SUFFICIENCY OF EXACT PENALTY MINIMIZATION

by

O. L. Mangasarian

Computer Sciences Technical Report #508

August 1983

Revised February 29, 1984

UNIVERSITY OF WISCONSIN-MADISON
COMPUTER SCIENCES DEPARTMENT

SUFFICIENCY OF EXACT PENALTY MINIMIZATION

O. L. Mangasarian

Technical Report #508

August 1983

ABSTRACT

By employing a recently obtained error bound for differentiable convex inequalities, it is shown that, under appropriate constraint qualifications, a minimum solution of an exact penalty function for a single value of the penalty parameter which exceeds a certain threshold, is also a solution of the convex program associated with the penalty function. No *a priori* assumption is made regarding the solvability of the convex program. If such a solvability assumption is made then we show that a threshold value of the penalty parameter can be used which is smaller than both the above-mentioned value and that of Zangwill. These various threshold values of the penalty parameter also apply to the well known big-M method of linear programming.

AMS (MOS) Subject Classifications: 90C30, 90C25

Key Words: Nonlinear programming, penalty functions, optimization, convex programming

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based on work sponsored by National Science Foundation Grant MCS-8200632.

SUFFICIENCY OF EXACT PENALTY MINIMIZATION

O. L. Mangasarian

1. Introduction

Consider the convex program

$$(1.1) \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \leqq 0$$

where $f: R^n \to R$, $g: R^n \to R^m$ are convex functions on the n-dimensional real Euclidean space R^n. It is well known [11,6] that if (1.1) has a solution $\bar{x}$ and if the constraints of (1.1) satisfy a constraint qualification then the exact penalty function

$$(1.2) \qquad P(x,\alpha):= f(x) + \alpha e g(x)_+ = f(x) + \alpha \sum_{i=1}^{m} \max\{0, g_i(x)\}$$

where e is a vector of ones in R^m, has a global minimum at $\bar{x}$ for each value of $\alpha \geqq \bar{\alpha}$ for some threshold value $\bar{\alpha}$. In [11, p. 356; 2, Theorem 40] it was shown that

$$(1.3) \qquad \bar{\alpha} = \bar{\alpha}_1 := \frac{f(x^1) - f(\bar{x}) + 1}{\min\limits_{1 \leqq i \leqq m} - g_i(x^1)}$$

where x^1 is any point satisfying the Slater constraint qualification

$$(1.4) \qquad g(x^1) < 0$$

In [6, Theorem 4.9] it was shown that

$$(1.5) \qquad \bar{\alpha} = \bar{\alpha}_2 := \|\bar{u}\|_\infty = \max_{1 \leqq i \leqq m} \bar{u}_i$$

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based on work supported by the National Science Foundation under Grant MCS-8200632.

where $\bar{u}$ is an optimal Lagrange multiplier for (1.1) provided that (1.4) holds. A minor modification of the proof of [6, Theorem 4.9] which invokes [10, Theorem 28.2] instead of [7, Theorem 5.4.8] extends (1.5) to the case where a relaxed Slater constraint qualification holds, that is

$$(1.6) \qquad g_{I_1}(x^2) < 0,\ g_{I_2}(x^2) \leqq 0 \quad \text{for some } x^2$$

where g_{I_1} is nonlinear and g_{I_2} is linear and $I_1 \cup I_2 = \{1,\ldots,m\}$. In contrast Zangwill's threshold (1.3) does not hold under the relaxed Slater constraint qualification (1.6) but must be replaced by a different value given by (2.2) below.

