LECTURE NOTE IV

Chapter 5

Constrained Optimality Criteria

• Constrained optimum is not always a stationary point.

5.1 Equality-constrained problems

$$\min_{\mathbf{x}} f(x_1, x_2, \cdots, x_N)$$

subject to $h_k(x_1, x_2, \cdots, x_N) = 0 \quad (k=1,\ldots,K)$

- *Variable elimination method*: Eliminate K variables using the equality constraints. However, the equality constraints should be solved explicitly for a given set of independent variables.

5.2 Lagrange Multipliers

$$\begin{aligned}&\min_{\mathbf{x}} f(x_1, x_2, \cdots, x_N)\\ &\text{subject to } h_1(x_1, x_2, \cdots, x_N) = 0\end{aligned} \quad \Rightarrow \quad \min L(x,v) = f(x) - v h_1(x)$$

$L(x,v)$: *Lagrangian function*

v: *Lagrange multiplier*

- Suppose for a given fixed value of v^*, if x^* satisfies the constraint,

$$\min L(x,v) = \min f(x)$$

- Challenge: How to find v^* so that x^* satisfies the constraint?

- **Example**:

$$\begin{aligned}&\min_{\mathbf{x}} f(\mathbf{x}) = x_1^2 + x_2^2\\ &\text{subject to } h_1(\mathbf{x}) = 2x_1 + x_2 - 2 = 0\end{aligned} \quad \Rightarrow \quad \min L(x,v) = x_1^2 + x_2^2 - v(2x_1 + x_2 - 2)$$

$$\frac{\partial L}{\partial x_1} = 2x_1 - 2v = 0 \quad \Rightarrow \quad x_1^* = v^*$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - v = 0 \quad \Rightarrow \quad x_2^* = v^*/2$$

$$\frac{\partial L}{\partial v} = 2x_1 + x_2 - 2 = 0 \quad \Rightarrow \quad 2v + v/2 = 2 \quad \Rightarrow \quad v^* = 4/5 \quad \Rightarrow \quad x_1^* = 4/5,\ x_2^* = 2/5$$

cf) Hessian of L: $\mathbf{H}_L = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is positive definite and L is a *convex function.*

5.3 Economic Interpretation of Lagrange multipliers

- If possible, **x** can be expressed as functions of **v** and **v** is adjusted to satisfy the constraints.
- Lagrange multiplier as *shadow price*:

$$\min_{\mathbf{x}} f(x_1, x_2) \quad \text{subject to } h_1(x_1, x_2) = b_1 \quad \Rightarrow \quad \min L(\mathbf{x}, v) = f(\mathbf{x}) - v[h_1(x) - b_1]$$

$$\frac{\partial L}{\partial x_1} = \left[\frac{\partial f}{\partial x_1}\right]^* - v^*\left[\frac{\partial h_1}{\partial x_1}\right]^* = 0, \quad \frac{\partial L}{\partial x_2} = \left[\frac{\partial f}{\partial x_2}\right]^* - v^*\left[\frac{\partial h_1}{\partial x_2}\right]^* = 0 \quad \text{and} \quad h_1(x_1^*, x_2^*) = b_1$$

Thus, $\mathbf{x}^*$ is a function of b_1.

$$\frac{\partial f^*}{\partial b_1} = \frac{\partial f^*}{\partial x_1^*}\frac{\partial x_1^*}{\partial b_1} + \frac{\partial f^*}{\partial x_2^*}\frac{\partial x_2^*}{\partial b_1} \quad \text{and} \quad \frac{\partial h_1^*}{\partial x_1^*}\frac{\partial x_1^*}{\partial b_1} + \frac{\partial h_1^*}{\partial x_2^*}\frac{\partial x_2^*}{\partial b_1} - 1 = 0$$

$$\frac{\partial f^*}{\partial b_1} = v^* + \sum_{j=1}^{2}\left[\frac{\partial f^*}{\partial x_j^*} - v^*\frac{\partial h_1^*}{\partial x_j^*}\right]\frac{\partial x_j^*}{\partial b_1} = v^*$$

$\Rightarrow$ v^* is the change in optimal value per unit increase in the right-hand-side constant of the constraint.

5.4 Kuhn-Tucker Conditions

Nonlinear programming (NLP):

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{subject to } h_k(\mathbf{x}) = 0 \quad (k = 1, \cdots, K)$$
$$g_j(\mathbf{x}) \geq 0 \quad (j = 1, \cdots, J)$$

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

- *Kuhn-Tucker condition (KTC)* for optimality

Assume f, g_j, h_k are differentiable.

$$\frac{\partial L}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) - \sum_{j=1}^{J} u_j \nabla g_j(\mathbf{x}) - \sum_{k=1}^{K} v_k \nabla h_k(\mathbf{x}) = 0$$

$$\frac{\partial L}{\partial \mathbf{u}} = \mathbf{g}(\mathbf{x}) \geq \mathbf{0} \quad (\mathbf{g}(\mathbf{x}) - \mathbf{s} = \mathbf{0} \text{ and } \mathbf{s} \geq \mathbf{0})$$

$$\frac{\partial L}{\partial \mathbf{v}} = \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

$u_j g_j(\mathbf{x}) = \mathbf{0}$ for $j = 1, 2, \ldots, J$ (complementary slackness condition)

$\mathbf{u} \geq \mathbf{0}$ (Shadow price $\mathbf{u} = \dfrac{\partial f}{\partial \mathbf{b}_I}$ will not be negative for narrower feasible region.)

- *Kuhn-Tucker problem (KTP)*

  Find vectors **x**, **u**, **v** to satisfy KTC.

