SOME APPLICATIONS OF PENALTY FUNCTIONS
IN MATHEMATICAL PROGRAMMING

by

O. L. Mangasarian

Computer Sciences Technical Report #544

June 1984

UNIVERSITY OF WISCONSIN-MADISON
COMPUTER SCIENCES DEPARTMENT

SOME APPLICATIONS OF PENALTY FUNCTIONS IN MATHEMATICAL PROGRAMMING

O. L. Mangasarian

Technical Report #544

June 1984

ABSTRACT

By using an exterior penalty function and recent boundedness and existence results for monotone complementarity problems, we give existence and boundedness results, for a pair of dual convex programs, of the following nature. If there exists a point which is feasible for the primal problem and which is interior to the constraints of the Wolfe dual, then the primal problem has a solution which is easily bounded in terms of the feasible point. Furthermore there exists no duality gap. We also show that by solving an exterior penalty problem for only two values of the penalty parameter we obtain an optimal point which is approximately feasible to any desired preassigned tolerance. This result is then employed to obtain an estimate of the perturbation parameter for a linear program which allows us to solve the linear program to any preassigned accuracy by an iterative scheme such as a successive overrelaxation (SOR) method.

AMS (MOS) Classification: 90C30, 90C25, 90C05

Key Words: Penalty functions, mathematical programming, duality, linear programs

---

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.

SOME APPLICATIONS OF PENALTY FUNCTIONS IN MATHEMATICAL PROGRAMMING

O. L. Mangasarian

1. Introduction

We consider in this work the constrained minimization problem

$$\text{(1.1)} \qquad \min_{x \in X} f(x), \; X := X_0 \cap X_1$$

where X_0 and X_1 are subsets of the n-dimensional real space R^n which have a nonempty intersection X, and $f\colon X_0 \to R$. Associated with the above problem is the classical exterior penalty problem [3,2,1]

$$\text{(1.2)} \qquad \min_{x \in X_0} P(x,\alpha) := f(x) + \alpha Q(x)$$

where α is in R_+, the nonnegative real line, and $Q(x)\colon X_0 \to R_+$ such that $Q(x) = 0$ for $x \in X$, else $Q(x) > 0$. We have two principal applications in mind regarding the penalty problem (1.2). The first application, which employs in addition to (1.2) the recent boundedness and existence results for monotone complementarity problems [10] and which is described in Section 3 of the paper, gives existence and boundedness results for a convex program obtained from (1.1) and the associated dual problem. In particular we show in Theorem 3.1 that if there exists a point which is feasible for a primal convex program and is interior to the constraints of its Wolfe dual [12,5], then the primal problem has a solution which is easily bounded in terms of the feasible point, and that there is no duality gap between the primal problem and its Wolfe dual. Theorem 3.2 shows that

---

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.

if there is a point which is interior to the constraints of a primal convex program which is also feasible for the associated Wolfe dual, then the Lagrangian dual [4,1] of the convex program has a nonempty solution set which is easily bounded in terms of the feasible point, and in addition there is no duality gap between the primal problem and its Lagrangian dual. In Section 4 our main concern is the recasting by means of an exterior penalty function of the standard linear programming problem as a quadratic minimization problem on the nonnegative orthant in the spirit of previous work [6,7,8]. The principal new result here is to show how to obtain a precise value of the penalty parameter which allows us to satisfy the Karush-Kuhn-Tucker optimality conditions [5] for the linear program to any preassigned degree of precision. Theorem 4.1 shows that this can be done by minimizing a convex quadratic function on the nonnegative orthant for only two values of the penalty parameter. Iterative methods developed in [6,7,8] can solve by this approach very large sparse linear programs which cannot be solved by a standard linear programming simplex package [8].

