_1(\mathbf{x}) = x_1 + x_2 + x_3 - 1 = 0 \qquad \text{Case 1}$$

or

$$h_1(\mathbf{x}) = x_1^2 x_3 + x_2 x_3^2 + x_1/x_2 = 0 \qquad \text{Case 2}$$

- variable-elimination method

$$x_3 = 1 - x_1 - x_2 \qquad \text{Case 1}$$

$$x_1 = \frac{-1 \pm \sqrt{1 - 4{x_2}^3{x_3}^3}}{2x_2x_3} \quad \text{or} \quad x_2x_3 = \frac{-{x_1}^2 \pm \sqrt{{x_1}^4 - 4x_1}}{2} \qquad \text{Case 2}$$

- method of Lagrange multiplier

4.2 Lagrange Multipliers

Lagrangian Function of the NLP Problem 1

$$L(\mathbf{x}; \mathbf{v}) = f(\mathbf{x}) - \mathbf{v}^T\mathbf{h}(\mathbf{x}) = f(\mathbf{x}) - \sum_{k=1}^{K} v_k h_k(\mathbf{x})$$

- $L(\mathbf{x};\mathbf{v})$: the Lagrangian function
- $\mathbf{v} = (v_1, \ldots, v_K)^T \in R^K$: the Lagrange multiplier. No sign restrictions on v_k's.

Example 4.2 (Geometric Interpretation of Lagrange Multipliers)

$$\min f(\mathbf{x}) = x_1^2 + x_2^2$$

subject to

$$h_1(\mathbf{x}) = 2x_1 + x_2 - 2 = 0$$

Solution *Using the variable-elimination method:* $x_2 = 2 - 2x_1$

$$\mathbf{x}^* = \left(\frac{4}{5}, \frac{2}{5}\right)$$

$$\min f(\mathbf{x}) = f(\mathbf{x}^*) = \frac{4}{5}$$

$$\nabla f(\mathbf{x}^*) = \left(\frac{8}{5}, \frac{4}{5}\right)$$

$$\nabla h_1(\mathbf{x}^*) = (2, 1)$$

and

$$\nabla f(\mathbf{x}^*) = \frac{4}{5}\nabla h_1(\mathbf{x}^*)$$

Note that, as shown in Figure 4.1, $\nabla f(\mathbf{x}^*)$ is parallel to[1] $\nabla h_1(\mathbf{x}^*)$, *i.e.,*

$$\nabla f(\mathbf{x}^*) = v_1 \nabla h_1(\mathbf{x}^*)$$

or

$$\nabla[\underbrace{f(\mathbf{x}) - v_1 h_1(\mathbf{x})}_{\equiv L(\mathbf{x};v_1)}]_{\mathbf{x}=\mathbf{x}^*} = 0$$

The Lagrangian function $L(\mathbf{x};\mathbf{v})$ is treated as a function of $\mathbf{x}$ where $\mathbf{v}$ is considered as a parameter whose value is *adjusted* to satisfy the constraint.

The values of $\mathbf{x}$ and $\mathbf{v}$ can be simultaneously determined from

$$\frac{\partial L}{\partial x_i} = 0 \qquad \forall i = 1, \ldots, N$$

$$h_k(\mathbf{x}) = \frac{\partial L}{\partial v_k} = 0 \qquad \forall k = 1, \ldots, K$$

[1]has either the same or the opposite direction with

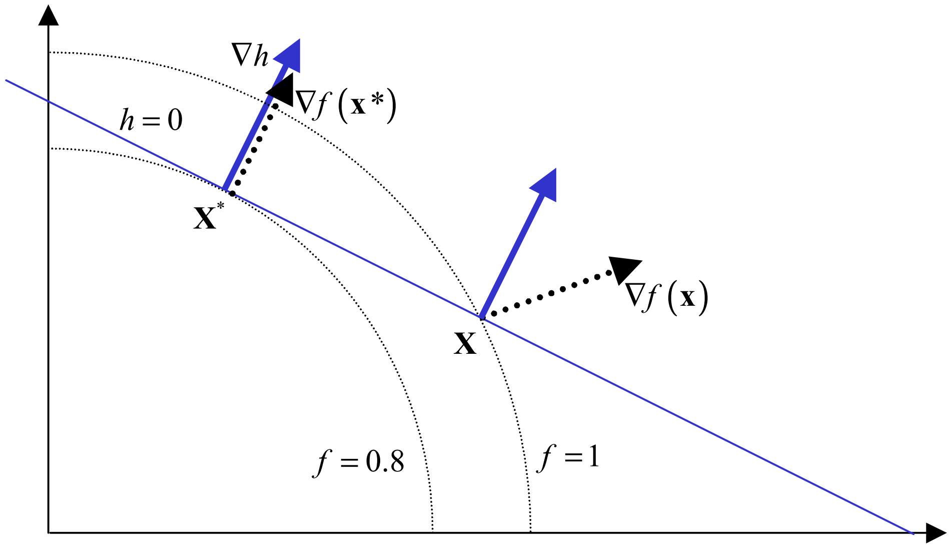

Figure 4.1: Geometric interpretation of Lagrange mulitpliers

4.3 Economic Interpretation of Lagrange Multipliers

$$\min f(x_1, x_2)$$

subject to

$$h_1(x_1, x_2) = b_1$$

Method of Lagrange multiplier:

$$L(x_1, x_2, v_1) = f(x_1, x_2) - v_1[h_1(x_1, x_2) - b_1]$$

$$\left(\frac{\partial L}{\partial x_i}\right)_{\mathbf{x}^*} = \left(\frac{\partial f}{\partial x_i}\right)_{\mathbf{x}^*} - v_1 \left(\frac{\partial h_1}{\partial x_i}\right)_{\mathbf{x}^*} = 0$$

for $i = 1, 2$. Change in $f(\mathbf{x}^*)$ due to change in b_1:

$$\left(\frac{\partial f}{\partial b_1}\right)_{\mathbf{x}^*} = \sum_{i=1}^{2} \left(\frac{\partial f}{\partial x_i}\right)_{\mathbf{x}^*} \left(\frac{\partial x_i^*}{\partial b_1}\right) \tag{4.1}$$