What is not well known and constitutes a principal concern of this work are converses to the results stated above. In [11, p. 356; 2, Theorem 40] Zangwill shows that if we assume *a priori* that the minimization problem (1.1) has a solution, the Slater constraint qualification (1.4) is satisfied and $\bar{x}$ minimizes the exact penalty function (1.2) for some $\alpha \geqq \bar{\alpha}_1$, then $\bar{x}$ solves the minimization problem (1.1). Note the *a priori* assumptions that (1.1) is solvable and that it satisfies the Slater constraint qualification. By contrast in [6, Theorem 4.1] without any *a priori* assumptions regarding the solvability of the minimization problem (1.1) or the satisfaction of a constraint qualification it was shown that if (1.1) is feasible, that is $g(x) \leqq 0$ for some x, and if $\bar{x}$ minimizes $P(x,\alpha)$ for <u>all</u> values of $\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha}$, then $\bar{x}$ solves the minimization problem (1.1). Note the distinction between these two sufficient conditions for $\bar{x}$ to solve the minimization problem (1.1). In Zangwill's result there are *a priori* assumptions that (1.1) is solvable and that its constraints satisfy the Slater constraint qualification, while the penalty function $P(x,\alpha)$ need be minimized for a single value of $\alpha \geqq \bar{\alpha}_1$. In [6, Theorem 4.1] no *a priori* assumption regarding the existence of a solution to (1.1) is made, however feasibility of (1.1) is assumed and $\bar{x}$ must be a solution to $\min_{x \in R^n} P(x,\alpha)$ for <u>all</u> $\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha}$, in order for $\bar{x}$ to be a solution to (1.1).

A primary purpose of this work is to combine the good features of these two results, namely the minimization of the penalty function for a <u>single</u> value of the penalty parameter and without an *a priori* assumption that the minimization problem has a solution. This is done in Theorem 3.1 where it is established that if for a single value of the penalty parameter $\alpha \geqq \bar{\alpha}_3$ for a well defined $\bar{\alpha}_3$, $\bar{x}$ minimizes the exact penalty function $P(x,\alpha)$ over R^n, then $\bar{x}$ is also a global solution of the minimization problem (1.1). Although no *a priori* assumption regarding the solvability of (1.1) is made in Theorem 3.1, both the relaxed Slater constraint qualification (1.6) and a mild asymptotic constraint qualification (3.2) are needed in order to invoke the recent [8, Theorem 2.1] absolute error bound for convex differentiable inequalities which plays a key role in the derivation of Theorem 3.1. Another result of this work is a two-way improvement of Zangwill's sufficiency result in Theorem 2.1, where the threshold value of $\bar{\alpha}$ is decreased from $\bar{\alpha}_1$ of (1.3) to $\bar{\alpha}_2$ of (1.5) and the Slater constraint qualification (1.4) is replaced by the relaxed constraint qualification (1.6). We also give in Corollary 2.3 a finite counterpart of the threshold value $\bar{\alpha}_1$ of (1.3) when the Slater constraint qualification (1.4) is replaced by the relaxed qualification (1.6) which renders $\bar{\alpha}_1$ infinite. Table 1 below gives a general outline of the relations between the various sufficiency results derived here and elsewhere for exact penalty functions and indicates the key assumptions needed for the different results to hold.

		A priori solvability of min. prob. (1.1):			
		Assumed	Not assumed		
Penalty function (1.2) minimized for:	All $\alpha \geq \bar{\alpha}$		Han-Mangasarian [6, Theorem 4.1]	Not assumed	Constraint qualification:
	A single $\alpha \geq \bar{\alpha}$	Zangwill [11, p. 356] Theorem 2.1	Theorem 3.1	Assumed	

Table 1: An outline of the key assumptions needed in the various sufficiency theorems establishing that each minimizer of an exact penalty function (1.2) solves the minimization problem (1.1).

In Section 3 of the paper we show that the big-M method of linear programming [1,9] is in fact equivalent to an exact penalty problem and hence the threshold values of the penalty parameter developed in this work apply to it as well as to a big-M formulation for convex programs. Such threshold values do not seem to have been given for the big-M method for linear programs.

