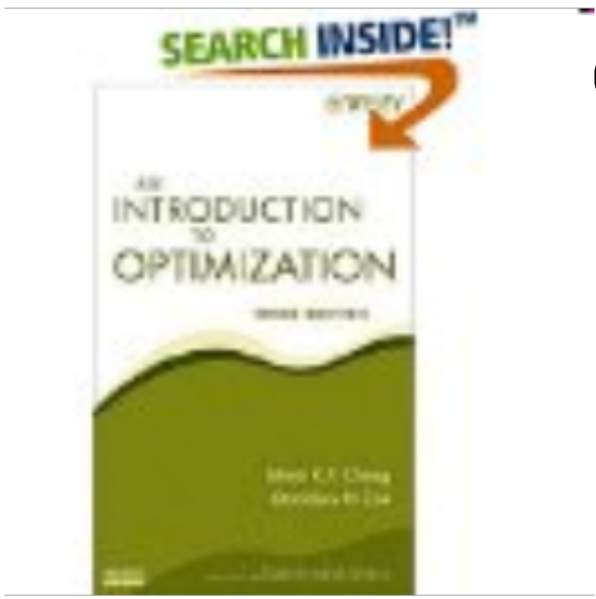
Part III
Nonlinear Constrained Optimization
Chapter 20
PROBLEMS WITH INEQUALITY CONSTRAINTS



Chia-Hui Chang (張嘉惠)
National Central University

2008.1.12

KARUSH-KUHN-TUCKER (KKT) CONDITION

Definition 20.1 An inequality constraint $g_j(\boldsymbol{x}) \leq 0$ is said to be *active* at $\boldsymbol{x}^*$ if $g_j(\boldsymbol{x}^*) = 0$. It is *inactive* at $\boldsymbol{x}^*$ if $g_j(\boldsymbol{x}^*) < 0$. ∎

Definition 20.2 Let $\boldsymbol{x}^*$ satisfy $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0}$, and let $J(\boldsymbol{x}^*)$ be the index set of active inequality constraints, that is,

$$J(\boldsymbol{x}^*) \triangleq \{j : g_j(\boldsymbol{x}^*) = 0\}.$$

Then, we say that $\boldsymbol{x}^*$ is a *regular point* if the vectors

$$\nabla h_i(\boldsymbol{x}^*), \quad \nabla g_j(\boldsymbol{x}^*), \quad 1 \leq i \leq m, \quad j \in J(\boldsymbol{x}^*)$$

are linearly independent. ∎

KKT Theorem

Theorem 20.1 ***Karush-Kuhn-Tucker (KKT) Theorem.*** *Let* $f, h, g \in C^1$. *Let* x^* *be a regular point and a local minimizer for the problem of minimizing* f *subject to* $h(x) = 0$, $g(x) \leq 0$. *Then, there exist* $\lambda^* \in \mathbb{R}^m$ *and* $\mu^* \in \mathbb{R}^p$ *such that*

1. $\mu^* \geq 0$;
2. $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$;
3. $\mu^{*T} g(x^*) = 0$.

□

- Implication
  - If $g_j(x^*) < 0$, then $u_j^* = 0$

min $f(x)$ s.t. $h(x) = 0$, $g(x) \leq 0$

Lagrange

$\mu_1 g_1 + \mu_2 g_2 + \cdots + \mu_p g_p = 0$

$Df(x^*) = -(\mu^{*T} Dg(x^*))$

Example 20.1 A graphical illustration of the Karush-Kuhn-Tucker (KKT) theorem is given in Figure 20.1. In this two-dimensional example, we have only inequality constraints $g_j(\boldsymbol{x}) \leq 0, j = 1, 2, 3$. Note that the point $\boldsymbol{x}^*$ in the figure is indeed a minimizer. The constraint $g_3(\boldsymbol{x}) \leq 0$ is inactive, that is, $g_3(\boldsymbol{x}^*) < 0$; hence $\mu_3^* = 0$. By the KKT theorem, we have

$$\nabla f(\boldsymbol{x}^*) + \mu_1^* \nabla g_1(\boldsymbol{x}^*) + \mu_2^* \nabla g_2(\boldsymbol{x}^*) = \boldsymbol{0},$$

or, equivalently,

$$\nabla f(\boldsymbol{x}^*) = -\mu_1^* \nabla g_1(\boldsymbol{x}^*) - \mu_2^* \nabla g_2(\boldsymbol{x}^*),$$

