Penalty and Barrier Methods for Constrained Optimization

Robert M. Freund

February, 2004

1 Introduction

Consider the constrained optimization problem P:

$$\begin{array}{rll} P: & \underset{x}{\text{minimize}} & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & & h_i(x) = 0, \quad i = 1, \ldots, k \\ & & x \in \Re^n, \end{array}$$

whose feasible region we denote by

$$\mathcal{F} := \{x \in \Re^n \mid g_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, k\}.$$

Barrier and penalty methods are designed to solve P by instead solving a sequence of specially constructed unconstrained optimization problems.

In a penalty method, the feasible region of P is expanded from $\mathcal{F}$ to all of $\Re^n$, but a large cost or "penalty" is added to the objective function for points that lie outside of the original feasible region $\mathcal{F}$.

In a barrier method, we presume that we are given a point x^o that lies in the interior of the feasible region $\mathcal{F}$, and we impose a very large cost on *feasible* points that lie ever closer to the boundary of $\mathcal{F}$, thereby creating a "barrier" to exiting the feasible region.

2 Penalty Methods

Consider the constrained optimization problem P:

$$\begin{array}{rll} P: & \underset{x}{\text{minimize}} & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & & h_i(x) = 0, \quad i = 1, \ldots, k \\ & & x \in \Re^n. \end{array}$$

By converting the constraints "$h_i(x) = 0$" to "$h_i(x) \leq 0, -h_i(x) \leq 0$", we can assume that P is of the form

$$
\begin{array}{lll}
P: & \underset{x}{\text{minimize}} & f(x) \\
& \text{s.t.} & g(x) \leq 0, \\
& & \\
& & x \in \Re^n,
\end{array}
$$

where we write $g(x) := (g_1(x), \ldots, g_m(x))^T$ for convenience.

Definition: A function $p(x) : \Re^n \to \Re$ is called a *penalty function* for P if $p(x)$ satisfies:

- $p(x) = 0$ if $g(x) \leq 0$ and
- $p(x) > 0$ if $g(x) \not\leq 0$.

Example:

$$
p(x) = \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2.
$$

We then consider solving the following *penalty program*:

$$
\begin{array}{lll}
P(c): & \underset{x}{\text{minimize}} & f(x) + cp(x) \\
& \text{s.t.} & x \in \Re^n
\end{array}
$$

for an increasing sequence of constants c as $c \to +\infty$. Note that in the program $P(c)$ we are assigning a penalty to the violated constraints. The scalar quantity c is called the *penalty parameter*.

Let $c_k \geq 0, k = 1, \ldots, \infty$, be a sequence of penalty parameters that satisfies $c_{k+1} > c_k$ for all k and $\lim_{k\to\infty} c_k = +\infty$. Let $q(c, x) = f(x) + cp(x)$, and let x^k be the exact solution to the program $P(c_k)$, and let x^* denote any optimal solution of P.

The following Lemma presents some basic properties of penalty methods:

Lemma 2.1 (Penalty Lemma).

1. $q(c_k, x^k) \leq q(c_{k+1}, x^{k+1})$
2. $p(x^k) \geq p(x^{k+1})$
3. $f(x^k) \leq f(x^{k+1})$
4. $f(x^*) \geq q(c_k, x^k) \geq f(x^k)$

Proof:

1. We have
$$q(c_{k+1}, x^{k+1}) = f(x^{k+1}) + c_{k+1}p(x^{k+1}) \geq f(x^{k+1}) + c_k p(x^{k+1})$$
$$\geq f(x^k) + c_k p(x^k) = q(c_k, x^k)$$

2. $$f(x^k) + c_k p(x^k) \leq f(x^{k+1}) + c_k p(x^{k+1})$$
and
$$f(x^{k+1}) + c_{k+1}p(x^{k+1}) \leq f(x^k) + c_{k+1}p(x^k)$$
Thus $(c^{k+1} - c^k)p(x^k) \geq (c^{k+1} - c^k)p(x^{k+1})$, whereby $p(x^k) \geq p(x^{k+1})$.

3. From the proof of (1.),
$$f(x^{k+1}) + c_k p(x^{k+1}) \geq f(x^k) + c_k p(x^k).$$
But $p(x^k) \geq p(x^{k+1})$, which implies that $f(x^{k+1}) \geq f(x^k)$.