Change in $h_1(\mathbf{x}^*) - b_1$ due to change in b_1:

$$\sum_{i=1}^{2} \left(\frac{\partial h_1}{\partial x_i}\right)_{\mathbf{x}^*} \left(\frac{\partial x_i^*}{\partial b_1}\right) - 1 = 0 \tag{4.2}$$

Eq. (4.1) $- v_1^* \times$ Eq. (4.2) yields

$$\left(\frac{\partial f}{\partial b_1}\right)_{\mathbf{x}^*} = v_1^* + \sum_{i=1}^{2} \underbrace{\left(\frac{\partial f}{\partial x_i} - v_1^* \frac{\partial h_1}{\partial x_i}\right)_{\mathbf{x}^*}}_{=0} \frac{\partial x_i^*}{\partial b_1} = v_1^*$$

The rate of change of the optimal value of f with respect to b_1 is given by the optimal value of the Lagrange multiplier v_1^*.

Example 4.3 (Canonical ensemble system) *In a pure component canonical ensemble system of number of molecules $\mathcal{N}$, volume $\mathcal{V}$, and energy $\mathcal{E}$, the energy is distrubuted so that the entropy of the system is maximized.*

$$\max S_m = -k_B \sum_{i=0}^{\infty} p_i \ln p_i$$

subject to

$$\sum_{i=0}^{\infty} p_i = 1$$

$$\sum_{i=0}^{\infty} p_i E_i = E_m$$

where

- *E_i is i-th energy level*
- *$E_m = \mathcal{E}/\mathcal{N}$ is the mean molecular internal energy*
- *k_B is Boltzmann constant*
- *p_i is the probability of molecules having energy level E_i*
- *$S_m = \mathcal{S}/\mathcal{N}$ is the mean molecular entropy*

Assume that p_i's can be considered as continuous variables.

$$L(p_i, \alpha, \beta) = -k_B \sum_i p_i \ln p_i - \alpha \left(\sum_i p_i - 1 \right) - \beta \left(\sum_i p_i E_i - E_m \right)$$

$$\frac{\partial L}{\partial p_i} = -k_B(1 + \ln p_i) - \alpha - \beta E_i = 0 \qquad i = 0, 1, \ldots$$

$$\ln p_i = -\frac{k_B + \alpha + \beta E_i}{k_B}$$

$$S_m = \sum_{i=1}^{\infty} p_i (k_B + \alpha + \beta E_i) = k_B + \alpha + \beta E_m$$

Interpretation of β

$$\beta = \left(\frac{\partial S_m}{\partial E_m} \right)_{\mathcal{V}, \mathcal{N}} = \frac{1}{T}$$

On the other hand,

$$p_i = \exp\left(-1 - \frac{\alpha}{k_B} \right) \exp\left(-\frac{\beta E_i}{k_B} \right)$$

$$\sum_i p_i = \exp\left(-1 - \frac{\alpha}{k_B} \right) \sum_i \exp\left(-\frac{\beta E_i}{k_B} \right) = 1$$

Assume $E_i = i\epsilon$ then

$$\sum_i p_i = \exp\left(-1 - \frac{\alpha}{k_B} \right) \frac{1}{1 - \exp(-\beta\epsilon / k_B)} = 1$$

or

$$\exp\left(-1 - \frac{\alpha}{k_B} \right) = 1 - \exp\left(-\frac{\beta\epsilon}{k_B} \right)$$

Mean molecular internal energy is

$$\sum_i p_i E_i = \exp\left(-1 - \frac{\alpha}{k_B} \right) \sum_i \exp\left(-\frac{\beta E_i}{k_B} \right) E_i = E_m$$

$$\begin{aligned}\sum_i \exp\left(-\frac{\beta E_i}{k_B}\right) E_i &= -k_B \frac{d}{d\beta} \sum_i \exp\left(-\frac{\beta E_i}{k_B}\right) \\ &= -k_B \frac{d}{d\beta}\left(\frac{1}{1-\exp(-\beta\epsilon/k_B)}\right) \\ &= -k_B \left(\frac{-\exp(-\beta\epsilon/k_B)\epsilon/k_B}{[1-\exp(-\beta\epsilon/k_B)]^2}\right)\end{aligned}$$

$$E_m = \frac{\exp(-\beta\epsilon/k_B)\epsilon}{1-\exp(-\beta\epsilon/k_B)} \approx \frac{k_B}{\beta} = k_B T \qquad (\epsilon << k_B/\beta)$$

or

$$\mathcal{E} = k_B \mathcal{N} T$$

4.4 Kuhn-Tucker Conditions

Karush[?] and later Kuhn and Tucker[?] have extended the method of Lagrange multiplier to include the general nonlinear programming (NLP) problem with both equality and inequality constraints.