5.5 Kuhn-Tucker Theorems

Theorem 5.1 (*Kuhn-Tucker necessity theorem*)

For NLP, let f, $\mathbf{g}$, $\mathbf{h}$ be differentiable function and $\mathbf{x}^*$ be a feasible solution to NLP. Let $I=\{j|g_j(\mathbf{x}^*)=0\}$. Furthermore, $\nabla\mathbf{g}(\mathbf{x}^*)\ \forall j\in I$ and $\nabla\mathbf{h}(\mathbf{x}^*)$ are linearly independent. If $\mathbf{x}^*$ is an optimal solution to the NLP, then there exists a $(\mathbf{u}^*,\mathbf{v}^*)$ such that $(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)$ solves KTP.

- *Constraint qualification (CQ)*: "$\nabla\mathbf{g}(\mathbf{x}^*)\ \forall j\in I$ and $\nabla\mathbf{h}(\mathbf{x}^*)$ are linearly independent"

  The CQ (or Slater CQ) is always satisfied:

  1. when all the inequality and equality constraints are linear
  2. when all the inequality constraints are concave functions and the equality constraints are linear and there exists at least one feasible **x** that is strictly inside the feasible region of the inequality constraints.

  If CQ is not met, there may not exist a solution to KTP.

Theorem 5.2 (*Kuhn-Tucker sufficiency theorem*)

For NLP, let f be convex, $\mathbf{g}$ be all concave functions, and $\mathbf{h}$ be linear. If there exists a solution $(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)$ that satisfies KTC, then $\mathbf{x}^*$ is an *optimal* solution to the NLP.

Remark 1: For practical problems, the CQ 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 Kuhn-Tucker point. (Recall the analogy to the unconstraint optimization case wherein the algorithms attempt to determine a stationary point.)

Remark 2: When the sufficiency conditions of Theorem 5.2 hold, a Kuhn-Tucker point automatically becomes the global minimum. Unfortunately, the sufficiency condition is difficult to verify, and often practical problems may not possess these nice properties. Note that the presence of one nonlinear equality constraint is enough to violate the assumptions of Theorem 5.2.

Remark 3: The sufficiency conditions of Theorem 5.2 have been generalized further to nonconvex inequality constraints, nonconvex objectives, and nonlinear equality constraints. These use generalizations of convex functions such as quasi-convex and pseudo-convex functions.

5.6 Saddle Point Conditions

Definition: A function $f(x,y)$ is said to have a *saddle point* at (x^*,y^*) if $f(x^*,y) \ge f(x^*,y^*) \ge f(x,y^*)$ for all x and y.

- Example:

$f(x,y)=x^2-xy+2y$ and $y \ge 0$ (x is unrestricted)

$$\frac{\partial f}{\partial x} = 2x - y = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = -x + 2 = 0 \quad \Rightarrow \quad x^*=2 \text{ and } y^*=4$$

$f(x^*,y)= 4-2y+2y=0$, $f(x,y^*)= x^2-4x+8=(x-2)^2+4$, and $f(x^*,y^*)=4$

Thus, $f(2,y) \ge f(2,4) \ge f(x,4)$ and f possess a saddle point at (2,4).

- Kuhn-Tucker Saddle point Problem (KTSP)

Find $(\mathbf{x}^*,\mathbf{u}^*)$ such that

$L(\mathbf{x}^*,\mathbf{u}) \le L(\mathbf{x}^*,\mathbf{u}^*) \le L(\mathbf{x},\mathbf{u}^*)$ for all $\mathbf{x}$ and $\mathbf{u} \ge \mathbf{0}$

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

Theorem 5.3 (*Sufficient optimality theorem*)

If $(\mathbf{x}^*,\mathbf{u}^*)$ is a saddle point of a KTSP, then $\mathbf{x}^*$ is an optimal solution to the NLP problem.

Pf) Since $L(\mathbf{x}^*,\mathbf{u}) \le L(\mathbf{x}^*,\mathbf{u}^*) \le L(\mathbf{x},\mathbf{u}^*)$,

$$f(\mathbf{x}^*) - \sum u_j g_j(\mathbf{x}^*) \le f(\mathbf{x}^*) - \sum u_j^* g_j(\mathbf{x}^*) \le f(\mathbf{x}) - \sum u_j^* g_j(\mathbf{x}) \quad \text{for all } \mathbf{u} \ge \mathbf{0}$$

$$f(\mathbf{x}^*) - \sum u_j g_j(\mathbf{x}^*) \le f(\mathbf{x}^*) \quad \text{and } \mathbf{u} \ge \mathbf{0} \quad \Rightarrow \quad \mathbf{g}(\mathbf{x}^*) \ge \mathbf{0} \text{ (feasible)}$$

$$f(\mathbf{x}^*) \le f(\mathbf{x}) - \sum u_j^* g_j(\mathbf{x}) \quad \Rightarrow \quad f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \text{if } \mathbf{g}(\mathbf{x}) \ge \mathbf{0} \text{ (minimum)}$$

Theorem 5.4 (*Necessary optimality theorem*)

Let $\mathbf{x}^*$ minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \ge \mathbf{0}$ and $\mathbf{x} \in S$. Assume S is a convex set, $f(\mathbf{x})$ is a convex function, and $\mathbf{g}(\mathbf{x})$ are concave functions on S. Assume also that there exists a point $\bar{\mathbf{x}} \in S$ such that $\mathbf{g}(\bar{\mathbf{x}}) > \mathbf{0}$. Then there exists a vector of multipliers $\mathbf{u}^* \ge \mathbf{0}$ such that $(\mathbf{x}^*,\mathbf{u}^*)$ is a saddle point of the Lagrangian function