We briefly describe now our notation. For a vector x in the n-dimensional real Euclidean space R^n, x_+ will denote the vector in R^n

with components $(x_+)_i = \max\{x_i, 0\}$, $i=1,\ldots,n$. For a vector norm $||x||$ on R^n, $||x||'$ will denote the dual norm on R^n, that is $||x||' = \max\limits_{||y||=1} xy$, where xy denotes the scalar product $\sum\limits_{i=1}^{n} x_i y_i$. The Cauchy-Schwarz inequality $|xy| \leq ||x|| \cdot ||y||'$ for x and y in R^n follows immediately from this definition of the dual norm. For $1 \leq p, q \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, the p-norm $||x||_p := (\sum\limits_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ and the q-norm are dual norms in R^n. For an $m \times n$ matrix A, A_i denotes the ith row, while $||A||_p$ denotes the matrix norm subordinate to the vector norm $||\cdot||_p$, that is $||A||_p = \max\limits_{||x||_p=1} ||Ax||_p$. The consistency condition $||Ax||_p \leq ||A||_p \; ||x||_p$ follows immediately from this definition of a matrix norm. We shall also use $||\cdot||$ to denote an arbitrary vector norm and its subordinate matrix norm. A vector of ones in any real Euclidean space will be denoted by e. For a differentiable function $g: R^n \to R^m$, $\nabla g(x)$ will denote the $m \times n$ Jacobian matrix evaluated at the point x in R^n. For a subset $I \subset \{1,\ldots,m\}$, $g_I(x)$ or $g_{i \in I}(x)$ will denote those components $g_i(x)$ such that $i \in I$. Similarly $\nabla g_I(x)$ will denote the rows $(\nabla g(x))_i$ of $\nabla g(x)$ such that $i \in I$. The set of vectors in R^n with nonnegative components will be denoted by R^n_+.

2. Exact Penalty Characterization Assuming Solvability of the Minimization Problem

In this section we completely characterize solutions of the minimization problem (1.1) in terms of minimizers of the exact penalty function (1.2) for a single value of the penalty parameter exceeding the threshold $\bar{\alpha}_2$. This is done under the assumptions that the minimization problem is solvable and that it satisfies the relaxed Slater constraint qualification (1.6). The necessity part of the following result Theorem 2.1 is an improvement over both [6, Theorem 4.9] and Zangwill's Theorem [11, p. 356] both of which require the Slater constraint qualification (1.4) instead of the relaxed qualification (1.6) needed here. This is a simple but important difference because it allows us to handle linearly constrained problems with no constraint qualification, and because Zangwill's threshold value $\bar{\alpha}_1$ becomes infinite under the relaxed constraint qualification (1.6). The new sufficiency part of Theorem 2.1 again improves over Zangwill's sufficiency result by using the relaxed Slater constraint qualification (1.6) instead of the Slater constraint qualification (1.4), and the smaller threshold value $\bar{\alpha}_2$ instead of $\bar{\alpha}_1$. It is interesting to note that the sufficiency part of Theorem 2.1 for the threshold value $\bar{\alpha}_2$ does not appear to have been given before even under the Slater constraint qualification. Now we state our result.

2.1 Theorem (Exact penalty characterization of solvable convex programs) Let $f:R^n \to R$ and $g:R^n \to R^m$ be convex functions on R^n. Let either $(\bar{x},\bar{u}) \in R^n \times R^m_+$ be a Karush-Kuhn-Tucker saddlepoint of the minimization problem (1.1), or let the relaxed Slater constraint qualification (1.6) hold and $\bar{x}$ be a solution of (1.1). A necessary (sufficient) condition for $\tilde{x} \in R^n$ to solve the minimization problem (1.1) is that $\tilde{x}$ minimizes $P(x,\alpha)$ over x in R^n for each (some) $\alpha \geqq \|\hat{u}\|_\infty$ $(\alpha > \|\hat{u}\|_\infty)$ where $\hat{u} \in R^m_+$ is any (some) dual optimal multiplier for (1.1).

Proof (Necessity) By assumption or by [10, Theorem 28.2] there exists a $\bar{u} \in R^m_+$ such that $(\bar{x},\bar{u})$ is a Karush-Kuhn-Tucker saddlepoint of (1.1). For any other dual optimal multiplier $\hat{u}$, $(\bar{x},\hat{u})$ is also a Karush-Kuhn-Tucker saddlepoint of (1.1) [4, p. 5]. Hence for $x \in R^n$ and $\alpha \geqq \|\hat{u}\|_\infty$

$$P(\bar{x},\alpha) = f(\bar{x}) = f(\bar{x}) + \hat{u}g(\bar{x}) \leqq f(x) + \hat{u}g(x)$$
$$\leqq f(x) + \hat{u}g(x)_+ \leqq f(x) + \|\hat{u}\|_\infty \|g(x)_+\|_1 \leqq P(x,\alpha)$$