Because of the key role played by exterior penalty functions in this work, we give in Section 2 some fundamental results regarding these functions in a form convenient for deriving our other results. Although some of these penalty results are known under more restrictive conditions [3,2], some are new. For example, Theorem 2.3 shows that by solving only two exterior penalty function minimization problems, we can obtain an optimal point which is feasible to any preassigned feasibility tolerance. Theorem 2.8 shows that under rather mild assumptions each accumulation point of a sequence of solutions of penalty functions, corresponding to an increasing unbounded sequence of positive numbers, solves the associated constrained optimization problem. Furthermore the corresponding sequence of products of the penalty parameter and the penalty term tends to zero.

We briefly describe our notation now. Vectors will be column or row vectors depending on the context. For a vector x in the n-dimensional real space R^n, $\|x\|$ will denote an arbitrary norm, while $\|x\|_p$ will denote the p-norm $\|x\|_p := (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ for $1 \leqq p < \infty$ and $\|x\|_\infty := \max_{1 \leqq i \leqq n} |x_i|$, where x_i is the i-th component of x; x_+ will denote the vector in R^n with components $(x_+)_i = \max\{x_i, 0\}$, $i=1,\ldots,n$. A vector of ones in any real space will be denoted by e. For a differentiable function $L: R^n \times R^m \to R$, $\nabla_x L(x,u)$ will denote the n-dimensional gradient vector $\frac{\partial L}{\partial x_i}(x,u)$, $i=1,\ldots,n$, while for $f: R^n \to R$, $\nabla f(x)$ will denote the n-dimensional gradient vector. The set of vectors in R^n with nonnegative components will be denoted by R^n_+.

2. Some Fundamental Properties of Exterior Penalty Functions

We collect in this section some fundamental properties of exterior penalty functions in a form convenient for our applications and under more general assumptions than usually given [3,1]. We begin with some elementary but important monotonicity properties for solutions of penalty problems.

2.1 Proposition Let $x_i \in X_0$ be a solution of $\min_{x\in X_0} P(x,\alpha_i)$ for $i=1,2$ with $\alpha_2 > \alpha_1 \geqq 0$. Then

$$(2.1) \qquad Q(x_2) \leqq Q(x_1),\ f(x_1) \leqq f(x_2),\ P(x_1,\alpha_1) \leqq P(x_2,\alpha_2)$$

Proof Addition of $P(x_2,\alpha_2) \leqq P(x_1,\alpha_2)$ and $P(x_1,\alpha_1) \leqq P(x_2,\alpha_1)$, gives, together with $\alpha_2 > \alpha_1$, the inequality $Q(x_2) \leqq Q(x_1)$, which in turn together with $P(x_1,\alpha_1) \leqq P(x_2,\alpha_1)$, and $\alpha_1 \geqq 0$, gives $f(x_1) \leqq f(x_2)$. We also have that

$$P(x_1,\alpha_1) \leqq P(x_2,\alpha_1) \leqq P(x_2,\alpha_2) \qquad \square$$

2.2 Proposition Let $\inf_{x\in X} f(x) > -\infty$, let $\alpha > 0$ and let $x(\alpha) \in X_0$ be such that $P(x(\alpha),\alpha) = \min_{x\in X_0} P(x,\alpha)$. Then

$$(2.2) \qquad f(x(\alpha)) \leqq \inf_{x\in X} f(x)$$

If $x(\alpha) \in X$ then

$$(2.3) \qquad f(x(\alpha)) = \min_{x\in X} f(x)$$

Proof For any $\varepsilon > 0$ pick $x(\varepsilon) \in X$ such that

$$f(x(\varepsilon)) < \inf_{x\in X} f(x) + \varepsilon$$

Then

$$\varepsilon + \inf_{x\in X} f(x) > f(x(\varepsilon)) = P(x(\varepsilon),\alpha) \geqq P(x(\alpha),\alpha) \geqq f(x(\alpha))$$

Since $x(\alpha)$ does not depend on ε, (2.2) follows by letting ε approach zero. If $x(\alpha)$ is also in X, then (2.3) is obviously a consequence of (2.2). □

The following simple theorem shows how, for any desired feasibility tolerance $\delta > 0$, solving the penalty problem (1.2) for only two values of the penalty parameter α will yield a point $x_2 \in X_0$ such that $Q(x_2) \leqq \delta$ and $f(x_2) \leqq \inf_{x\in X} f(x)$. Hence if δ chosen sufficiently small, x_2 is an approximately feasible optimal solution for the minimization problem (1.1).