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

satisfying $L(\mathbf{x}^*,\mathbf{u}) \le L(\mathbf{x}^*,\mathbf{u}^*) \le L(\mathbf{x},\mathbf{u}^*)$ for all $\mathbf{x} \in S$ and $\mathbf{u} \ge \mathbf{0}$

Pf) Since $f(\mathbf{x}) > f(\mathbf{x}^*)$, $\mathbf{g}(\mathbf{x}^*) \ge \mathbf{0}$, and $u_j^* g_j(\mathbf{x}^*)=0$ for all $j=1, \ldots, J$,

$$f(\mathbf{x}^*) - \sum u_j g_j(\mathbf{x}^*) \le f(\mathbf{x}^*) = f(\mathbf{x}^*) - \sum u_j^* g_j(\mathbf{x}^*) \quad \Rightarrow \quad L(\mathbf{x}^*,\mathbf{u}) \le L(\mathbf{x}^*,\mathbf{u}^*)$$

From Farkas Lemma, $f(\mathbf{x})-f(\mathbf{x}^*)<0$ and $\mathbf{g}(\mathbf{x}) \ge \mathbf{0}$ with $\mathbf{u}^*$ have no solution if

$$f(\mathbf{x}) - f(\mathbf{x}^*) - \sum u_j^* g_j(\mathbf{x}) \ge 0 \quad \Rightarrow \quad L(\mathbf{x}^*,\mathbf{u}^*) \le L(\mathbf{x},\mathbf{u}^*)$$

Farkas Lemma: For NLP, we assume that the Slater CQ is satisfied. The inequality system $f(\mathbf{x})<0$, $\mathbf{g}_j(\mathbf{x}) = 0$ ($j = 1, \cdots, J$) has no solution if and only if there exists a vector $\mathbf{u}=\mathbf{0}$ such that

$$f(\mathbf{x}) - \sum u_j g_j(\mathbf{x}) \geq 0$$

These systems are called alternative systems, i.e., exactly one of them has a solution.

Theorem 5.5

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

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

ii) $\mathbf{g}(\mathbf{x}) = \mathbf{0}$

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

5.7 Second-Order Optimality Conditions

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{subject to } h_k(\mathbf{x}) = 0 \quad (k = 1, \cdots, K)$$

- The first-order KTC is $\nabla f(\mathbf{x}) - \sum_{k=1}^{K} v_k \nabla h_k(\mathbf{x}) = 0$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$.

- Let $\mathbf{x}^*$ be a Kuhn-Tucker point. Using the Taylor series expansion,

$$\Delta f(\mathbf{x}^*) = f(\mathbf{x}^* + \Delta\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)\Delta\mathbf{x} + 0.5\Delta\mathbf{x}^T \mathbf{H}_f \Delta\mathbf{x} + O_3$$

$$\Delta h_k(\mathbf{x}^*) = h_k(\mathbf{x}^* + \Delta\mathbf{x}) - h_k(\mathbf{x}^*) = \nabla h_k(\mathbf{x}^*)\Delta\mathbf{x} + 0.5\Delta\mathbf{x}^T \mathbf{H}_k \Delta\mathbf{x} + O_3$$

$$\Delta f^* - \sum_k v_k \Delta h_k^* = \left[\nabla f^* - \sum_k v_k \nabla h_k^*\right]\Delta\mathbf{x} + 0.5\,\Delta\mathbf{x}^T \left[\mathbf{H}_f - \sum_k v_k \mathbf{H}_k\right]\Delta\mathbf{x} + O_3$$

For $(\mathbf{x} + \Delta\mathbf{x})$ to be feasible, $\Delta h_k^* = 0$ and assuming the CQ is satisfied at $\mathbf{x}^*$, the KT necessary condition implies that

$$\nabla f^* - \sum_k v_k \nabla h_k^* = 0$$

$$\therefore \Delta f^* \approx 0.5\Delta\mathbf{x}^T \left[\mathbf{H}_f - \sum_k v_k \mathbf{H}_k\right]\Delta\mathbf{x} \geq 0 \quad \text{for } \mathbf{x}^* \text{ to be minimum.}$$

Theorem 5.6 (*Second-order necessity theorem*)

For NLP with equality and inequality constraints, let f, $\mathbf{g}$, $\mathbf{h}$ be twice differentiable functions and $\mathbf{x}^*$ be a feasible solution to NLP. Let $I = \{j | g_j(\mathbf{x}^*)=0\}$. Furthermore,

$\nabla \mathbf{g}(\mathbf{x}^*)\ \forall j \in I$ and $\nabla \mathbf{h}(\mathbf{x}^*)$ are linearly independent. Then the necessary conditions that $\mathbf{x}^*$ be a local minimum to the NLP 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 $\mathbf{y}$ satisfying $\nabla g_j^* \mathbf{y} = 0\ \forall j \in I$ and $\nabla h_k^* \mathbf{y} = 0\ \forall k$,

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)\mathbf{y} \ge 0 \quad \text{where} \quad L(\mathbf{x},\mathbf{u},\mathbf{v}) = f(\mathbf{x}) - \sum_{j=1}^{J} u_j g_j(\mathbf{x}) - \sum_{k=1}^{K} v_k h_k(\mathbf{x}).$$

- Example:

For $\begin{array}{l} \min f(\mathbf{x}) = (x_1-1)^2 + x_2^2 \\ \text{Subject to } g(\mathbf{x}) = -x_1 + x_2^2 \ge 0 \end{array}$, is $\mathbf{x}^*=(0,0)$ optimal?