(Sufficiency) Let $\hat{u} \in R^m_+$ be some dual optimal multiplier for (1.1). Since $(\bar{x},\bar{u})$ is a Karush-Kuhn-Tucker saddlepoint for (1.1) it follows by the necessity part of this theorem that for $\beta := \|\hat{u}\|_\infty$

$$P(\bar{x},\beta) = \min_{x \in R^n} P(x,\beta)$$

Let $\tilde{x}$ be a solution of $\min_{x \in R^n} P(x,\alpha)$ for some $\alpha > \|\hat{u}\|_\infty = \beta$.

Hence

$$f(\bar{x}) + \alpha e g(\bar{x})_+ \geqq f(\tilde{x}) + \alpha e g(\tilde{x})_+$$

and

$$f(\tilde{x}) + \beta e g(\tilde{x})_+ \geqq f(\bar{x}) + \beta e g(\bar{x})_+$$

Addition of the last two inequalities gives upon noting that $g(\bar{x})_+ = 0$

$$(\alpha - \beta) e g(\tilde{x})_+ \leqq 0$$

Since $\alpha > \beta$ this implies that $g(\tilde{x}) \leqq 0$ and hence $\tilde{x}$ is feasible for (1.1). For any other feasible point x

$$f(x) = P(x,\alpha) \geqq P(\tilde{x},\alpha) = f(\tilde{x}). \qquad \square$$

The following corollary shows that under the Slater constraint qualification the threshold value $\bar{\alpha}_2 := \|\hat{u}\|_\infty$ of Theorem 2.1 is smaller than that of Zangwill's $\bar{\alpha}_1$ as defined in (1.3).

2.2 Corollary Let $f:R^n \to R$ and $g:R^n \to R^m$ be convex functions on R^n, let x^1 be any point in R^n satisfying the Slater constraint qualification $g(x^1) < 0$, and let $\bar{x}$ be a solution of the minimization problem (1.1). Then for any dual optimal multiplier $\hat{u} \in R^m_+$ for (1.1)

$$\text{(2.1)} \qquad \|\hat{u}\|_\infty \leqq \|\hat{u}\|_1 \leqq \frac{f(x^1) - f(\bar{x})}{\min\limits_{1 \leqq i \leqq m} -g_i(x^1)} < \frac{f(x^1) - f(\bar{x}) + 1}{\min\limits_{1 \leqq i \leqq m} -g_i(x^1)} =: \bar{\alpha}_1$$

Proof Since $\bar{x}$ is a solution of (1.1) and the Slater constraint qualification is satisfied it follows that $\bar{x}$ and some $\bar{u} \in R^m_+$ constitute a Karush-Kuhn-Tucker saddlepoint for (1.1) and by [4, p. 5] so does $(\bar{x},\hat{u})$. Consequently

$$f(\bar{x}) = f(\bar{x}) + \hat{u}g(\bar{x}) \leqq f(x^1) + \hat{u}g(x^1) \leqq f(x^1) - \|\hat{u}\|_1 \min_{1 \leqq i \leqq m} -g_i(x^1)$$

from which (2.1) follows. $\square$

We establish now another upper bound for the threshold value $\bar{\alpha}_2 := \|\hat{u}\|_\infty$ of Theorem 2.1 under the relaxed Slater constraint qualification (1.6).