NLP Problem 2

$$\min f(\mathbf{x}) \tag{4.3}$$

subject to

$$g_j(\mathbf{x}) \geq 0 \qquad \forall j = 1, \ldots, J \tag{4.4}$$

$$h_k(\mathbf{x}) = 0 \qquad \forall k = 1, \ldots, K \tag{4.5}$$

$$\mathbf{x} \in R^N$$

Note that NLP Problem 1 is a special case of NLP Problem 2 when $J = 0$.

Definition 4.1 (Feasible Solution) $\bar{\mathbf{x}}$ *is a* feasible solution *to NLP Problem 2 when* $g_j(\bar{\mathbf{x}}) \geq 0$ *for* $j = 1, \ldots, J$ *and* $h_k(\bar{\mathbf{x}}) = 0$ *for* $k = 1, \ldots, K$.

Definition 4.2 (Local Minimum) $\mathbf{x}^*$ *is a* local minimum *to NLP Problem 2 when* $\mathbf{x}^*$ *is feasible and* $f(\mathbf{x}^*) \leq f(\bar{\mathbf{x}})$ *for all feasible* $\bar{\mathbf{x}}$ *in some small neighborhood* $\delta(\mathbf{x}^*)$ *of* $\mathbf{x}^*$

Definition 4.3 (Strict Local Minimum) $\mathbf{x}^*$ *is a* strict *(*unique *or* isolated*)* local minimum *when* $\mathbf{x}^*$ *is feasible and* $f(\mathbf{x}^*) < f(\bar{\mathbf{x}})$ *for all feasible* $\bar{\mathbf{x}} \neq \mathbf{x}^*$ *in some small neighborhood* $\delta(\mathbf{x}^*)$ *of* $\mathbf{x}^*$

Definition 4.4 *The inequality constraint* $g_j(\mathbf{x}) \geq 0$ *is said to be an* active *or* binding constraint *at the point* $\bar{\mathbf{x}}$ *if* $g_j(\bar{\mathbf{x}}) = 0$*; it is said to be* inactive *or* nonbinding *if* $g_j(\bar{\mathbf{x}}) > 0$.

Kuhn and Tucker have developed the necessary and sufficient optimality conditions for NLP Problem 2 assuming that the functions f, g_j, and h_k are differentiable. The optimality conditions, commonly known as the *Kuhn-Tucker conditions* (KTC) may be stated in the form of finding a solution to a system of *nonlinear equations.* Hence, they are also referred as the *Kuhn-Tucker problem* (KTP).

4.4.1 KTC or KTP

Find vectors $\mathbf{x} \in R^N$, $\mathbf{u} \in R^J$, and $\mathbf{v} \in R^K$ that satisfy

$$\begin{aligned}
\nabla_{(\mathbf{x})} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) &= \nabla f(\mathbf{x}) - \sum_{j=1}^{J} u_j \nabla g_j(\mathbf{x}) - \sum_{k=1}^{K} v_k \nabla h_k(\mathbf{x}) = 0 && (4.6) \\
g_j(\mathbf{x}) &\geq 0 \qquad \forall j = 1, \ldots, J && (4.7) \\
h_k(\mathbf{x}) &= 0 \qquad \forall k = 1, \ldots, K && (4.8) \\
u_j g_j(\mathbf{x}) &= 0 \qquad \forall j = 1, \ldots, J && (4.9) \\
u_j &\geq 0 \qquad \forall j = 1, \ldots, J && (4.10)
\end{aligned}$$

Eq. (4.9) is known as the *complementary slackness condition* in KTP.

Definition 4.5 (Kuhn-Tucker Point) *A* Kuhn-Tucker point *to NLP Problem 2 is a vector* $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ *satisfying Eqs. (4.6) – (4.10)*

Example 4.4

$$\min f(\mathbf{x}) = {x_1}^2 - x_2$$

subject to

$$\begin{aligned}
h_1(\mathbf{x}) &= x_1 + x_2 - 6 = 0 \\
g_1(\mathbf{x}) &= x_1 - 1 \geq 0 \\
g_2(\mathbf{x}) &= 26 - {x_1}^2 - {x_2}^2 \geq 0
\end{aligned}$$

KTP

$$\begin{aligned}
&f(\mathbf{x}) = x_1^2 - x_2 && \nabla^T f(\mathbf{x}) = (2x_1,\ -1) \\
&g_1(\mathbf{x}) = x_1 - 1 && \nabla^T g_1(\mathbf{x}) = (1,\ 0) \\
&g_2(\mathbf{x}) = 26 - x_1^2 - x_2^2 && \nabla^T g_2(\mathbf{x}) = (-2x_1,\ -2x_2) \\
&h_1(\mathbf{x}) = x_1 + x_2 - 6 && \nabla^T h_1(\mathbf{x}) = (1,\ 1)
\end{aligned}$$

- $\nabla L = 0$

$$2x_1 - u_1 + 2x_1 u_2 - v_1 = 0$$

$$-1 + 2x_2 u_2 - v_1 = 0$$

- $g_j \geq 0$

$$x_1 - 1 \geq 0$$

$$26 - x_1^2 - x_2^2 \geq 0$$

- $h_k = 0$

$$x_1 + x_2 - 6 = 0$$

- $u_j g_j = 0$ (complementary slackness condition)

$$u_1(x_1 - 1) = 0$$
$$u_2(26 - x_1^2 - x_2^2) = 0$$

- $u_j \geq 0$

$$u_1 \geq 0$$
$$u_2 \geq 0$$

Solution Although the minimum is located at

$$x_1^* = 1 \quad x_2^* = 5$$

and $f^* = -4$:

$$\nabla g_1(\mathbf{x}^*) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \nabla g_2(\mathbf{x}^*) = \begin{pmatrix} -2 \\ -10 \end{pmatrix} \quad \nabla h_1(\mathbf{x}^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