$$\nabla f(\mathbf{x}) = [2(x_1-1) \quad 2x_2],\ \nabla g(\mathbf{x}) = [-1 \quad 2x_2] \text{ and } I=\{1\}$$

Since $\nabla g(\mathbf{x}^*) = [-1 \quad 0]^T$ is linearly independent, the CQ is satisfied at $\mathbf{x}^*$. The first-order KTC are:

$$2(x_1-1) + u_1 = 0,\ 2x_2 - 2x_2 u_1 = 0,\ u_1(-x_1 + x_2^2) = 0 \text{ and } u_1 \ge 0$$

The solution $(\mathbf{x}^*,\mathbf{u}^*)$=(0,0,2) satisfies these conditions and it is a Kuhn-Tucker point. (a candidate as the optimal solution) Next, check the second-order necessary conditions to test whether it is a local minimum to the NLP problem.

$$\mathbf{H}_L = \begin{bmatrix} 2 & 0 \\ 0 & 2-2u_1 \end{bmatrix} \text{ and } \mathbf{H}_L(x^*,u^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

Consider only vectors $\mathbf{y}$ satisfying $\nabla g_1^* \mathbf{y} = [-1 \quad 0]\mathbf{y} = 0 \Rightarrow \mathbf{y} = [0 \quad y_2]$, verify if

$$[0 \quad y_2]\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ y_2 \end{bmatrix} = -2y_2^2 \ge 0$$

However, this condition is not satisfied unless y_2=0 and this point is not a local minimum for the NLP.

Theorem 5.7 (*Second-order sufficiency theorem*)

For NLP with equality and inequality constraints, let *f*, **g**, **h** be twice differentiable functions and $\mathbf{x}^*$ be a feasible solution to NLP. Let $I=\{j|g_j(\mathbf{x}^*)=0\}$. Then the sufficient conditions that $\mathbf{x}^*$ be a local minimum to the NLP 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 $\mathbf{y}$ satisfying

$$\nabla g_j^* \mathbf{y} = 0\ \forall j \in I_1 = \{j \mid g_j(x^*) = 0, u_j^* > 0\}$$

$$\nabla g_j^* \mathbf{y} \ge 0\ \forall j \in I_2 = \{j \mid g_j(x^*) = 0, u_j^* = 0\} \quad (I_1 \cup I_2 = I)$$

$$\nabla h_k^* \mathbf{y} = 0\ \forall k,$$

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)\mathbf{y} > 0 \quad \text{where} \quad L(\mathbf{x},\mathbf{u},\mathbf{v}) = f(\mathbf{x}) - \sum_{j=1}^{J} u_j g_j(\mathbf{x}) - \sum_{k=1}^{K} v_k h_k(\mathbf{x}).$$

Remark 1: The conditions on $\nabla g_j^* \mathbf{y}$ is equivalent to

$$\nabla f^* \mathbf{y} \le 0 \text{ and } \nabla g_j^* \mathbf{y} \ge 0 \; \forall j \in I .$$

Remark 2: The vector **y** is in the tangent plane of equality constraints. ($\nabla h_k^* \mathbf{y} = 0 \;\; \forall k$)

Remark 3: When the functions are differentiable and the constraints satisfy the CQ, the KTC is the *necessary conditions*.

Remark 4: When the objective function is convex, the inequality constraints are concave, and the equality constraints are linear, the KTC becomes the *sufficient conditions for global optima*.

Remark 5: If the functions were not differentiable, the *saddle point optimality conditions* can be applied.

Remark 6: Since there can be several Kuhn-Tucker points, the *second-order optimality conditions* should be applied, which impose additional restrictions.

Remark 7: The *second-order sufficiency conditions* do not require the convexity of the function and the linearity of the equality constraints.

Chapter 6

Transformation Methods

- NLP problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad (\mathbf{x} \in R^N)$$

$$\text{subject to } h_k(\mathbf{x}) = 0 \quad (k = 1, \cdots, K)$$

$$g_j(\mathbf{x}) \geq 0 \quad (j = 1, \cdots, J)$$

$$\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$$

- Main idea:
  - Solve a constrained optimization problem by solving a sequence of unconstrained optimization problems, and in the limit, the solutions of the unconstrained problems will converge to the solution of the constrained problem.
  - Use an auxiliary function that incorporates the objective function together with "*penalty*" terms that measure violations of the constraints.

- Transformation methods
  - The original constrained problem is *transformed* into a sequence of unconstrained problem via the *penalty function*.
  - If only one unconstrained optimization is required, the penalty function is *exact*.
  - *Interior point method*: the sequence $\mathbf{x}^{(k)}$ contains feasible points. (*Barrier method*)
  - *Exterior point method*: the sequence $\mathbf{x}^{(k)}$ contains infeasible points. (*Penalty method*)
  - *Mixed point method*: the sequence $\mathbf{x}^{(k)}$ contains both feasible and infeasible points.

6.1 The Penalty Concept

- The *penalty function*

$$P(\mathbf{x}, \mathbf{R}) = f(\mathbf{x}) + \Omega(\mathbf{R}, \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}))$$

  - $\mathbf{R}$: a set of *penalty parameters*
  - Ω : *penalty term*
  - $\Omega = 0$ if $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and $\Omega > 0$ if $\mathbf{g}(\mathbf{x}) < \mathbf{0}$ and/or $\mathbf{h}(\mathbf{x})$? $\mathbf{0}$.