2.3 Theorem Let $\delta > 0$, $\alpha_1 > 0$, let $\inf_{x\in X} f(x) > -\infty$, let $\hat{x} \in X$ and let $P(x_1,\alpha_1) = \min_{x\in X_0} P(x,\alpha_1)$. If $f(\hat{x}) \leqq f(x_1)$ then $\hat{x}$ solves $\min_{x\in X} f(x)$, else for

(2.4) $$\alpha_2 > \alpha_1 \quad \text{and} \quad \alpha_2 \geqq \frac{f(\hat{x}) - f(x_1)}{\delta}$$

it follows that

(2.5) $$x_2 \in X_0,\ Q(x_2) \leqq \delta,\ f(x_2) \leqq \inf_{x\in X} f(x)$$

where

$$P(x_2,\alpha_2) = \min_{x\in X_0} P(x,\alpha_2),\ x_2 \in X_0$$

Proof First note that if $f(\hat{x}) \leqq f(x_1)$ then by (2.2) $\hat{x}$ solves $\min_{x\in X} f(x)$. Suppose now that $f(\hat{x}) > f(x_1)$ and (2.4) holds. Then

(2.6) $$f(x_2) + \alpha_2 Q(x_2) \leqq f(\hat{x}) + \alpha_2 Q(\hat{x}) = f(\hat{x})$$

Hence by (2.4), (2.1) and (2.6) respectively it follows that

$$\delta \geqq \frac{f(\hat{x}) - f(x_1)}{\alpha_2} \geqq \frac{f(\hat{x}) - f(x_2)}{\alpha_2} \geqq Q(x_2)$$

which establishes the first inequality of (2.5). The second inequality of (2.5) follows from (2.2). □

2.4 Remark Theorem 2.3 can be applied to obtain an approximate solution of (1.1) in the sense of (2.5) as follows:

(a) Choose $\delta > 0$, $\alpha_1 > 0$, $\hat{x} \in X$.

(b) Compute $x_1 \in X_0$ such that: $P(x_1, \alpha_1) = \min_{x \in X_0} P(x, \alpha_1)$. If $f(\hat{x}) \leqq f(x_1)$, stop, $\hat{x}$ solves (1.1).

(c) Choose α_2 such that $\alpha_2 > \alpha_1$ and $\alpha_2 \geqq \frac{f(\hat{x}) - f(x_1)}{\delta}$

(d) Compute $x_2 \in X_0$ such that: $P(x_2, \alpha_2) = \min_{x \in X_0} P(x, \alpha_2)$.

If α_2 of step (c) is too large, an $\bar{\alpha}_1$ such that $\alpha_1 < \bar{\alpha}_1 < \alpha_2$ can be chosen to replace α_1 and steps (a)-(b)-(c) are repeated. Also $\hat{x}$ may be replaced when possible by some $\tilde{x} \in [\hat{x}, x_1] \cap X$ such that $f(\tilde{x}) < f(\hat{x})$.

The next result shows that for a sequence of solutions $\{x_i\}$ of the penalty problem (1.2) for an increasing unbounded sequence of penalty parameters $\{\alpha_i\}$, the sequence of penalties $\{Q(x_i)\}$ converges to 0 and the sequence $\{f(x_i)\}$ converges to a lower bound for $\inf_{x \in X} f(x)$, provided the latter is finite. We do not require that the sequence $\{x_i\}$ have an accumulation point here.

2.5 Theorem Let $\inf_{x \in X} f(x) > -\infty$, let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers, let $\{x_i\}$ be a corresponding sequence of points in X_0 not in X such that $P(x_i, \alpha_i) = \min_{x \in X_0} P(x, \alpha_i)$.