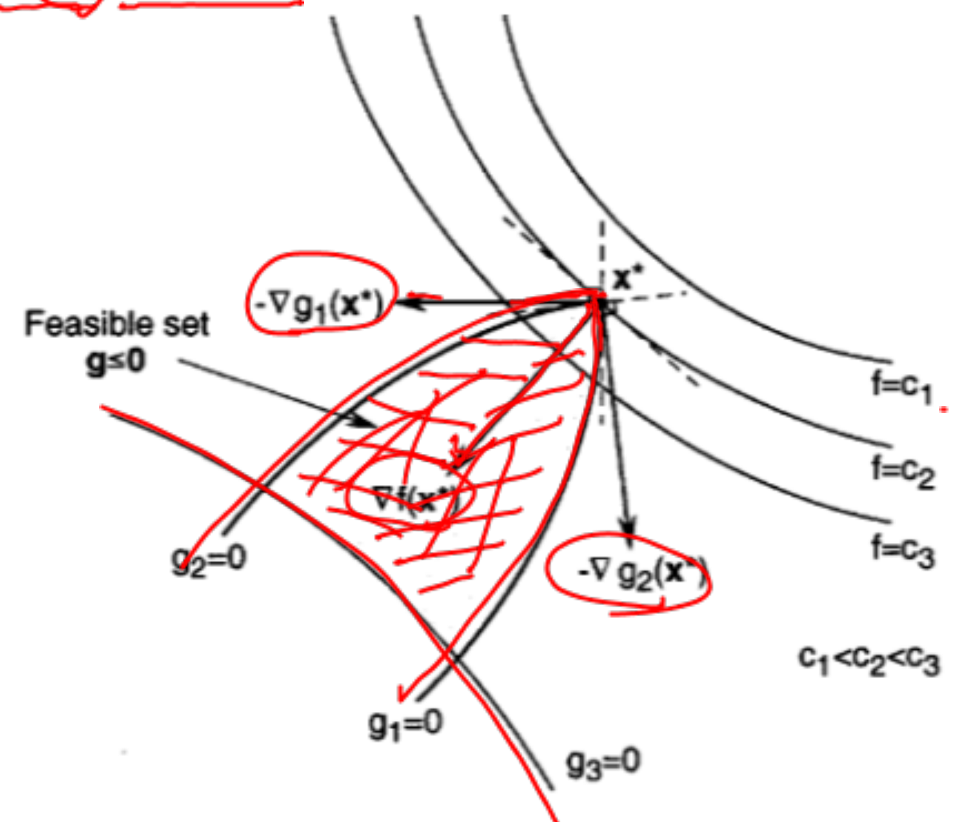


Figure 20.1 Illustration of the Karush-Kuhn-Tucker (KKT) theorem

KKT necessary condition

1. $\boldsymbol{\mu}^* \geq \mathbf{0}$;
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} Dh(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} Dg(\boldsymbol{x}^*) = \mathbf{0}^T$;
3. $\boldsymbol{\mu}^{*T} \boldsymbol{g}(\boldsymbol{x}^*) = 0$;
4. $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$;
5. $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$.