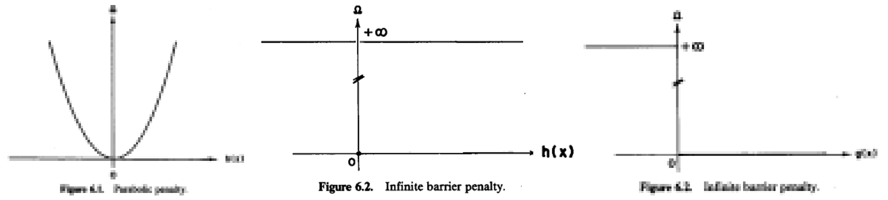
- The transformation methods to be useful:
  - The subproblem solution should approach a solution of NLP, that is,

$$\lim_{k \to T < \infty} x^{(k)} = x^*$$

- The problem of minimizing $P(\mathbf{x},\mathbf{R})$ should be similar in difficulty to minimizing $f(\mathbf{x})$. That is, the method will be less than useful if the unconstrained problems are excessively difficult to solve, no matter how strong the theoretical basis of convergence.
- $\mathbf{R}^{(k+1)}=F(\mathbf{R}^{(k)})$ should be simple. It seems reasonable to hope that the calculation overhead associated with updating the penalty parameters should be small compared to the effort associated with solving the unconstrained subproblems. (Note: This may in fact not be desirable for problems with very complex objective and constraint functions. In this case, considerable effort updating the penalty parameters may be justified.)

• The various *penalty terms*

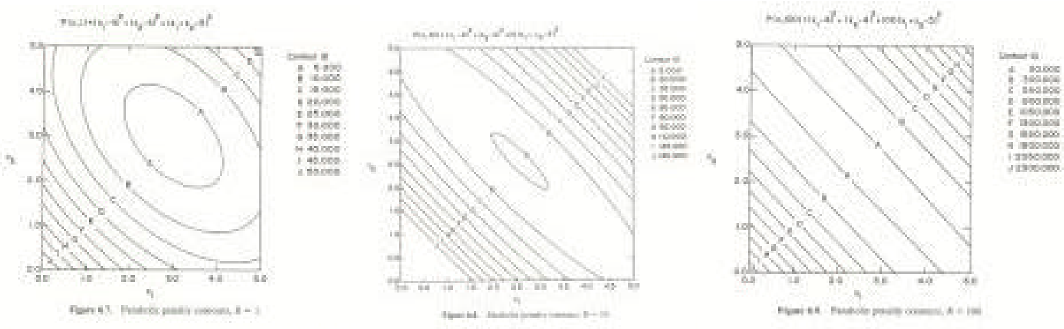
- For equality constraints (penalty methods)



1. Parabolic penalty (quadratic-loss penalty): exterior

$$W = R\mathbf{h}(\mathbf{x})^T\mathbf{h}(\mathbf{x}) \quad \text{or} \quad W = R\left\|\mathbf{h}(\mathbf{x})\right\|_l \quad (R>0,\ l=0)$$

- Initially, use small R so that the initial subproblem can be solved easily.
- Gradually increase R so that the violation gets penalized.



2. Infinite penalty: exterior

$$W = \begin{cases} \infty & \text{if } \mathbf{h}(\mathbf{x}) \neq 0 \\ 0 & \text{if } \mathbf{h}(\mathbf{x})=0 \end{cases}$$

- Simplest form but discontinuous along the boundary
- Practically, assign large number instead of infinity

Remark: Approximation of Lagrange multiplier from penalty method

Lagrange multiplier: $v_j = \nabla f(\mathbf{x}) / \nabla h_j(\mathbf{x})$

Penalty function: $P(\mathbf{x}, R) = f(\mathbf{x}) + 0.5R\sum_{j=1}^{J}[h_j(\mathbf{x})]^2$

$$\nabla P(\mathbf{x}, R) = \nabla f(\mathbf{x}) + R\sum_{j=1}^{J} h_j(\mathbf{x})[\nabla h_j(\mathbf{x})] = 0$$

$$\Rightarrow v_j = -\lim_{R\to\infty} Rh_j(\mathbf{x}(R))$$

Example:

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

$$\text{subject to } h(\mathbf{x}) = x_1 + 2x_2 - 4 = 0$$

Using parabolic penalty function

$$\min P(\mathbf{x}) = -x_1 x_2 + 0.5R(x_1 + 2x_2 - 4)^2$$

$$\frac{\partial P(\mathbf{x})}{\partial x_1} = -x_2 + R(x_1 + 2x_2 - 4) = 0$$

$$\frac{\partial P(\mathbf{x})}{\partial x_2} = -x_1 + 2R(x_1 + 2x_2 - 4) = 0$$

$$\Rightarrow x_1(R) = \frac{8R}{4R-1} \text{ and } x_2(R) = \frac{4R}{4R-1}$$

$$\Rightarrow h(R) = x_1(R) + 2x_2(R) - 4 = \frac{4}{4R-1}$$

If $R \to \infty$, x_1=2 and x_2=1 (Solution to constrained problem)

(Lagrange multiplier: $u(R) = -Rh(x(R)) = -\frac{4R}{4R-1} = -1$)

- For inequality constraints

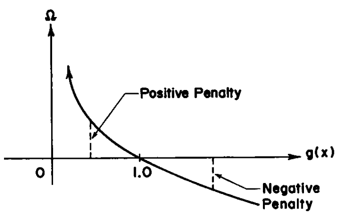


Figure 6.3. Log penalty.

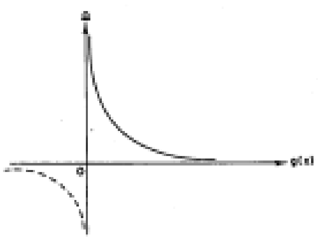


Figure 6.4. Inverse penalty.