KTC is given by

- $\nabla L = 0$ (2 eqn's)

$$2 - u_1 + u_2 - v_1 = 0$$
$$-1 + 10u_2 - v_1 = 0$$

- $h_1 = 0$ and
- both g_1 and g_2 are active:

$$g_1 = g_2 = 0$$

Thus we have 5 equations that should be satisfied by 4 unknowns — x_1, x_2, u_1, and u_2 — hencd the KTC is *overspecified.* This is a very unusual case: 3 constraints — one equality constraint and two active inequality constraints — meet at a point X_a on a plain (R^2) as shown in Figure 4.2. As can be seen in the next example of maximizing $f(\mathbf{x})$, KTC can be properly defined for a rather normal situation where two constraints — $h_1 = 0$ and $g_2 = 0$ — meet at a point X_b (also see Figure 4.2).

Example 4.5 *Maximation of $f(\mathbf{x})$ in Example 4.4*

$$x_1^* = 5 \quad x_2^* = 1 \quad f^* = 24$$
$$10 - u_1 + 10u_2 - v_1 = 0$$
$$-1 + u_2 - v_1 = 0$$
$$g_1 = 4 \quad g_2 = h_1 = 0$$
$$4u_1 = 0$$

Solution to KTP:

$$u_1 = 0 \quad \underbrace{u_2 = -\frac{11}{9}}_{\max f} \quad v_1 = -\frac{20}{9}$$

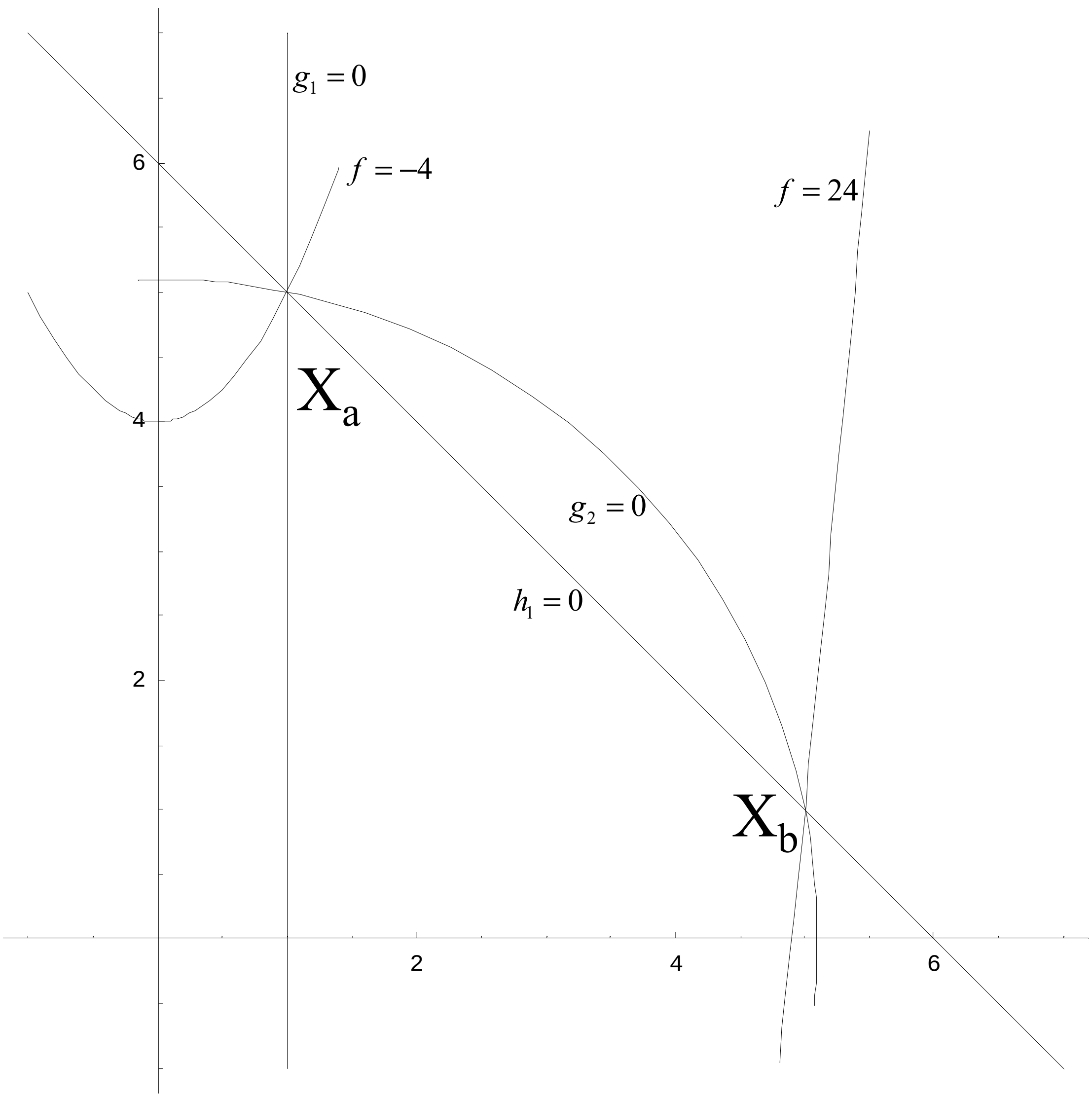

Figure 4.2: Example 4.4

4.4.2 Interpretation of KTC

Lagrangian function

$$L(\mathbf{x}; \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \sum_j u_j g_j(\mathbf{x}) - \sum_k v_k h_k(\mathbf{x})$$

and Eq. (4.6) is the gradient of L with respect to $\mathbf{x}$

Complementary Slackness Condition Since u_j is the shadow price of the constraint $g_j(\mathbf{x}) \geq 0$, it reflects the change in f^* due to the change in b_j when $g_j \geq 0$ is rewritten as $g_j \geq b_j$