Then

(2.7) $$\lim_{i\to\infty} Q(x_i) = 0 \quad \text{and} \quad \lim_{i\to\infty} f(x_i) \leqq \inf_{x\in X} f(x).$$

Proof By (2.1), the sequence $\{Q(x_i)\}$ is nonincreasing and bounded below by 0 and hence converges to $\bar{Q} \geqq 0$ and $Q(x_i) \geqq \bar{Q}$, $i=1,2,\ldots$. If $\bar{Q} > 0$ we get from (2.5) by picking i sufficiently large such that $\alpha_i \geqq 2(f(\hat{x}) - f(x_i))/\bar{Q}$ where $\hat{x} \in X$, that $\bar{Q} \leqq Q(x_i) \leqq \bar{Q}/2$ which is a contradiction. Hence $\bar{Q} = 0$ and $\lim_{i\to\infty} Q(x_i) = 0$. Now again by (2.1), the sequence $\{f(x_i)\}$ is nondecreasing, and by (2.2) it is bounded above by $\inf_{x\in X} f(x)$. Hence $\{f(x_i)\}$ converges to $\bar{f}$ and

$$f(x_i) \leqq \bar{f} \leqq \inf_{x\in X} f(x) \qquad \square$$

To make the inequality in (2.7) an equality we need additional assumptions such as those given in the following corollary.

2.6 Corollary If in addition to the assumptions of Theorem 2.5, f is Lipschitz continuous on X_0, that is for some $K > 0$

(2.8) $$|f(y) - f(x)| \leqq K\|y - x\|_2 \quad \text{for all} \quad x, y \in X_0$$

and there exists a constant $\mu > 0$ such that for each $x \in X_0$ there exists an $\hat{x}(x) \in X$ such that

(2.9) $$\|x - \hat{x}(x)\|_2 \leqq \mu Q(x)$$

then

(2.10) $$\lim_{i\to\infty} f(x_i) = \inf_{x\in X} f(x)$$

Proof For each x_i there exists an $\hat{x}_i \in X$ such that

$$\|x_i - \hat{x}_i\|_2 \leq \mu Q(x_i)$$

Hence by (2.8) and (2.9)

$$0 \leq |f(x_i) - f(\hat{x}_i)| \leq K\|x_i - \hat{x}_i\|_2 \leq K\mu Q(x_i) \tag{2.11}$$

Since by (2.7) $\lim_{i\to\infty} Q(x_i) = 0$, it follows from (2.11) that

$$\lim_{i\to\infty} f(\hat{x}_i) = \lim_{i\to\infty} f(x_i) \tag{2.12}$$

From (2.11) and $\hat{x}_i \in X$ we get the inequalities

$$f(x_i) + K\mu Q(x_i) \geq f(\hat{x}_i) \geq \inf_{x\in X} f(x)$$

Taking the limit as $i \to \infty$ and using (2.7) gives

$$\inf_{x\in X} f(x) \geq \lim_{i\to\infty} f(x_i) \geq \inf_{x\in X} f(x)$$

Hence $\lim_{i\to\infty} f(x_i) = \inf_{x\in X} f(x)$. □

2.7 Remark Condition (2.9) is satisfied if the feasible region X is convex and satisfies an appropriate constraint qualification [9, Theorem 2.1]. In particular (2.9) holds in the special case when $X_0 = R^n$ and X_1 is defined by linear inequalities [9, Remark 2.2].

We observe that in both Theorem 2.5 and Corollary 2.6 the sequence $\{x_i\}$ need not have an accumulation point. A stronger result is obtained if $\{x_i\}$ has an accumulation point.