Proof of KKT Theorem

- Let x^* be a regular local minimizer of □ on the set {x: h(x)=0, g(x)□0}
- Then x^* is also regular on the set {x: h(x)=0, $g_j(x)=0$, j□J(x^*)}
- From Lagrange's theorem

$$Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T,$$

- Implication: all j□ J(x^*), we have $u_j^*=0$;

Proof of KKT Theorem (Cont.)

- Show u_j^* □ 0 by contradiction
  - Suppose $u_j^*<0$, then $Dg_j(x^*)y$ □ 0
  - Consider Lagrange condition

$$Df(x^*) + \lambda^{*T} Dh(x^*) + \mu_j^* Dg_j(x^*) + \sum_{i \neq j} \mu_i^* Dg_i(x^*) = 0^T.$$

  - Implication

$$Df(x^*)y = -\mu_j^* Dg_j(x^*)y.$$

$$Df(x^*)y < 0.$$

$$\frac{d}{dt} g_j(x(t^*)) = Dg_j(x^*)y < 0,$$

Because the points $x(t)$, $t \in (t^*, t^* + \min(\delta, \varepsilon)]$, are in $\hat{S}$, they are feasible points with lower objective function values than x^*. This contradicts the assumption that x^* is a local minimizer, and hence the proof is completed. ■

Maximization Problem

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \end{array}$$

min $-f(x)$

1. $\boldsymbol{\mu}^* \geq \mathbf{0}$;
2. $-Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}^T$;
3. $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$;
4. $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$;
5. $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$.

Minimize A similar problem

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \mathbf{0}, \end{array}$$

min $-f(x)$

$-g(x) \leq 0$

1. $\boldsymbol{\mu}^* \leq \mathbf{0}$;
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}^T$;
3. $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$;
4. $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$;
5. $\boldsymbol{g}(\boldsymbol{x}^*) \geq \mathbf{0}$.

Example 20.3 In Figure 20.3, the two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are feasible points, that is, $g(\boldsymbol{x}_1) \geq 0$ and $g(\boldsymbol{x}_2) \geq 0$, and they satisfy the KKT condition.

The point $\boldsymbol{x}_1$ is a maximizer. The KKT condition for this point (with KKT multiplier μ_1) is:

1. $\mu_1 \geq 0$;
2. $\nabla f(\boldsymbol{x}_1) + \mu_1 \nabla g(\boldsymbol{x}_1) = \boldsymbol{0}$;
3. $\mu_1 g(\boldsymbol{x}_1) = 0$;
4. $g(\boldsymbol{x}_1) \geq 0$.

$f(x)$

s.t $g(x) \geq 0$

The point $\boldsymbol{x}_2$ is a minimizer of f. The KKT condition for this point (with KKT multiplier μ_2) is:

1. $\mu_2 \leq 0$;
2. $\nabla f(\boldsymbol{x}_2) + \mu_2 \nabla g(\boldsymbol{x}_2) = \boldsymbol{0}$;
3. $\mu_2 g(\boldsymbol{x}_2) = 0$;
4. $g(\boldsymbol{x}_2) \geq 0$.

Example 20.4 Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) \\ \text{subject to} & x_1, x_2 \geq 0, \end{array}$$