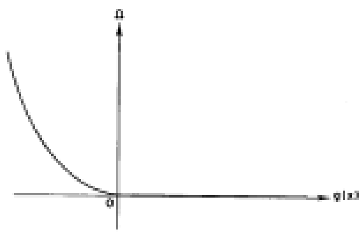


Figure 6.5. Bracket operator.

1. *Logarithmic barrier function* (*Logarithmic penalty*) : interior

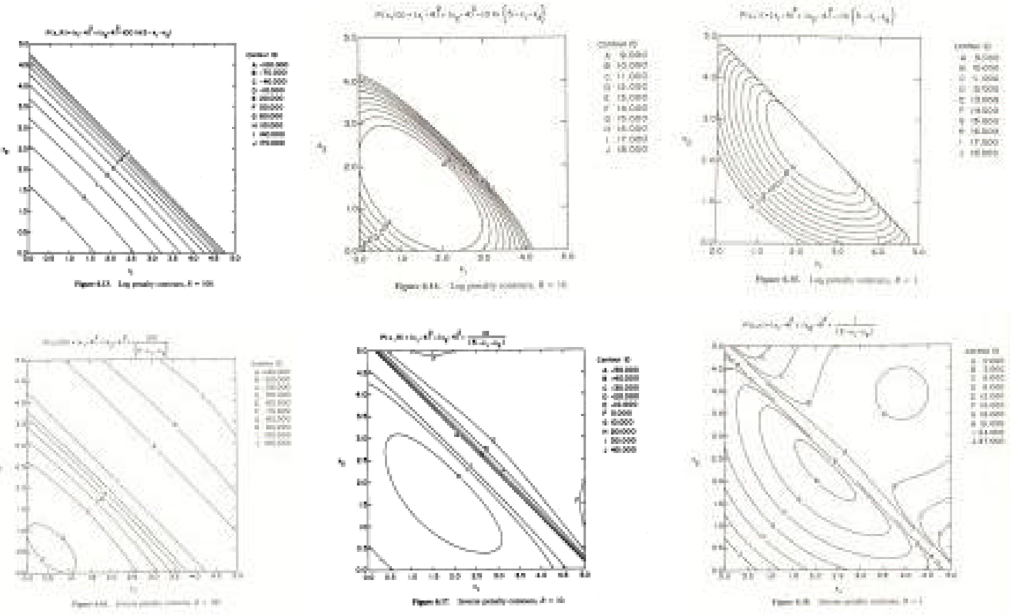
$$W = -R\sum_j \ln g_j(\mathbf{x}) \quad (0 < g_j(\mathbf{x}) < 1) \quad \text{otherwise,} \quad W = 0 \quad (g_j(\mathbf{x}) > 1)$$

2. *Inverse barrier function* (*Inverse penalty*) : interior

$$W = R\sum_j [1/g_j(\mathbf{x})] \quad (g_j(\mathbf{x}) > 0)$$

- Start at a feasible solution so that inequalities are not violated
- Discontinuous near the boundary

- Recovery procedure is needed if the search point is an exterior point
- Initially, use large R so that the initial subproblem can be solved easily.
- Gradually, decrease R monotonically toward zero so that the penalty term for violation becomes zero at the optimal solution.



Remark: Why not use small R from the beginning?

- No matter how small the R is, the solution of the penalty function is not the optimal solution and it is very difficult to solve
- Start from large R (easier to solve) and decrease R and repeatedly solve the subproblem by using the previous solution.

Remark: Approximation of Lagrange multiplier from logarithmic barrier method

Lagrange multiplier: $u_j = \nabla f(\mathbf{x}) / \nabla g_j(\mathbf{x})$

Penalty function: $P(\mathbf{x}, R) = f(\mathbf{x}) + R\sum_{j=1}^{J} \ln[g_j(\mathbf{x})]$

$$\nabla P(\mathbf{x}, R) = \nabla f(\mathbf{x}) + R\sum_{j=1}^{J} (1/g_j(\mathbf{x}))[\nabla g_j(\mathbf{x})] = 0$$

$$\Rightarrow v_j = \lim_{R \to 0} R / g_j(\mathbf{x}(R))$$

Example:

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

subject to $g_1(\mathbf{x}) = 1 + x_1 - 2x_2^2 \geq 0$ and $g_2(\mathbf{x}) = x_2 \geq 0$

Using logarithmic barrier function

$\min P(\mathbf{x}) = x_1 - 2x_2 - R[\ln(1 + x_1 - 2x_2^2) + \ln(x_2)]$

$$\frac{\partial P(\mathbf{x})}{\partial x_1} = 1 - R\frac{1}{(1+x_1-2x_2^2)} = 0$$

$$\frac{\partial P(\mathbf{x})}{\partial x_2} = -2 + R\frac{2x_2}{(1+x_1-2x_2^2)} - R\frac{1}{x_2} = 0$$

$$\Rightarrow -2 + 2x_2 - R/x_2 = 0 \text{ and } x_2^2 - x_2 - R/2 = 0$$

$$\Rightarrow x_2(R) = (1+\sqrt{1+2R})/2 \text{ and } x_1(R) = (3R - 1 + \sqrt{1+2R})/2$$

If $R \to 0$, x_1=0 and x_2=1 (Solution to constrained problem)