4. $f(x^k) \leq f(x^k) + c_k p(x^k) \leq f(x^*) + c_k p(x^*) = f(x^*)$.

q.e.d.

The next result concerns convergence of the penalty method.

Theorem 2.1 (Penalty Convergence Theorem). *Suppose that $f(x)$, $g(x)$, and $p(x)$, are continuous functions. Let $\{x^k\}, k = 1, \ldots, \infty$, be a sequence of solutions to $P(c^k)$. Then any limit point $\bar{x}$ of $\{x^k\}$ solves P .*

Proof: Let $\bar{x}$ be a limit point of $\{x^k\}$. From the continuity of the functions involved, $\lim_{k\to\infty} f(x^k) = f(\bar{x})$. Also, from the Penalty Lemma,

$$q^* := \lim_{k\to\infty} q(c_k, x^k) \leq f(x^*).$$

Thus

$$\lim_{k\to\infty} c_k p(x^k) = \lim_{k\to\infty} \left[q(c_k, x^k) - f(x^k)\right] = q^* - f(\bar{x}).$$

But $c_k \to \infty$, which implies from the above that

$$\lim_{k\to\infty} p(x^k) = 0.$$

Therefore, from the continuity of $p(x)$ and $g(x)$, $p(\bar{x}) = 0$, and so $g(\bar{x}) \leq 0$, that is, $\bar{x}$ is a feasible solution of P. Finally, from the Penalty Lemma, $f(x^k) \leq f(x^*)$ for all k, and so $f(\bar{x}) \leq f(x^*)$, which implies that $\bar{x}$ is an optimal solution of P.
q.e.d.

An often-used class of penalty functions is:

$$p(x) = \sum_{i=1}^{m} \left[\max\{0, g_i(x)\}\right]^q, \quad \text{where} \quad q \geq 1. \tag{1}$$

We note the following:

- If $q = 1$, $p(x)$ in (1) is called the "linear penalty function". This function may not be differentiable at points where $g_i(x) = 0$ for some i.
- Setting $q = 2$ is the most common form of (1) that is used in practice, and is called the "quadratic penalty function" (for obvious reasons).

- If we denote

$$g^+(x) = (\max\{0, g_i(x)\}, \ldots, \max\{0, g_m(x)\})^T,$$

then the quadratic penalty function can be written as

$$p(x) = (g^+(x))^T(g^+(x)).$$

2.1 Karush-Kuhn-Tucker Multipliers in Penalty Methods

The penalty function $p(x)$ is in actuality a function only of $g^+(x)$, where $g^+(x) = \max\{0, g_i(x)\}$ (the nonnegative part of $g_i(x)$), $i = 1, \ldots, m$. Then we can write $p(x) = \gamma(g^+(x))$, where $\gamma(y)$ is a function of $y \in (\Re^m)^+$.

Two examples of this type of penalty function are

$$\gamma(y) = \sum_{i=1}^{m} y_i,$$

which corresponds to the linear penalty function, and

$$\gamma(y) = \sum_{i=1}^{m} y_i^2,$$

which corresponds to the quadratic penalty function.

Note that even if $\gamma(y)$ is continuously differentiable, $p(x)$ might not be continuously differentiable, since $g^+(x)$ is not differentiable at points x where $g_i^+(x) = 0$ for some i. However, if we assume the following:

$$\frac{\partial \gamma(y)}{\partial y_i} = 0 \quad \text{at} \quad y_i = 0, \quad i = 1, \ldots, m, \tag{2}$$

then $p(x)$ is differentiable whenever the functions $g_i(x)$ are differentiable, $i = 1, \ldots, m$, and we can write

$$\nabla p(x) = \sum_{i=1}^{m} \frac{\partial \gamma(g^+(x))}{\partial y_i} \nabla g_i(x). \tag{3}$$

Now let x^k solve $P(c_k)$. Then x^k will satisfy

$$\nabla f(x^k) + c_k \nabla p(x^k) = 0,$$

that is,

$$\nabla f(x^k) + c_k \sum_{i=1}^{m} \frac{\partial \gamma(g^+(x^k))}{\partial y_i} \nabla g_i(x^k) = 0.$$

Let us define

$$u_i^k = c_k \frac{\partial \gamma(g^+(x^k))}{\partial y_i}. \tag{4}$$

Then

$$\nabla f(x^k) + \sum_{i=1}^{m} u_i^k \nabla g_i(x^k) = 0,$$

and so we can interpret the u^k as a sort of vector of Karush-Kuhn-Tucker multipliers. In fact, we have:

Lemma 2.2 *Suppose $\gamma(y)$ is continuously differentiable and satisfies (2), and that $f(x)$ and $g(x)$ are differentiable. Let u^k be definied by (4). Then if $x^k \to \bar{x}$, and $\bar{x}$ satisfies the linear independence condition for gradient vectors of active constraints, then $u^k \to \bar{u}$, where $\bar{u}$ is a vector of Karush-Kuhn-Tucker multipliers for the optimal solution $\bar{x}$ of P.*

Proof: From the Penalty Convergence Theorem, $\bar{x}$ is an optimal solution of P. Let $I = \{i \mid g_i(\bar{x}) = 0\}$ and $N = \{i \mid g_i(\bar{x}) < 0\}$. For $i \in N$, $g_i(x^k) < 0$ for all k sufficiently large, so $u_i^k = 0$ for all k sufficiently large, whereby $\bar{u}_i = 0$ for $i \in N$.

From (4) and the definition of a penalty function, it follows that $u_i^k \geq 0$ for $i \in I$, for all k sufficiently large, .

Suppose $u^k \to \bar{u}$ as $k \to \infty$. Then $\bar{u}_i = 0$ for $i \in N$. From the continuity of all functions involved,

$$\nabla f(x^k) + \sum_{i=1}^{m} u_i^k \nabla g(x^k) = 0$$

implies

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g(\bar{x}) = 0.$$

From the above remarks, we also have $\bar{u} \geq 0$ and $\bar{u}_i = 0$ for all $i \in N$. Thus $\bar{u}$ is a vector of Karush-Kuhn-Tucker multipliers. It therefore remains to show that $u^k \to \bar{u}$ for some unique $\bar{u}$.

Suppose $\{u^k\}_{k=1}^{\infty}$ has no accumulation point. Then $\|u^k\| \to \infty$. But then define

$$v^k = \frac{u^k}{\|u^k\|},$$

and then $\|v^k\| = 1$ for all k, and so the sequence $\{v^k\}_{k=1}^{\infty}$ has some accumulation point $\bar{v}$. For all $i \in N$, $v_i^k = 0$ for k large, whereby $\bar{v}_i = 0$ for $i \in N$, and

$$\sum_{i \in I} v_i^k \nabla g_i(x^k) = \sum_{i=1}^{m} v_i^k \nabla g_i(x^k) = \sum_{i=1}^{m} \left(\frac{u_i^k}{\|u^k\|} \right) \nabla g_i(x^k) = \frac{-\nabla f(x^k)}{\|u^k\|}$$

for k large. As $k \to \infty$, we have $x^k \to \bar{x}$, $v^k \to \bar{v}$, and $\|u^k\| \to \infty$, and so the above becomes

$$\sum_{i \in I} \bar{v}_i \nabla g_i(\bar{x}) = 0,$$

and $\|\bar{v}\| = 1$, which violates the linear independence condition. Therefore $\{u^k\}$ is a bounded sequence, and so has at least one accumulation point.

Now suppose that $\{u^k\}$ has two accumulation points, $\tilde{u}$ and $\bar{u}$. Note $\bar{u}_i = 0$ and $\tilde{u}_i = 0$ for $i \in N$, and so

$$\sum_{i \in I} \bar{u}_i \nabla g_i(\bar{x}) = -\nabla f(\bar{x}) = \sum_{i \in I} \tilde{u}_i \nabla g_i(\bar{x}),$$

so that

$$\sum_{i \in I} (\bar{u}_i - \tilde{u}_i) \nabla g_i(\bar{x}) = 0.$$

But by the linear independence condition, $\bar{u}_i - \tilde{u}_i = 0$ for all $i \in I$, and so $\bar{u}_i = \tilde{u}_i$. This then implies that $\bar{u} = \tilde{u}$.

q.e.d.

Remark. The quadratic penalty function satisfies the condition (2), but that the linear penalty function does not satisfy (2).

2.2 Exact Penalty Methods

The idea in an *exact penalty method* is to choose a penalty function $p(x)$ and a constant c so that the optimal solution $\tilde{x}$ of $P(c)$ is also an optimal solution of the original problem P.