- If g_j is inactive ($g_j(\mathbf{x}^*) > 0$), then change in b_j will not results in any change f^* and $u_j = 0$.
- If g_j is active ($g_j(\mathbf{x}^*) = 0$), then increase in b_j means the NLP is more restrictive so that f^* will increase and hence $u_j \geq 0$

Hence at least one of u_j or $g_j(\mathbf{x})$ must be zero at the optimum, which results in the complementary slackness condition, $u_j g_j(\mathbf{x}) = 0$.

4.5 Kuhn-Tucker Theorems

Theorem 4.1 (KT Necessity Theorem) *Consider the NLP Problem 2. Let f, g, and h be differentiable functions and $\mathbf{x}^*$ be a feasible solution to NLP. Let $\mathbf{I} = \{j | g_j(\mathbf{x}^*) = 0\}$. Furthermore, $\nabla g_j(\mathbf{x}^*)$ for $j \in \mathbf{I}$ and $\nabla h_k(\mathbf{x}^*)$ for $k = 1, \ldots, K$ are linearly independent. If $\mathbf{x}^*$ is an optimal solution to NLP, then there exists a $(\mathbf{u}^*, \mathbf{v}^*)$ such that $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ solves the KTP given by Eqs. (4.6) – (4.10).*

Constraint Qualification The condition that $\nabla g_j(\mathbf{x}^*)$ for $j \in \mathbf{I}$ and $\nabla h_k(\mathbf{x}^*)$ for $k = 1, \ldots, K$ are linearly independent at the optimum is known as a *constraint qualification.* For the following NLP problems, the constraint qualification is always satisfied:

- When all the inequality and equality constraints are linear.
- When all the inequality constraints are concave functions and the equality constraints are linear and there exists at least one feasible $\mathbf{x}$ that is strictly inside the feasible region of the inequality constraints. In other words, there exists an $\bar{\mathbf{x}}$ such that $g_j(\bar{\mathbf{x}}) > 0$ for $j = 1, \ldots, J$ and $h_k(\bar{\mathbf{x}}) = 0$ for $k = 1, \ldots, K$.

When the CQ is satisfied at the optimum, there exist a solution to the KTP — If the KTP does not have a solution, either CQ is violated or x^* is not an optimum.

When the CQ is violated at the optimum, there may not exist a solution to the KTP. A few example of CQ being violated are

- In Example 4.4, more than $N-K$ inequality constraints are active at the optimum. KTP possesses multiple solution.
- In Example 4.6, $g_1 = 0$ and $g_3 = 0$ share the same tangent line at the optimum. KTP does not have any solution.

Example 4.6

$$\min f(\mathbf{x}) = (x_1-3)^2 + x_2^2$$

subject to

$$\begin{aligned} g_1(\mathbf{x}) &= (1-x_1)^3 - x_2 \geq 0 \\ g_2(\mathbf{x}) &= x_1 \geq 0 \\ g_3(\mathbf{x}) &= x_2 \geq 0 \end{aligned}$$

From the Figure 4.3,

$$\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$L = (x_1-3)^2 + x_2^2 - u_1[(1-x_1)^3 - x_2] - u_2 x_1 - u_3 x_2$$

KTC is given by

$$\frac{\partial L}{\partial x_1} = 2(x_1-3) + 3u_1(1-x_1)^2 - u_2 = -4 - u_2 = 0 \tag{4.11}$$

$$\frac{\partial L}{\partial x_2} = x_2 + u_1 - u_3 = u_1 - u_3 = 0$$

$$\begin{aligned} u_1 g_1 &= u_1[(1-x_1)^3 - x_2] = u_1 \times 0 = 0 \\ u_2 g_2 &= u_2 x_1 = u_2 \times 1 = 0 \\ u_3 g_3 &= u_3 x_2 = u_3 \times 0 = 0 \end{aligned} \tag{4.12}$$

Both g_1 and g_3 are active but ∇g_1 and ∇g_3 are linearly dependent to each other as shown below

$$\nabla g_1(\mathbf{x}^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad \nabla g_3(\mathbf{x}^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

As a consequence, the KTP is not solvable since Eqs. (4.11) and (4.12) contradict to each other.

Theorem 4.2 (KT Sufficiency Theorem) *Consider the NLP problem given by Eqs. (4.3) – (4.5). Let the objective function $f(\mathbf{x})$ be convex, the inequality constraints $g_j(\mathbf{x})$ be all concave functions for $j = 1, \ldots, J$, and the equality constraints $h_k(\mathbf{x})$ for $k = 1, \ldots, K$ be linear. If there exists a solution $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ that satisfies the KT condition given by Eqs. (4.6) – (4.10), then $\mathbf{x}^*$ is an* optimal *solution to the NLP problem.*