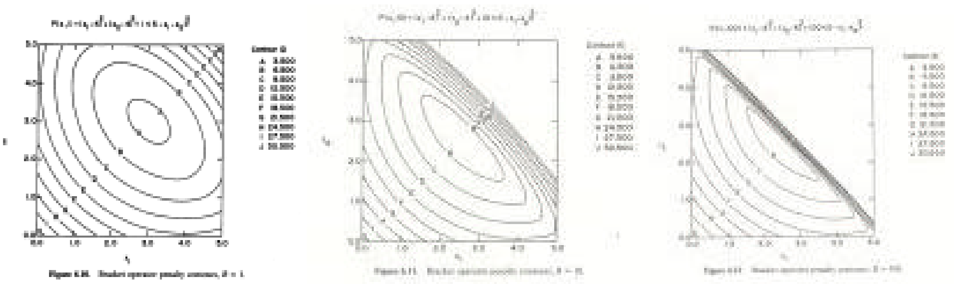
$$v_1(R) = \lim_{R\to 0} R/g_1(R) = \lim_{R\to 0} 2R/[2+(3R-1+\sqrt{1+2R})-(1+R+\sqrt{1+2R})] = 1$$

$$v_2(R) = \lim_{R\to 0} R/g_2(R) = \lim_{R\to 0} 2R/(1+\sqrt{1+2R}) = 0$$

3. *Bracket penalty* (*quadratic-loss penalty*) : mixed

$$W = 0.5R\sum_j [\min(g_j(\mathbf{x}),0)]^2$$

- First derivative continuous, but second derivative is discontinuous (not suitable for Newton type optimization)
- Initially, choose R as small positive and
- Gradually increase R so that the violation gets penalized.



- The convergence
  - Consider only the barrier methods (penalty methods can be analyzed in a similar way) applied to the problem

$$\min f(\mathbf{x})$$
$$\text{subject to } g(\mathbf{x}) \geq 0$$

Let S and S^0 denote, respectively, the feasible region and its interior, i.e.

$$S = \{\mathbf{x} \mid g_i(\mathbf{x}) \geq 0, i = 1, \cdots m\}$$

$$S^0 = \{\mathbf{x} \mid g_i(\mathbf{x}) > 0, i = 1, \cdots m\}$$

Assumptions

1) f and all g_i (i=1,…,m) are continuous. W is a continuous function on S^0, and $W \to +\infty$ when x approaches the boundary of S.

2) For any a , the set $\{x \mid x \in S, f(x) \le a\}$ is bounded.

3) S^0 is not empty.

4) Any $y \in S$ can be approached by a sequence $\{x_k\}$ in $x_k \to y$, $x_k \in S^0$.

- **Theorem** (Convergence of the Barrier Methods)

Let $P(\mathbf{x}, R) = f(\mathbf{x}) + RW(\mathbf{x})$; $R_1 \ge R_2 \ge \cdots \ge \lim_{k\to\infty} R_k = 0$ and

x_k is a global minimizer of problem $\min_{x \in S^0} P(x, R_k)$. Then, for $k = 1, 2, ...$,

a) $P(x_{k+1}, R_{k+1}) \le P(x_k, R_k)$

b) $W(x_{k+1}) \ge W(x_k)$

c) $f(x_{k+1}) \le f(x_k)$

d) $f(x^*) \le f(x_k) \le P(x_k, R_k)$

e) If the subsequence $\{x_k \mid k \in K\}$ converges to $\hat{x}$, then $\hat{x}$ must be an optimal solution of the inequality constrained problem.

Pf)

a) Since x_i is a global minimizer of problem $\min_{x \in S^0} P(x, R_i)$,

$$P(x_{k+1}, R_{k+1}) = f(x_{k+1}) + R_{k+1} W(x_{k+1}) \le f(x_k) + R_{k+1} W(x_k) \qquad (R_k = R_{k+1})$$
$$\le f(x_k) + R_k W(x_k) = P(x_k, R_k)$$

b) Summing $f(x_k) + R_k W(x_k) \le f(x_{k+1}) + R_k W(x_{k+1})$ and $f(x_{k+1}) + R_{k+1} W(x_{k+1}) \le f(x_k) + R_{k+1} W(x_k)$, then by rearranging

$(R_k - R_{k+1}) W(x_k) \le (R_k - R_{k+1}) W(x_{k+1})$

$W(x_k) \le W(x_{k+1})$ $\quad (R_k - R_{k+1} = 0)$

c) From a), $f(x_{k+1}) \le f(x_{k+1}) + R_{k+1}(W(x_{k+1}) - W(x_k)) \le f(x_k)$

d) $f(x^*) \le f(x_k) \le f(x_k) + R_k W(x_k) = P(x_k, R_k)$

e) Let $\hat{x}$ be the limit of $\{x_k\}$ that satisfies constraints. For any $e > 0$, there exist $\hat{x}$ such that $f(x^*) + e \ge f(\hat{x})$ for $x \in S^0$.

$$f(x^*) + e + R_\infty W(\hat{x}) \ge f(\hat{x}) + R_\infty W(\hat{x}) = \lim_{k\to\infty} P(x_k, R_k) \ge f(x^*)$$

$$f(x^*) + ae \ge \lim_{k\to\infty} P(x_k, R_k) \ge f(x^*) \ (a > 0) \Rightarrow \lim_{k\to\infty} P(x_k, R_k) = f(x^*)$$

Since $f(x^*) \le f(\hat{x})$ and

$$\lim_{k\to\infty} P(x_k, R_k) = f(\hat{x}) + R_k W(\hat{x}) = f(x^*) \Rightarrow f(\hat{x}) \le f(x^*) \ (\because R_k W(\hat{x}) \ge 0)$$

$$\therefore f(x^*) \le f(\hat{x}) \le f(x^*) \Rightarrow f(\hat{x}) = f(x^*)$$

- Exact Penalty Method

  The idea in an exact penalty method is to choose a penalty function $W(x)$ and a constant R so that the optimal solution x^* of $P(x,R)$ is also an optimal solution of the original problem.