Theorem 2.2 *Suppose P is a convex program for which the Karush-Kuhn-Tucker conditions are necessary. Suppose that*

$$p(x) := \sum_{i=1}^{m} (g_i(x))^+.$$

Then as long as c is chosen sufficiently large, the sets of optimal solutions of $P(c)$ and P coincide. In fact, it sufficies to choose $c > \max_i\{u_i^*\}$, where u^* is a vector of Karush-Kuhn-Tucker multipliers.

Proof: Suppose $\hat{x}$ solves P. For any $x \in R^n$ we have:

$$\begin{aligned}
q(c,x) = f(x) + c\sum_{i=1}^{m} (g_i(x))^+ &\geq f(x) + \sum_{i=1}^{m} (u_i^* g_i(x))^+ \\
&\geq f(x) + \sum_{i=1}^{m} u_i^* g_i(x) \\
&\geq f(x) + \sum_{i=1}^{m} u_i^*(g_i(\hat{x}) + \nabla g_i(\hat{x})^T(x-\hat{x})) \\
&= f(x) + \sum_{i=1}^{m} u_i^* \nabla g_i(\hat{x})^T(x-\hat{x}) \\
&= f(x) - \nabla f(\hat{x})^T(x-\hat{x}) \\
&\geq f(\hat{x}) = f(\hat{x}) + c\sum_{i=1}^{m} (g_i(\hat{x}))^+ = q(c,\hat{x}).
\end{aligned}$$

Thus $q(c,\hat{x}) \leq q(c,x)$ for all x, and therefore $\hat{x}$ solves $P(c)$.

Next suppose that $\bar{x}$ solves $P(c)$. Then if $\hat{x}$ solves P, we have:

$$f(\bar{x}) + c\sum_{i=1}^{m}(g_i(\bar{x}))^+ \leq f(\hat{x}) + c\sum_{i=1}^{m}(g_i(\hat{x}))^+ = f(\hat{x}),$$

and so

$$f(\bar{x}) \leq f(\hat{x}) - c\sum_{i=1}^{m}(g_i(\bar{x}))^+. \tag{5}$$

However, if $\bar{x}$ is not feasible for P, then

$$\begin{aligned} f(\bar{x}) &\geq f(\hat{x}) + \nabla f(\hat{x})^T(\bar{x} - \hat{x}) \\ &= f(\hat{x}) - \sum_{i=1}^{m} u_i^* \nabla g_i(\hat{x})^T(\bar{x} - \hat{x}) \\ &\geq f(\hat{x}) + \sum_{i=1}^{m} u_i^*(g_i(\hat{x}) - g_i(\bar{x})) \\ &= f(\hat{x}) - \sum_{i=1}^{m} u_i^* g_i(\bar{x}) > f(\hat{x}) - c\sum_{i=1}^{m}(g_i(\bar{x}))^+, \end{aligned}$$

which contradicts (5). Thus $\bar{x}$ is feasible for P. That being the case, $f(\bar{x}) \leq f(\hat{x}) - c\sum_{i=1}^{m}(g_i(\bar{x}))^+ = f(\hat{x})$ from (5), and so $\bar{x}$ solves P.
q.e.d.

2.3 Penalty Methods for Inequality and Equality Constraints

The presentation of penalty methods has assumed either that the problem P has no equality constraints, or that the equality constraints have been converted to inequality constraints. For the latter, the conversion is easy to do, but the conversion usually violates good judgement in that it unnecessarily complicates the problem. Furthermore, it can cause the linear independence condition to be automatically violated for every feasible solution. Therefore, instead let us consider the constrained optimization problem P with both inquality and equality constraints:

$$
\begin{array}{rll}
P: & \underset{x}{\text{minimize}} & f(x) \\
 & \text{s.t.} & g(x) \le 0, \\
 & & h(x) = 0, \\
 & & x \in \Re^n,
\end{array}
$$

where $g(x)$ and $h(x)$ are vector-valued functions, that is, $g(x) := (g_1(x), \ldots, g_m(x))^T$ and $h(x) := (h_1(x), \ldots, h_k(x))^T$ for notational convenience.

Definition: A function $p(x) : \Re^n \to \Re$ is called a *penalty function* for P if $p(x)$ satisfies:

- $p(x) = 0$ if $g(x) \le 0$ and $h(x) = 0$
- $p(x) > 0$ if $g(x) \not\le 0$ or $h(x) \ne 0$.