2.3 Corollary Let $f: R^n \to R$ and $g: R^n \to R^m$ be differentiable convex functions on R^n, let x^2 be any point in R^n satisfying the relaxed Slater constraint qualification (1.6), and let $\bar{x}$ be a solution of the minimization problem (1.1). Then there exists a dual optimal multiplier $\hat{u} \in R^m_+$ for (1.1) such that

$$(2.2) \qquad \|\hat{u}\|_\infty \leqq \|\hat{u}\|_1 \leqq \frac{f(x^2) - f(\bar{x})}{\min\limits_{i \in I_1} - g_i(x^2)}$$

$$+ \left(\|\nabla f(\bar{x})\|_1 + \frac{f(x^2) - f(\bar{x})}{\min\limits_{i \in I_1} - g_i(x^2)} \|\nabla g_{I_1}(\bar{x})\|_1 \right) \max_{\substack{g(x) \leqq 0 \\ I(x) \in J(x)}} \left\| A^T_{I(x)} \left(A_{I(x)} A^T_{I(x)} \right)^{-1} \right\|_1$$

where

$$(2.3) \qquad J(x) = \{ I \,|\, I \subset I_2,\ A_I x = b_I,\ A_{i \in I} \text{ lin. indep.} \}$$

and $g_{I_2}(x) = A_{I_2} x - b_{I_2}$.

Proof Since $\bar{x}$ is a solution of (1.1) and the relaxed Slater constraint qualification is satisfied it follows that $\bar{x}$ and some $\bar{u} \in R^m_+$ constitute a Karush-Kuhn-Tucker saddlepoint for (1.1). Since f and g are differentiable it follws that

$$(2.4) \qquad \nabla f(\bar{x}) + \bar{u}_{I_1} \nabla g_{I_1}(\bar{x}) + \bar{u}_{I_2} A_{I_2} = 0,\ \bar{u} g(\bar{x}) = 0,\ g(\bar{x}) \leqq 0,\ \bar{u} \geqq 0$$

By the fundamental theorem on the existence of basic feasible solutions [3, Theorem 2.11] it follows that there exists $\hat{u} \in R^m_+$ such that $(\bar{x}, \hat{u})$ is a Karush-Kuhn-Tucker saddlepoint of (1.1) and

(2.5) $$\nabla f(\bar{x}) + \hat{u}_{I_1}\nabla g_{I_1}(\bar{x}) + \hat{u}_{I(\bar{x})}A_{I(\bar{x})} = 0, \quad \hat{u}_{i\notin I_1 \cup I(\bar{x})} = 0$$

where $I(\bar{x})$ belongs to $J(\bar{x})$ as defined by (2.3). Hence

(2.6) $$\hat{u}_{I(\bar{x})} = -(\nabla f(\bar{x}) + \hat{u}_{I_1}\nabla g_{I_1}(\bar{x}))A^T_{I(\bar{x})}(A_{I(\bar{x})}A^T_{I(\bar{x})})^{-1}$$

Consequently

(2.7) $$\|\hat{u}_{I(\bar{x})}\|_1 \leqq (\|\nabla f(\bar{x})\|_1 + \|\hat{u}_{I_1}\|_1 \|\nabla g_{I_1}(\bar{x})\|_1)\|A^T_{I(\bar{x})}(A_{I(\bar{x})}A^T_{I(\bar{x})})^{-1}\|_1$$

From the saddlepoint property we have that

$$f(\bar{x}) \leqq f(x^2) + \hat{u}_{I_1}g_{I_1}(x^2) + \hat{u}_{I(\bar{x})}g_{I(\bar{x})}(x^2) \leqq f(x^2) - \|\hat{u}_{I_1}\|_1 \min_{i\in I_1} - g_i(x^2)$$

Hence

(2.8) $$\|\hat{u}_{I_1}\|_1 \leqq \frac{f(x^2) - f(\bar{x})}{\min\limits_{i\in I_1} - g_i(x^2)}$$

Combining (2.7) and (2.8) gives

$$\|\hat{u}\|_\infty \leqq \|\hat{u}\|_1 \leqq \frac{f(x^2) - f(\bar{x})}{\min\limits_{i\in I_1} - g_i(x^2)}$$

$$+ \left(\|\nabla f(\bar{x})\|_1 + \frac{f(x^2) - f(\bar{x})}{\min\limits_{i\in I_1} - g_i(x^2)} \|\nabla g_{I_1}(\bar{x})\|_1\right) \left\|A^T_{I(\bar{x})}\left(A_{I(\bar{x})}A^T_{I(\bar{x})}\right)^{-1}\right\|_1$$

Inequality (2.2) follows from the above upon replacing the last term by its maximum over all feasible x. □