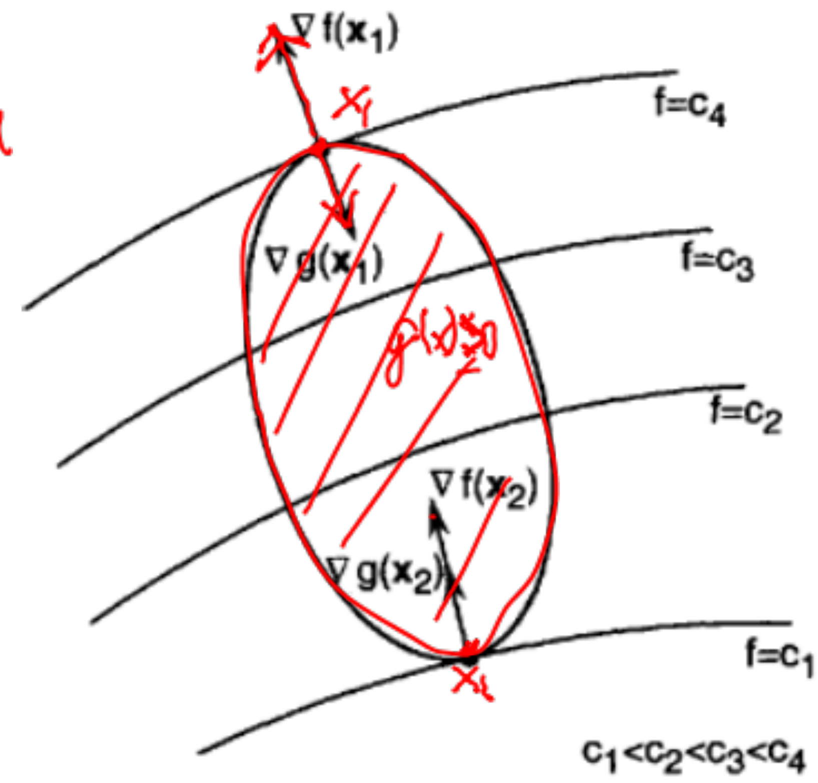
where

$$f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1.$$

The KKT condition for this problem is

1. $\boldsymbol{\mu} = [\mu_1, \mu_2]^T \leq \mathbf{0}$;
2. $Df(\boldsymbol{x}) + \boldsymbol{\mu}^T = \mathbf{0}^T$;
3. $\boldsymbol{\mu}^T \boldsymbol{x} = 0$;
4. $\boldsymbol{x} \geq \mathbf{0}$.