The main class of penalty functions for this general problem are of the form:

$$p(x) = \sum_{i=1}^{m} [\max\{0, g_i(x)\}]^q + \sum_{i=1}^{k} |h_i(x)|^q, \quad \text{where} \quad q \ge 1.$$

All of the results of this section extend naturally to problems with equality constraints and for penalty functions of the above form. For example, in the analogous result of Theorem 2.2, it suffices to choose $c > \max\{u_1^*, \ldots, u_m^*, |v_1^*|, \ldots, |v_k^*|\}$.

3 Barrier Methods

Definition. A *barrier function* for P is any function $b(x) : \Re^n \to \Re$ that satisfies

- $b(x) \ge 0$ for all x that satisfy $g(x) < 0$, and

- $b(x) \to \infty$ as $\lim_x \max_i\{g_i(x)\} \to 0$.

The idea in a barrier method is to dissuade points x from ever approaching the boundary of the feasible region. We consider solving:

$$\begin{array}{lll} B(c): & \text{minimize} & f(x) + \frac{1}{c}b(x) \\ & \text{s.t.} & g(x) < 0, \\ & & x \in \Re^n. \end{array}$$

for a sequence of $c_k \to +\infty$. Note that the constraints "$g(x) < 0$" are effectively unimportant in $B(c)$, as they are never binding in $B(c)$.

Example:

$$b(x) = \sum_{i=1}^{m} \frac{1}{-g_i(x)}$$

Suppose $g(x) = (x - 4, 1 - x)^T, \quad x \in \Re^1$. Then

$$b(x) = \frac{1}{4-x} + \frac{1}{x-1}.$$

Let $r(c, x) = f(x) + \frac{1}{c}b(x)$. Let the sequence $\{c_k\}$ satisfy $c_{k+1} > c_k$ and $c_k \to \infty$ as $k \to \infty$. Let x^k denote the exact solution to $B(c^k)$.

The following Lemma presents some basic properties of barrier methods.

Lemma 3.1 (Barrier Lemma).

1. $r(c_k, x^k) \geq r(c_{k+1}, x^{k+1})$

2. $b(x^k) \leq b(x^{k+1})$

3. $f(x^k) \geq f(x^{k+1})$

4. $f(x^*) \le f(x^k) \le r(c_k, x^k)$.

Proof:

1.
$$r(c_k, x^k) = f(x^k) + \frac{1}{c_k} b(x^k) \ge f(x^k) + \frac{1}{c_{k+1}} b(x^k)$$
$$\ge f(x^{k+1}) + \frac{1}{c_{k+1}} b(x^{k+1}) = r(c_{k+1}, x^{k+1})$$

2.
$$f(x^k) + \frac{1}{c_k} b(x^k) \le f(x^{k+1}) + \frac{1}{c_k} b(x^{k+1})$$

and

$$f(x^{k+1}) + \frac{1}{c_{k+1}} b(x^{k+1}) \le f(x^k) + \frac{1}{c_{k+1}} b(x^k).$$

Summing and rearranging, we have

$$\left(\frac{1}{c_k} - \frac{1}{c_{k+1}}\right) b(x^k) \le \left(\frac{1}{c_k} - \frac{1}{c_{k+1}}\right) b(x^{k+1}).$$

Since $c_k < c_{k+1}$, it follows that $b(x^{k+1}) \ge b(x^k)$.

3. From the proof of (1.),

$$f(x^k) + \frac{1}{c_{k+1}} b(x^k) \ge f(x^{k+1}) + \frac{1}{c_{k+1}} b(x^{k+1}).$$

But from (2.), $b(x^k) \le b(x^{k+1})$. Thus $f(x^k) \ge f(x^{k+1})$.

4. $f(x^*) \le f(x^k) \le f(x^k) + \frac{1}{c_k} b(x^k) = r(c_k, x^k)$.

q.e.d.

Let $N(\epsilon, x)$ denote the ball of radius ϵ centered at the point x. The next result concerns convergence of the barrier method.

Theorem 3.1 (Barrier Convergence Theorem). *Suppose $f(x)$, $g(x)$, and $b(x)$ are continuous functions. Let $\{x^k\}$, $k = 1, \ldots, \infty$, be a sequence of solutions of $B(c^k)$. Suppose there exists an optimal solution x^* of P for which $N(\epsilon, x^*) \cap \{x \mid g(x) < 0\} \neq \emptyset$ for every $\epsilon > 0$. Then any limit point $\bar{x}$ of $\{x^k\}$ solves P.*

Proof: Let $\bar{x}$ be any limit point of the sequence $\{x^k\}$. From the continuity of $f(x)$ and $g(x)$, $\lim_{k\to\infty} f(x^k) = f(\bar{x})$ and $\lim_{k\to\infty} g(x^k) = g(\bar{x}) \leq 0$. Thus $\bar{x}$ is feasible for P.