It is evident that the last term in (2.2) may be difficult to compute because of its combinatorial aspect. However if there are only a few linear constraints, or if the point x^2 is interior to most of the linear constraints, in which case these constraints can be lumped with the nonlinear constraints, it may not be too difficult to compute the bound of (2.2). Obviously since $\bar{x}$ is unknown beforehand, $f(\bar{x})$ must be replaced by a lower bound (as must be done for Zangwill's bound $\bar{\alpha}_1$) and $\|\nabla f(\bar{x})\|_1$ and $\|\nabla g_{I_1}(\bar{x})\|_1$, by upper bounds in (2.2).

3. Exact Penalty Characterization Without Assuming Solvability of the Minimization Problem

In this section we characterize solutions of the minimization problem (1.1) in terms of minimizers of the exact penalty function (1.2) without any *a priori* assumption regarding the existence of solutions to (1.1) as was the case in the previous section. We do however need the relaxed Slater constraint qualification (1.6) and a mild asymptotic constraint qualification (3.2) below, which is automatically satisfied if all the constraints are linear. It is interesting to note that the threshold value $\bar{\alpha}_3$ of the penalty parameter in Theorem 3.1 below exceeds or equals the threshold value $\bar{\alpha}_2 := \|\bar{u}\|_\infty$ of Theorem 2.1.

3.1 Theorem (Exact penalty characterization of feasible convex programs) Let $f:R^n \to R$ and $g:R^n \to R^m$ be differentiable convex functions on R^n. Let the relaxed Slater constraint qualification (1.6) hold, let

$$(3.1) \qquad 0 \neq \beta := \sup_x \{\|\nabla f(x)\|_1 \mid g(x) \leqq 0\} < \infty$$

and let the following asymptotic constraint qualification [8] hold:

(3.2) For each nonempty $I \subset \{1,\ldots,m\}$ and each sequence of points $\{x^i\}$ such that: $g(x^i) \leqq 0$, $g_I(x^i) = 0$ and $\nabla g_{j\in I}(x^i)$ are linearly independent, each accumulation point $(\overline{\nabla g}_{I_0}, \overline{\nabla g}_{I_1}, \overline{\nabla g}_{I_2})$ of the sequence $\{\nabla g_{j\in I_0}(x^i)/\|\nabla g_{j\in I_0}(x^i)\|, \nabla g_{I_1}(x^i), \nabla g_{I_2}(x^i)\}$ satisfies

$$\overline{\nabla g}_{I_0} z > 0, \ \overline{\nabla g}_{I_1} z > 0, \ \overline{\nabla g}_{I_2} z \geqq 0 \ \text{for some} \ z \in R^n$$

where $I_0 \cup I_1 \cup I_2$ is a partition of I such that the sequence $\{\nabla g_j(x^i)\}$ is unbounded for $j \in I_0$ and bounded for $j \in I_1$, $g_{j\in I_0 \cup I_1}$ is nonlinear and $g_{j\in I_2}$ is linear.

A necessary (sufficient) condition for $\bar{x} \in R^n$ to solve the minimization problem (1.1) is that $\bar{x}$ minimizes $P(x,\alpha)$ over x in R^n for all $\alpha \geqq \bar{\alpha}_3$ (<u>some</u> $\alpha > \bar{\alpha}_3$) where

$$(3.3) \quad \bar{\alpha}_3 := \beta \sup_{w,p,I} \{\|w_I\|_\infty \mid g(p) \leqq 0,\ w_I > 0,\ g_I(p) = 0,\ \|w_I \nabla g_I(p)\|_1 = 1,\ \nabla g_{j\in I}(p) \text{ lin. ind., } I \subset \{1,\dots\dots,m\}\}$$

<u>Proof</u> We first note that the finiteness of $\bar{\alpha}_3$ is ensured by the asymptotic constraint qualification [8, Theorem 2.1].