  Theorem:

  Suppose a constrained optimization problem is a convex program for which the Karush-Kuhn-Tucker conditions are necessary.

  Suppose that $W(x)=\sum_{j=1}^{J}\min(g_j(x),0)+\sum_{i=1}^{I}|h_i(x)|$

  Then as long as R is chosen sufficiently large, the sets of optimal solutions of $P(x,R)$ and the original problem coincide. In fact, it suffices to choose $R > \max(u_j^*; v_i^*)$, where (u^*, v^*) is a vector of Karush-Kuhn-Tucker multipliers.

  Remark: Unfortunately, the resulting problem with exact penalty (large penalty parameter) generates quite ill-conditioned problem which is very difficult to solve.

6.3 Method of Multipliers (MOM)

- The standard penalty approach generates progressively ill-conditioned subproblems which limits the utility of the method for practical applications.
- The fixed parameter penalty methods has been suggested
- Huard's method of centers:

$$P(x,R)=[R-f(x)]\prod_{j=1}^{J} g_j(x)$$

where R is a moving truncation at each maximization stage, say, $R_k = f(x_{k-1})$.

This is equivalent to *parameterless penalty form* for minimization

$$P(x) = -\ln[f(x_{k-1})-f(x)]-\sum_{j=1}^{J} g_j(x) \quad \text{(Parameter-free method)}$$

- MOM

$$P(x,s^{(k)},t^{(k)})=f(x)+R\sum_{j=1}^{J}\left\{\min(g_j(x)+s_j^{(k)},0)^2-[s_j^{(k)}]^2\right\}$$

$$+R\sum_{i=1}^{I}\left\{\left[h_i(x)+t_i^{(k)}\right]^2-\left[t_i^{(k)}\right]^2\right\}$$

where $s_j^{(k+1)}=\min(g_j(x^{(k)})+s_i^{(k)},0) \quad (j=1,\cdots,J)$

$$t_i^{(k+1)}=h_i(x^{(k)})+t_i^{(k)} \quad (i=1,\cdots,I)$$

- R is constant for each iteration.
- The method will terminate if $x^{(k)}$ fails to change.

Remark:

1. $\lim\limits_{k \to K < \infty} x^{(k)} = x^*$ (a Kuhn-Tucker point)
2. Minimization of $P(x)$ is similar in difficulty to $\min f(x)$ fro reasonable value of R.
3. The updating rules for s and t are simple, requiring essentially no additional calculation.

Example:

$\min f(x) = (x_1 - 4)^2 + (x_2 - 4)^2$

subject to $h(x) = x_1 + x_2 - 5 = 0$

$$P(x) = (x_1 - 4)^2 + (x_2 - 4)^2 + \frac{1}{R}(x_1 + x_2 - 5 + t)^2 - \frac{1}{R}t^2$$

$$\frac{\partial P}{\partial x_1} = 2(x_1 - 4) + \frac{2}{R}(x_1 + x_2 - 5 + t) = 0$$

$$\frac{\partial P}{\partial x_2} = 2(x_2 - 4) + \frac{2}{R}(x_1 + x_2 - 5 + t) = 0$$

$$\Rightarrow x_1 = x_2 = \frac{5 + 4R - t}{2 + R}$$

For convenience, choose $R = 1$ to get $\Rightarrow x_1 = x_2 = 3 - \frac{t}{3}$

Start with $t^{(0)} = 0$, $t^{(1)} = h(3,3) + t^{(0)} = 1$,

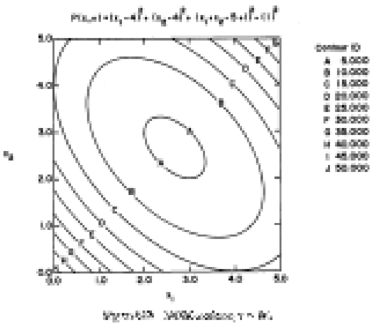
$t^{(2)} = h(2.6667, 2.6667) + t^{(1)} = 1.333$

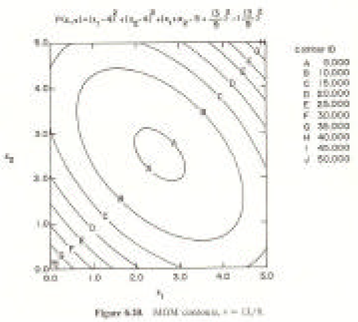
$t^{(3)} = h(2.5555, 2.5555) + t^{(2)} = 1.444$

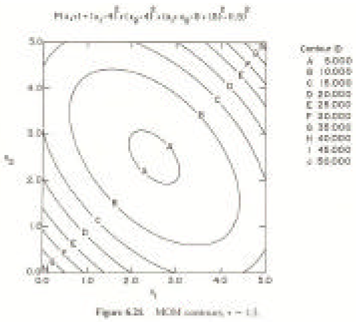
$t^{(4)} = h(2.5185, 2.5185) + t^{(3)} = 1.4818$

$t^{(5)} = h(2.5062, 2.5062) + t^{(4)} = 1.5$

$t^{(6)} = h(2.5, 2.5) + t^{(5)} = 1.5$ (Stop) $\Rightarrow x_1 = x_2 = 2.5$







- MOM type codes: ACDPAC, GAPFPR, VF01A, SALQDR, SALQDF, SALMNF, …