For any $\epsilon > 0$, there exists $\tilde{x}$ such that $g(\tilde{x}) < 0$ and $f(\tilde{x}) \leq f(x^*)+\epsilon$. For each k,

$$f(x^*) + \epsilon + \frac{1}{c_k} b(\tilde{x}) \geq f(\tilde{x}) + \frac{1}{c_k} b(\tilde{x}) \geq r(c_k, x^k).$$

Therefore for k sufficiently large, $f(x^*)+2\epsilon \geq r(c_k, x^k)$, and since $r(c_k, x^k) \geq f(x^*)$ from (4.) of the Barrier Lemma, then

$$f(x^*) + 2\epsilon \geq \lim_{k\to\infty} r(c_k, x^k) \geq f(x^*).$$

This implies that

$$\lim_{k\to\infty} r(c_k, x^k) = f(x^*).$$

We also have

$$f(x^*) \leq f(x^k) \leq f(x^k) + \frac{1}{c_k} b(x^k) = r(c_k, x^k).$$

Taking limits we obtain

$$f(x^*) \leq f(\bar{x}) \leq f(x^*),$$

whereby $\bar{x}$ is an optimal solution of P.
q.e.d.

A typical class of barrier functions are:

$$b(x) = \sum_{i=1}^{m} (-g_i(x))^{-q}, \quad \text{where} \quad q > 0.$$

3.1 Karush-Kuhn-Tucker Multipliers in Barrier Methods

Let

$$b(x) = \gamma(g(x)),$$

where $\gamma(y) : \Re^m \rightarrow \Re$, and assume that $\gamma(y)$ is continuously differentiable for all $y < 0$. Then

$$\nabla b(x) = \sum_{i=1}^{m} \frac{\partial \gamma(g(x))}{\partial y_i} \nabla g_i(x),$$

and if x^k solves $B(c_k)$, then $\nabla f(x^k) + \frac{1}{c_k} \nabla b(x^k) = 0$, that is,

$$\nabla f(x^k) + \frac{1}{c_k} \sum_{i=1}^{m} \frac{\partial \gamma(g(x^k))}{\partial y_i} \nabla g_i(x^k) = 0. \tag{6}$$

Let us define

$$u_i^k = \frac{1}{c_k} \frac{\partial \gamma(g(x^k))}{\partial y_i}. \tag{7}$$

Then (6) becomes:

$$\nabla f(x^k) + \sum_{i=1}^{m} u_i^k \nabla g_i(x^k) = 0. \tag{8}$$

Therefore we can interpret the u^k as a sort of vector of Karush-Kuhn-Tucker multipliers. In fact, we have:

Lemma 3.2 *Let P satisfy the conditions of the Barrier Convergence Theorem. Suppose $\gamma(y)$ is continuously differentiable and let u^k be defined by (7). Then if $x^k \to \bar{x}$, and $\bar{x}$ satisfies the linear independence condition for gradient vectors of active constraints, then $u^k \to \bar{u}$, where $\bar{u}$ is a vector of Karush-Kuhn-Tucker multipliers for the optimal solution $\bar{x}$ of P.*

Proof:

Let $x^k \to \bar{x}$ and let $I = \{i \mid g_i(\bar{x}) = 0\}$ and $N = \{i \mid g_i(\bar{x}) < 0\}$. For all $i \in N$,

$$u_i^k = \frac{1}{c_k} \frac{\partial \gamma(g(x_k))}{\partial y_i} \to 0 \quad \text{as} \quad k \to \infty,$$

since $c_k \to \infty$ and $g_i(x^k) \to g_i(\bar{x}) < 0$, and $\frac{\partial \gamma(g(\bar{x}))}{\partial y_i}$ is finite. Also $u_i^k \geq 0$ for all i, and k sufficiently large.