(Necessity) Let $\bar{x}$ be a solution of (1.1) and let $\bar{u} \in R^m_+$ be an optimal dual multiplier for (1.1) chosen as indicated below. We will show that $\bar{\alpha}_3 \geqq \|\bar{u}\|_\infty$ and hence by the necessity part of Theorem 2.1, $\bar{x}$ minimizes $P(x,\alpha)$ for $\alpha \geqq \bar{\alpha}_3$. If $\nabla f(\bar{x}) = 0$ we take $\bar{u} = 0$ and evidently $\bar{\alpha}_3 \geqq \|\bar{u}\|_\infty = 0$. Suppose now $\nabla f(\bar{x}) \neq 0$. Take $\bar{u} = (\bar{u}_L, \bar{u}_K)$ where $\bar{u}_L > 0$ and corresponding to "basic" $g_{j\in L}(\bar{x}) = 0$ such that $\nabla g_{j\in L}(\bar{x})$ are linearly independent and $\bar{u}_K = 0$. Hence by the Karush-Kuhn-Tucker conditions [7]

$$\nabla f(\bar{x}) + \bar{u}_L \nabla g_L(\bar{x}) = 0$$

and consequently

$$\left\| \frac{\bar{u}_L}{\|\nabla f(\bar{x})\|_1} \nabla g_L(\bar{x}) \right\|_1 = 1$$

Hence by the definition (3.3) of $\bar{\alpha}_3$ and the definition (3.1) of β

$$\bar{\alpha}_3 \geqq \beta \left\| \frac{\bar{u}_L}{\|\nabla f(\bar{x})\|_1} \right\|_\infty = \beta \frac{\|\bar{u}\|_\infty}{\|\nabla f(\bar{x})\|_1} \geqq \|\bar{u}\|_\infty =: \bar{\alpha}_2.$$

(Sufficiency) Let $\bar{x}$ be a solution of $\min_{x\in R^n} P(x,\alpha)$ for some $\alpha > \bar{\alpha}_3$. We first show, by contradiction, that $g(\bar{x}) \leqq 0$. For if $\bar{x}$ is infeasible, then by [8, Theorem 2.1] there exist a feasible $p(\bar{x})$ such that

$$\|\bar{x} - p(\bar{x})\|_\infty \leqq \frac{\bar{\alpha}_3}{\beta}\|g(\bar{x})_+\|_1 = \frac{\bar{\alpha}_3}{\beta} eg(\bar{x})_+ \tag{3.4}$$

Then for $\alpha > \bar{\alpha}_3$

$$\begin{aligned} f(p(\bar{x})) &= P(p(\bar{x}),\alpha) \\ &\geqq P(\bar{x},\alpha) \quad \text{(Since } \bar{x} \text{ minimizes } P(x,\alpha) \text{ over } x \in R^n) \\ &= f(\bar{x}) + \alpha eg(\bar{x})_+ \\ &> f(\bar{x}) + \beta\|\bar{x} - p(\bar{x})\|_\infty \quad \text{(By (3.4), } \alpha > \bar{\alpha}_3 \text{ and } g(\bar{x})_+ \neq 0) \\ &\geqq f(\bar{x}) + \|\nabla f(p(\bar{x}))\|_1 \|\bar{x} - p(\bar{x})\|_\infty \quad \text{(By (3.1))} \\ &\geqq f(\bar{x}) - \nabla f(p(\bar{x}))(\bar{x} - p(\bar{x})) \quad \text{(By the Cauchy-Schwarz inequality)} \\ &\geqq f(p(\bar{x})) \quad \text{(By the convexity of } f) \end{aligned}$$

which is a contradiction. Hence $g(\bar{x}) \leqq 0$ and $\bar{x}$ is feasible. For any other feasible x and $\alpha > \bar{\alpha}_3$

$$f(\bar{x}) = P(\bar{x},\alpha) \leqq P(x,\alpha) = f(x)$$

and hence $\bar{x}$ solves (1.1). □

Obviously the threshold value $\bar{\alpha}_3$ given by (3.2) is difficult to compute in general. However besides providing an existence result for the minimization problem (1.1), it is useful to know that such a threshold value exists and to know how it depends on the problem parameters, especially when one is engaged in an unconstrained exact penalty function minimization either on R^n, as a substitute for the original constrained optimization problem, or on R^1 as part of an iterative method [5]. In both cases an α such that $\alpha > \bar{\alpha}_3$ would be a useful upper bound to the penalty parameters employed. This would avoid the use of arbitrarily large penalty parameters that may lead to numerical difficulties.