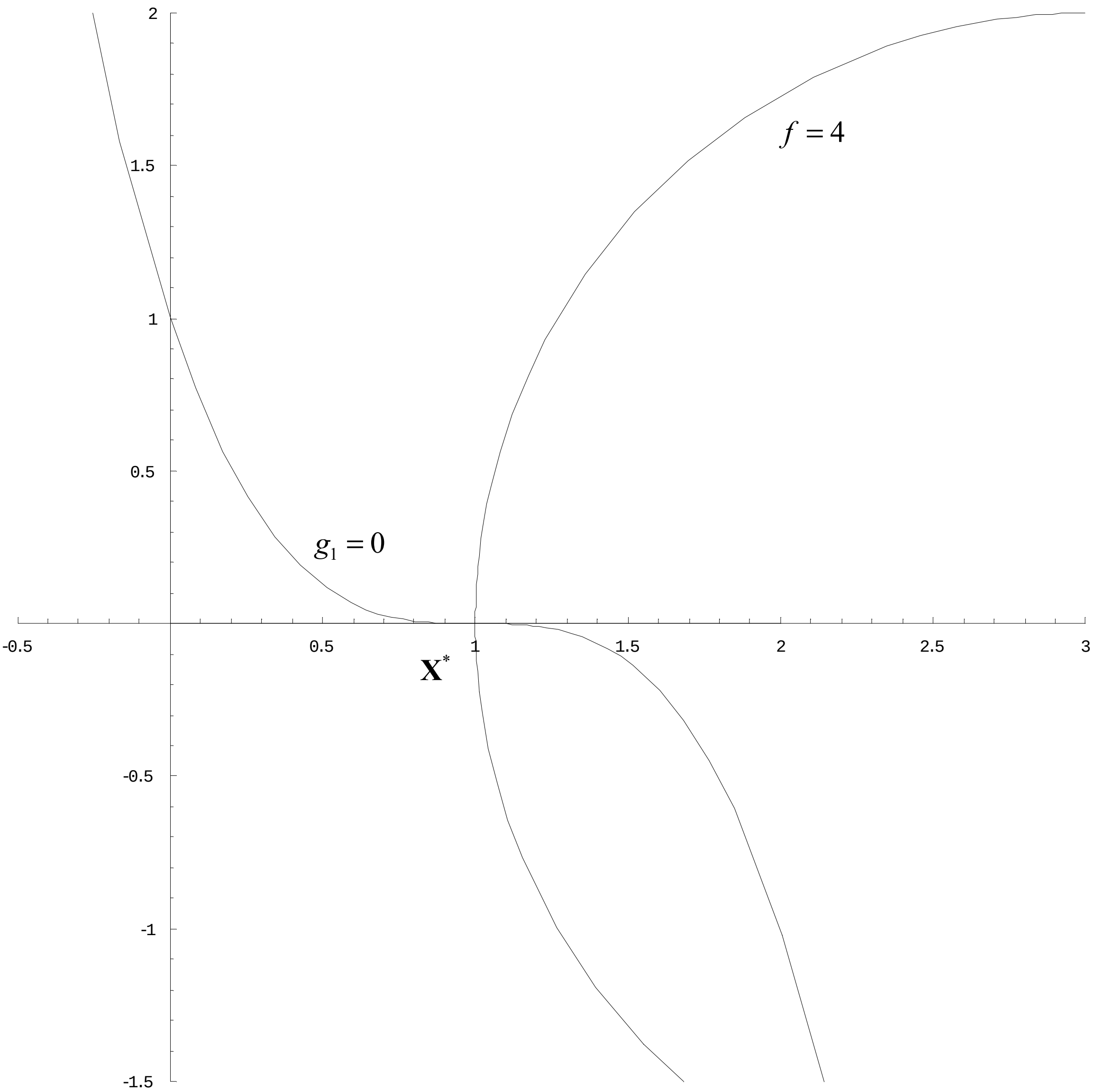

Figure 4.3: Example 4.6

Remarks

1. For practical problems, the constraint qualification will generally hold. If the functions are differentiable, a Kuhn-Tucker point is a possible candidate for the optimum. Hence, many of the NLP methods attempt to converge to a Kuhn-Tucker point.

2. When the sufficiency conditions of Theorem 4.2 hold, a Kuhn-Tucker point automatically becomes global minimum. Unfortunately, the sufficiency conditions are difficult to verify, and often practical problems may not possess these nice property. Note that the presence of one nonlinear equality constraint is enough to violate theassumptions of Theorem 4.2.

3. The sufficiency conditions of Theorem 4.2 have been generalized further to nonconvex inequality constraints, non convex objective functions, and nonlinear equality constraints. These use generalization of convex functions such as quasi-convex and pseudo-convex functions[2]

4.6 Saddlepoint Conditions

Definition 4.6 (Saddlepoint) *A function* $f(\mathbf{x}, \mathbf{y})$ *is said to have a* saddlepoint *at* $(\mathbf{x}^*, \mathbf{y}^*)$ *if* $f(\mathbf{x}^*, \mathbf{y}) \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq f(\mathbf{x}, \mathbf{y}^*)$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$.

$$\min x^2$$

subject to

$$x - 2 \geq 0$$

$$L(x, u) = x^2 - u(x - 2)$$

$$x^* = 2 \quad u^* = 4$$

$$\underbrace{L(2, u)}_{4} \leq \underbrace{L(2, 4)}_{4} \leq \underbrace{L(x, 4)}_{(x-2)^2+4}$$

Kuhn-Tucker saddlepoint problem (KTSP) Find $(\mathbf{x}^*, \mathbf{u}^*)$ such that

$$L(\mathbf{x}^*, \mathbf{u}) \leq L(\mathbf{x}^*, \mathbf{u}^*) \leq L(\mathbf{x}, \mathbf{u}^*)$$

$$\text{all } u_j \geq 0 \text{ and all } \mathbf{x} \in \mathbf{S}$$

where

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) - \sum_j u_j g_j(\mathbf{x})$$

Theorem 4.3 (Sufficient Optimality Theorem) *If* $(\mathbf{x}^*, \mathbf{u}^*)$ *is a saddlepoint solution of a KTSP, then* $\mathbf{x}^*$ *is an optimal solution to the NLP problem.*

[2] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

Remarks

1. No convexity assumptions of the functions have been made in Theorem 4.3.

2. No constraint qualification is invoked.

3. Nonlinear equality constraints of the form $h_k(\mathbf{x}) = 0$ for $k = 1, \ldots, K$ can be handled easily by redefining the Lagrangian function as

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \sum_j u_j g_j(\mathbf{x}) - \sum_k v_k h_k(\mathbf{x})$$

   Here the variables v_k for $k = 1, \ldots, K$ will be unrestricted in sign.