$$Df(\boldsymbol{x}) = [2x_1 + x_2 - 3, \; x_1 + 2x_2].$$



Second Order Conditions

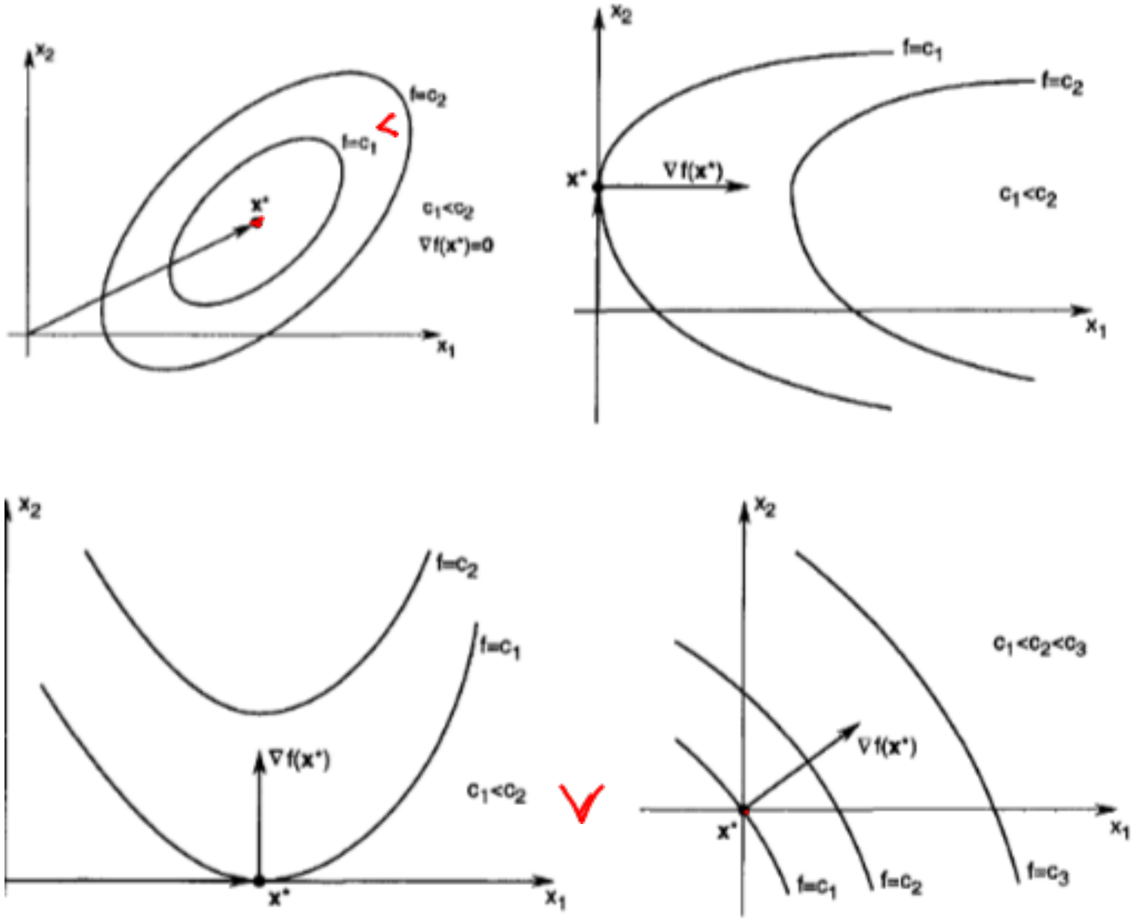


Figure 20.4 Some possible points satisfying the KKT condition for problems with positive constraints (adapted from [9])

$x_1 \geq 0$ $x_2 \geq 0$

Theorem 20.3 ***Second-Order Sufficient Conditions.*** *Suppose* $f, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{C}^2$ *and there exist a feasible point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *and vectors* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *and* $\boldsymbol{\mu}^* \in \mathbb{R}^p$*, such that:*

1. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^T$, $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$; *and*
2. *For all* $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, $\boldsymbol{y} \neq \boldsymbol{0}$, *we have* $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} > 0$.

Then, $\boldsymbol{x}^*$ *is a strict local minimizer of* f *subject to* $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$. □

Proof. For a proof of this theorem, we refer the reader to [64, p. 317]. ■

Example 20.5 We wish to minimize $f(\boldsymbol{x}) = (x_1 - 1)^2 + x_2 - 2$ subject to

$$\begin{aligned} h(\boldsymbol{x}) &= x_2 - x_1 - 1 = 0, \\ g(\boldsymbol{x}) &= x_1 + x_2 - 2 \leq 0. \end{aligned}$$

$Df(x) = [2(x_1 - 1), 1]$

$\mu^T g(x) = 0$

For all $\boldsymbol{x} \in \mathbb{R}^2$, we have

$$Dh(\boldsymbol{x}) = [-1, 1], \quad Dg(\boldsymbol{x}) = [1, 1].$$

Thus, $\nabla h(\boldsymbol{x})$ and $\nabla g(\boldsymbol{x})$ are linearly independent and hence all feasible points are regular. We first write the KKT condition. Because $Df(\boldsymbol{x}) = [2x_1 - 2, 1]$, we have

$$\begin{aligned} Df(\boldsymbol{x}) + \lambda Dh(\boldsymbol{x}) + \mu Dg(\boldsymbol{x}) = [2x_1 - 2 - \lambda + \mu, 1 + \lambda + \mu] &= \mathbf{0}^T \\ \mu(x_1 + x_2 - 2) &= 0 \\ \mu &\geq 0 \\ x_2 - x_1 - 1 &= 0 \\ x_1 + x_2 - 2 &\leq 0. \end{aligned}$$

$$\begin{cases} 2x_1 - 2 - \lambda + \mu = 0 \\ 1 + \lambda + \mu = 0 \\ \mu(x_1 + x_2 - 2) = 0 \\ x_2 - x_1 - 1 = 0 \end{cases}$$