2.8 Theorem Let $\inf_{x\in X} > -\infty$, and let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers. Let $\{x_i\}$ be a corresponding sequence of points in X_0 not in X such that $P(x_i,\alpha_i) = \min_{x\in X_0} P(x,\alpha_i)$ with an accumulation point

$\bar{x} \in X_0$. If f and Q are lower semicontinuous at $\bar{x}$, then $Q(\bar{x}) = 0$ and $\bar{x}$ solves $\min_{x \in X} f(x)$. Furthermore

$$\lim_{j\to\infty} \alpha_{i_j} Q(x_{i_j}) = 0 \quad \text{for} \quad x_{i_j} \to \bar{x} \in X_0. \tag{2.13}$$

<u>Proof</u> Let $x_{i_j} \to \bar{x} \in X_0$. From (2.7) and the lsc of Q we have

$$0 = \lim_{j\to\infty} Q(x_{i_j}) \geqq Q(\bar{x}) \geqq 0$$

Hence $Q(\bar{x}) = 0$ and $\bar{x} \in X$. From (2.7) and the lsc of f we have

$$f(\bar{x}) \leqq \lim_{j\to\infty} f(x_{i_j}) \leqq \inf_{x \in X} f(x)$$

Since $\bar{x} \in X$, it follows that $\bar{x}$ solves $\min_{x \in X} f(x)$. To establish (2.13) note that

$$0 \geqq P(x_{i_j}, \alpha_{i_j}) - P(\bar{x}, \alpha_{i_j}) = f(x_{i_j}) - f(\bar{x}) + \alpha_{i_j} Q(x_{i_j})$$

Hence

$$f(\bar{x}) - f(x_{i_j}) \geqq \alpha_{i_j} Q(x_{i_j}) \geqq 0$$

By letting $j \to \infty$ and recalling that f is lsc at $\bar{x}$ it follows that $\lim_{j\to\infty} \alpha_{i_j} Q(x_{i_j}) = 0$. □

3. Bounds and Existence for Dual Convex Programs

We consider in this section the convex primal program

$$(3.1) \qquad \min_{x \in X} f(x), \; X = \{x \mid x \in R^n_+, \; g(x) \leqq 0\}$$

where $f: R^n \to R$, $g: R^n \to R^m$ are differentiable and convex on R^n. The Wolfe dual [12,5] associated with this problem is

$$(3.2) \qquad \max_{(x,u,v) \in Y} L(x,u) - vx, \; Y = \{(x,u,v) \mid x \in R^n, \; u \in R^m_+, \; v \in R^n_+, \; \nabla_x L(x,u) - v = 0\}$$

and the Lagrangian dual [4,1] is

$$(3.3) \qquad \max_{(u,v) \geqq 0} \; \inf_{x \in R^n} L(x,u) - vx$$

where $L(x,u) := f(x) + ug(x)$ is the usual Lagrangian. Note that (3.2) is equivalent to

$$(3.2') \qquad \max_{(x,u) \in Z} L(x,u) - x\nabla_x L(x,u), \; Z = \{(x,u) \mid x \in R^n, \; u \in R^m_+, \; \nabla_x L(x,u) \geqq 0\}$$

Note that (3.1) can be identified with problem (1.1) by setting $X_0 = R^n_+$ and $X_1 = \{x \mid g(x) \leqq 0\}$.

Our primary objective here is to give simple conditions for the separate existence of a solution to each of primal and Lagrangian dual problems and to bound their solutions. Loosely speaking we shall establish existence of a solution and a bound for the primal (Lagrangian dual) problem under a primal and Wolfe-dual feasibility assumption together with a Wolfe-dual (primal) constraint interiority assumption. Our principal tools will be the recent boundedness and existence results for monotone complementarity problems and convex programs of [10] and the penalty function results outlined in the previous section. We begin with an existence and boundedness result for the primal problem (3.1).

3.1 <u>Theorem</u> (Primal feasibility & Wolfe dual interior-feasibility $\Longrightarrow$ Primal solution existence-boundedness & zero duality gap with Wolfe dual) Let f and g be differentiable and convex on R^n and let $(\hat{x},\hat{u})$ satisfy

$$\hat{x}\in X,\ (\hat{x},\hat{u})\in Z,\ \nabla_x L(\hat{x},\hat{u}) > 0$$

Then there exists a primal optimal solution $\bar{x}$ to (3.1) which is bounded by

$$\text{(3.4)} \qquad \|\bar{x}\|_1 \leqq \frac{-\hat{u}g(\hat{x}) + \hat{x}\nabla_x L(\hat{x},\hat{u})}{\min_i\ (\nabla_x L(\hat{x},\hat{u}))_i}$$