Suppose $u^k \to \bar{u}$ as $k \to \infty$. Then $\bar{u} \geq 0$, and $\bar{u}_i = 0$ for all $i \in N$. From the continuity of all functions involved, (8) implies that

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) = 0, \quad \bar{u} \geq 0, \quad \bar{u}^T g(\bar{x}) = 0.$$

Thus $\bar{u}$ is a vector of Kuhn-Tucker multipliers. It remains to show that $u^k \to \bar{u}$ for some unique $\bar{u}$. The proof that $u^k \to \bar{u}$ for some unique $\bar{u}$ is exactly as in Lemma 2.2.
q.e.d.

4 Exercises

1. Show that the penalty method problem can be formulated as follows:

$$\text{find} \quad \sup_{c \geq 0} \inf_{x \in X} \{f(x) + cp(x)\} \; .$$

Next show that this problem is equivalent to:

$$\text{find} \quad \inf_{x \in X} \sup_{c \geq 0} \{f(x) + cp(x)\} \; .$$

2. To use a barrier function method, we must find a point $x \in X$ that satisfies $g_i(x) < 0, i = 1, \ldots, m$. Consider the following procedure for computing such a point:

   Initialization: Select $x_1 \in X$, set $k \leftarrow 1$

   Main Iteration:

   Step 1. Let $I := \{i \mid g_i(x_k) < 0\}$. If $I = \{1, \ldots, m\}$, stop with x_k satisfying $g_i(x) < 0, i = 1, \ldots, m$. Otherwise, select $j \notin I$ and go to Step 2.

   Step 2. Use a barrier function method to solve the following problem starting from x_k:

$$\begin{array}{ll} \text{minimize}_x & g_j(x) \\ \\ \text{s.t.} & g_i(x) \leq 0 \text{ for } i \in I \\ \\ & x \in X \ . \end{array}$$

   Let x_{k+1} be an optimal solution of this problem. If $g_j(x_{k+1}) \geq 0$, stop; there is no solution x satisfying $x \in X$ and $g_i(x) < 0, i = 1, \ldots, m$. Otherwise, set $k \leftarrow k + 1$ and repeat Step 1.

   Show that the above procedure stops in at most m main iterations with a point $x \in X$ that satisfies $g_i(x) < 0, i = 1, \ldots, m$, or with the valid conclusion that no such point exists.

3. Consider the problem to minimize $c^t x$ subject to $Ax = b$ and $x \geq 0$, where $A \in \Re^{m \times n}$ and $\text{rank}(A) = m$. Suppose we are given a feasible solution $\bar{x} > 0$ and consider the following *logarithmic barrier* problem parameterized by the barrier parameter :

$$\begin{array}{lll} B(\theta): & \text{minimize}_x & c^T x - \theta \sum_{j=1}^{n} \ln(x_j) \\ \\ & \text{s.t.} & Ax = b \\ & & x > 0. \end{array}$$

   Let us denote the objective function of $B(\theta)$ by $f_\theta(x)$.

(a) Consider the second-order Taylor series approximation of $f_\theta(x)$ at $x = \bar{x}$:

$$PT: \quad \text{minimize}_d \quad f_\theta(\bar{x}) + \nabla f_\theta(\bar{x})^T d + \tfrac{1}{2} d^T \nabla^2 f_\theta(\bar{x}) d$$

$$\text{s.t.} \qquad Ad = 0 \ .$$

Let $\bar{X}$ denote the $n \times n$ diagonal matrix whose diagonal entries are precisely the components of $\bar{x}$:

$$\bar{X} := \begin{pmatrix} \bar{x}_1 & 0 & \dots & 0 \\ 0 & \bar{x}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{x}_n \end{pmatrix} .$$

Using this notation, what is the formula for $\nabla f_\theta(\bar{x})$? for $\nabla^2 f_\theta(\bar{x})$?

(b) Derive a closed form solution for the solution $\hat{d}$ of PT.

(c) The standard linear optimization dual problem is given by:

$$D: \quad \text{maximize}_{\pi,s} \quad b^T \pi$$

$$\text{s.t.} \qquad A^T \pi + s = c$$
$$s \geq 0.$$

Now consider the following problem:

$$PR: \quad \text{minimize}_{\pi,s} \quad \|\tfrac{1}{\theta} \bar{X} s - e\|$$

$$\text{s.t.} \qquad A^T \pi + s = c \ .$$

Construct a closed form solution $(\hat{\pi}, \hat{s})$ to PR, and show that the direction of part (3b) satisfies:

$$\hat{d} = \bar{x} - \frac{1}{\theta} \bar{X}^2 \hat{s} \ .$$