Existence of Saddlepoints There exist necessary optimality theorems that guarantee the existence of a saddlepoint solution without the assumption of differentiability. However, they assume that the constraint qualification is met and that the functions are convex.

Theorem 4.4 (Necessary Optimality Theorem) *Let* $\mathbf{x}^*$ *minimizes* $f(x)$ *subject to* $g_j(\mathbf{x}) \geq 0$, $j = 1, \ldots, J$ *and* $\mathbf{x} \in \mathbf{S}$. *Assume* $\mathbf{S}$ *is a convex set,* $f(\mathbf{x})$ *is a convex function, and* $g_j(\mathbf{x})$ *are concave functions on* $\mathbf{S}$. *Assume also that there exists a point* $\bar{\mathbf{x}} \in \mathbf{S}$ *such that* $g_j(\bar{\mathbf{x}}) > 0$ *for all* $j = 1, \ldots, J$. *Then there exists a vector of multipliers* $\mathbf{u}^* \geq 0$ *such that* $(\mathbf{x}^*, \mathbf{u}^*)$ *is a saddlepoint of the Lagrangian function*

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) - \sum_j u_j g_j(\mathbf{x})$$

satisfying

$$L(\mathbf{x}^*, \mathbf{u}) \leq L(\mathbf{x}^*, \mathbf{u}^*) \leq L(\mathbf{x}, \mathbf{u}^*)$$

for all $\mathbf{x} \in \mathbf{S}$ *and* $\mathbf{u} \geq 0$.

Theorem 4.5 *A solution* $(\mathbf{x}^*, \mathbf{u}^*)$ *with* $\mathbf{u}^* \geq 0$ *and* $\mathbf{x}^* \in \mathbf{S}$ *is a saddlepoint of a KTSP if and only if the following conditions are satisfied:*

1. $\mathbf{x}^*$ *minimizes* $L(\mathbf{x}, \mathbf{u}^*)$ *over all* $\mathbf{x} \in \mathbf{S}$

2. $g_j(\mathbf{x}^*) \geq 0$ *for* $j = 1, \ldots, J$

3. $u_j g_j(\mathbf{x}^*) = 0$ *for* $j = 1, \ldots, J$

4.7 Second-Order Optimality Conditions

Theorem 4.6 (Second-Order Necessity Theorem) *Consider NLP Problem 2. Let* f, g_j, *and* h_k *be twice-differentiable functions, and let* $\mathbf{x}^*$ *be feasible for the NLP. Let the active constraint at* $\mathbf{x}^*$ *be* $\mathbf{I} = \{j | g_j(\mathbf{x}^*) = 0\}$. *Furthermore, assume that* $\nabla g_j(\mathbf{x}^*)$ *for* $j \in \mathbf{I}$ *and* $\nabla h_k(\mathbf{x}^*)$ *for* $k = 1, \ldots, K$ *are linearly independent. Then the* necessary conditions *that* $\mathbf{x}^*$ *be a* local minimum *to NLP Problem 2 are that*

1. *There exists* $(\mathbf{u}^*, \mathbf{v}^*)$ *such that* $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ *is a Kuhn-Tucker point.*

2. *For every vector* $\mathbf{y} \in R^N$ *satisfying*

$$\mathbf{y}^T \nabla g_j(\mathbf{x}^*) = 0 \qquad \textit{for } j \in \mathbf{I}$$

$$\mathbf{y}^T \nabla h_k(\mathbf{x}^*) = 0 \qquad \textit{for } k = 1, \ldots, K$$

it follows that

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)\mathbf{y} \geq 0$$

$\mathbf{H}_L(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ *is the Hessian matrix of the second partial derivatives of* L *with respect to* $\mathbf{x}$ *evaluated at* $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$.

Theorem 4.7 (Second-Order Sufficiency Theorem) *Sufficient conditions that a point* $\mathbf{x}^*$ *is a* strict local minimum *of NLP Problem 2, where* f, g_j, *and* h_k *are twice-differentiable functions, are that*

1. *There exists* $(\mathbf{u}^*, \mathbf{v}^*)$ *such that* $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)$ *is a Kuhn-Tucker point.*

2. *For every nonzero vector* $\mathbf{y} \in \mathbf{R}^N$ *satisfying*

$$\mathbf{y}^T \nabla g_j(\mathbf{x}^*) = 0 \textit{ for } j \in \mathbf{I}_1 = \{j | g_j(\mathbf{x}^*) = 0, u_j^* > 0\}$$

$$\mathbf{y}^T \nabla g_j(\mathbf{x}^*) \geq 0 \textit{ for } j \in \mathbf{I}_2 = \{j | g_j(\mathbf{x}^*) = 0, u_j^* = 0\}$$

$$\mathbf{y}^T \nabla h_k(\mathbf{x}^*) = 0 \textit{ for } k = 1, \ldots, K$$

$$\mathbf{y} \neq \mathbf{0}$$

it follows that

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*)\mathbf{y} > 0$$

4.8 Summary

The necessary and sufficient conditions for $\mathbf{x}^*$ to be a local minimum of NLP Problem 2:

1. The necessary conditions for $\mathbf{x}^*$ to be a local minimum of $f(\mathbf{x})$ are:
   (a) f, g_j, h_k are all twice differentiable at $\mathbf{x}^*$
   (b) The so-called *second-order constraint qualification* holds[3]; the sufficient conditions for this requirement are that the gradient of the binding constraints $(g_j(\mathbf{x}^*) = 0)$, $\nabla g_j(\mathbf{x}^*)$, and the equality constraints, $\nabla h_k(\mathbf{x}^*)$, because $h_k(\mathbf{x}^*) = 0$, are linearly independent.
   (c) The Lagrange multipliers exist; they do if (b) holds.

[3]It contains information about the curvature of the constraints that is taken into account at $\mathbf{x}^*$ as explained in Example 8.4 of Edgar and Himmelblau

(d) The constraints are satisfied at $\mathbf{x}^*$

$$g_j(\mathbf{x}^*) \geq 0$$

$$h_k(\mathbf{x}^*) = 0$$

(e) The Lagrange multipliers u_j^* are not negative

$$u_j^* \geq 0$$

(f) The binding (active) inequality constraints are zero; the inactive inequality constraints are > 0, and the associated u_j's are 0 at $\mathbf{x}^*$

$$u_j^* g_j(\mathbf{x}^*) = 0$$

(g) The Lagrangian function is at a stationary point

$$\nabla L(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*) = 0$$

(h) The Hessian matrix of L is *positive semidefinite* for those $\mathbf{y}$'s forr which $\mathbf{y}^T \nabla g_j(\mathbf{x}^*) = 0$, and $\mathbf{y}^T \nabla h_k(\mathbf{x}^*) = 0$, that is, for all the active constraints

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*) \mathbf{y} \geq 0$$

2. The sufficient conditions for $\mathbf{x}^*$ to be a local minimum are:

   (a) The necessary conditions (a), (b) by implication, (c), (d), (e), (f), and (g)

   (b) Second-order sufficiency theorem 4.7[13]

4.9 Assignments

- Reklaits, *et. al.*'s Chapter 5: Constrained Optimality Criteria [19]
- Edgar & Himmelblau's §8.1 – 8.2 [5]
- Rao's §2.4 – 2.5 [17]

Bibliography

[1] Beveridge, G. S. & R. S. Schechter, 1970, *Optimization: Theory and Practice*, McGraw-Hill.

[2] Biegler, L. T., E. I. Grossmann & A. W. Westerberg, 1997, *Systematic Methods of Chemical Process Design*, Prentice-Hall.

[3] Bryson, Jr., A. E. & Y.-C. Ho, 1975, *Applied Optimal Control: Optimization, Estimation, and Control*, Revised Printing, Hemisphere.

[4] Chong, E. K. P. & S. H. Żak, 1996, *An Introduction to Optimization*, Wiley.

[5] Edgar, T. F. & D. M. Himmelblau, 1989, *Optimization of Chemical Processes*, McGraw-Hill.

[6] Edgeworth, F. Y., 1987, *Mathematical Psychics*, University Microfilms International (Out-of Print Books on Demand, the original edition in 1881).

[7] Floudas, C. A. & P. M. Pardalos (eds.), 2001, *Encyclipedia of Optimization*, Kluwer.

[8] Floudas, C. A., 2000, *Deterministic Global Optimization: Theory, Methods and Applications*, Kluwer.

[9] Floudas, C. A. & P. M. Pardalos (eds.), 2000, *Optimization in Computational Chemistry and Molecular Biology*, Kluwer.

[10] Gelfand, I. M. & S. V. Fomin, 1963, *Calculus of Variation*, Revised English editon translated and edited by R. A. Silverman, Prentice-Hall.

[11] Horst, R. & P. M. Pardolos (eds.), 1995, *Handbook of Global Optimization*, Kluwer.

[12] Horst, R. & H. Tuy, 1996, *Global Optimization: Deterministic Approach*, 3rd ed., Springer.

[13] McCormick, G. P., 1967, "Second order conditions for constrained minima," *SIAM J. Appl. Math.*, **15**(3), 641 – 652.

[14] Miettinen. K. M., 1999, *Nonlinear Multiobjective Optimization*, Kluwer.

[15] Pareto, V., 1964, *Cours d'Economie Politique*, Libraire Droz, Genève (the first edition in 1896).

[16] Pareto, V., 1971, *Manual of Political Economy*, MacMillan Press (the original edition in French in 1927).

[17] Rao, S. S., 1996, *Engineering Optimization: Theory and Practice*, 3rd ed., Wiley.

[18] Ray, W. H., 1981, *Advanced Process Control*, McGraw-Hill.

[19] Reklaitis, G. V., A. Ravindran & K. M. Ragsdell, 1983, *Engineering Optimization: Methods and Applications*, Wiley.

Optimization Journals

- AIAA Journal
- ASCE Journal of Structural Engineering
- ASME Journal of Mechanical Design
- Computers and Chemical Engineering
- Computers and Operations Research
- Computers and Structures
- Engineering Optimization
- International Journal for Numerical Methods in Engineering
- Journal of Optimization Theory and Applications
- Management Sceience
- Operations Research
- SIAM Journal of Optimization
- Structural Optimization