4. An Application: The Big-M Method for Convex Programs

In linear programming, a well known method [9, 1] for solving a linear program without an explicit phase I procedure is to add nonnegative artificial variables to the constraints and then add a penalty to the objective function involving the artificial variables. If the penalty parameter is "sufficiently large", then the artificial variables will be driven to zero and an optimal solution will be obtained, if one exists. In this section we will make the "sufficiently large" concept precise by using the results of the two previous sections and extend the idea of the big-M method to convex programs. We first state a simple lemma whose elementary proof we omit.

4.1 Lemma Let $f:R^n \to R$, $g:R^n \to R^m$ and let $\alpha > 0$. Then the problems

$$(4.1) \qquad \min_{x \in R^n} \; f(x) + \alpha e g(x)_+ =: \min_{x \in R^n} P(x,\alpha)$$

$$(4.2) \qquad \min_{(x,z) \in R^{n+m}} \; f(x) + \alpha e z \quad \text{s.t.} \quad g(x) \leqq z, \; z \geqq 0$$

are equivalent in the following sense: For each solution $\bar{x}$ of (4.1), $(\bar{x},\ \bar{z} := g(\bar{x})_+)$ solves (4.2), and for each solution $(\tilde{x},\tilde{z})$ of (4.2), $\tilde{x}$ solves (4.1).

The formulation of (4.2) is the big-M formulation and is used in linear programming because it is easy to obtain a feasible point for it by taking any x in R^n and $z := g(x)_+$. Formulation (4.2) can be used also for the very same reason in convex programming. Theorem 2.1 tells us that if we know *a priori* that problem (1.1) has a solution, f and g are convex and the relaxed Slater constraint qualification (1.6) is satisfied then the penalty parameter α of the big-M formulation (4.2) must satisfy $\alpha > \bar{\alpha}_2 := \|\bar{u}\|_\infty$ where $\bar{u}$ is any optimal dual multiplier to (1.1). Note that if g is linear, then the relaxed Slater constraint qualification (1.6) is satisfied by any feasible point x. If we have no *a priori* knowledge that (1.1) is solvable, but that it is merely feasible, that f, g are differentiable and convex, and that (3.1) and the constraint qualifications (1.6) and (3.2) are satisfied then the penalty parameter α of the big-M method (4.2) must satisfy $\alpha > \bar{\alpha}_3$ where $\bar{\alpha}_3$ is defined by (3.3). Note that if g is linear then (1.6) and (3.2) are automatically satisfied, and if in addition f is nonconstant and linear, then (3.1) is also automatically satisfied.

References

1. M. S. Bazaraa and J. J. Jarvis: "Linear programming and network flows", John Wiley, New York, 1977.

2. A. V. Fiacco and G. P. McCormick: "Nonlinear programming: Sequential unconstrained minimization techniques", John Wiley, New York, 1968.

3. D. Gale: "The theory of linear economic models", McGraw-Hill, New York, 1960.

4. A. M. Geoffrion: "Duality in nonlinear programming: A simplified applications oriented development", SIAM Review 13, 1971, 1-37.

5. S.-P. Han: "A globally convergent method for nonlinear programming", Journal of Optimization Theory and Applications 22, 1977, 297-309.

6. S.-P. Han and O. L. Mangasarian: "Exact penalty functions in nonlinear programming", Mathematical Programming 17, 1979, 251-269.

7. O. L. Mangasarian: "Nonlinear programming", McGraw-Hill, New York, 1969.

8. O. L. Mangasarian: "A condition number for differentiable convex inequalities", University of Wisconsin, Madison, Computer Sciences Technical Report #491, Revised July 1983.

9. K. Murty: "Linear and combinatorial programming", John Wiley, New York, 1976.

10. R. T. Rockafellar: "Convex analysis", Princeton University Press, Princeton, New Jersey, 1970.

11. W. Zangwill: "Non-linear programming via penalty functions", Management Science 13, 1967, 344-358.