In addition there exists no duality gap between the primal problem (3.1) and the Wolfe dual (3.2), that is:

$$\text{(3.5)} \qquad \min_{x\in X} f(x) = f(\bar{x}) = \sup_{(x,u,v)\in Y} L(x,u) - vx$$

<u>Proof</u> Consider the penalty function problem associated with (3.1)

$$\text{(3.6)} \qquad \min_{x\geqq 0} f(x) + \alpha e g(x)_+$$

or equivalently

$$\text{(3.6')} \qquad \min_{(x,z)\geqq 0} f(x) + \alpha e z \quad \text{s.t. } g(x) - z \leqq 0$$

The Wolfe dual associated with (3.6') is

$$(3.7)\qquad \max_{(x,z,u,v,w)} \; L(x,u) + z\,(\alpha e - u - w) - vx$$

$$\text{s.t. } \nabla_x L(x,u) - v = 0,\ \alpha e - u - w = 0,\ u, v, w \geqq 0$$

which is equivalent to

$$(3.7')\qquad \max_{(x,u)} L(x,u) - x\nabla_x L(x,u) \quad \text{s.t. } \nabla_x L(x,u) \geqq 0,\ \alpha e \geqq u \geqq 0$$

Note that the only difference between (3.7') and (3.2') is the constraint $\alpha e \geqq u$. Now, for any $\varepsilon > 0$, the point $(\hat{x},\ \hat{z} := e\varepsilon,\ \hat{u})$ satisfies a "Slater" constraint qualification for the dual problems (3.6')-(3.7') for $\alpha > \|\hat{u}\|_\infty$. Hence these probems have equal extrema and a solution $(x(\alpha),\ z(\alpha),\ u(\alpha))$ such that $x(\alpha)$ is bounded by [10, Theorem 2.3]

$$(3.8)\qquad \|x(\alpha)\|_1 \leqq \frac{\hat{u}(-g(\hat{x}) + \varepsilon e) + \hat{x}\nabla_x L(\hat{x},\hat{u}) + \varepsilon e(\alpha e - \hat{u})}{\min_i\ (\nabla_x L(\hat{x},\hat{u}))_i}$$

Since the left side of (3.8) does not depend on ε, we can let $\varepsilon \to 0$ in (3.8) and we have

$$(3.9)\qquad \|x(\alpha)\|_1 \leqq \frac{-\hat{u}g(\hat{x}) + \hat{x}\nabla_x L(\hat{x},\hat{u})}{\min_i\ (\nabla_x L(\hat{x},\hat{u}))_i}$$

Note now that by the weak duality theorem [5] applied to (3.1) and (3.3) we have

$$\inf_{x\in X} f(x) \geqq L(\hat{x},\hat{u}) - \hat{x}\nabla_x L(\hat{x},\hat{u}) > -\infty$$

Hence for an unbounded increasing sequence of positive numbers $\{\alpha_i\}$ exceeding $\|\hat{u}\|_\infty$, it follows [10, Theorem 2.3] that there exists a sequence of points $\{x(\alpha_i), u(\alpha_i)\}$ with $x(\alpha_i)$ bounded as in (3.9), such that each $x(\alpha_i)$ solves the penalty function problem (3.6) with $\alpha = \alpha_i$ and $(x(\alpha_i), u(\alpha_i))$ solves its dual (3.7'). Since $\{x(\alpha_i)\}$ is bounded it has an accumulation point $\bar{x}$ which is bounded by (3.9). Since $ez(\alpha_i) = e(g(x(\alpha_i)))_+$ is the penalty term for (3.6'), it follows by Theorem (2.8) that $e\bar{z} = eg(\bar{x})_+ = 0$, that $\bar{x}$ solves $\min_{x \in X} f(x)$ and that

$$\text{(3.10)} \qquad \lim_{j\to\infty} \alpha_{i_j} ez(\alpha_{i_j}) = \lim_{j\to\infty} \alpha_{i_j} e(g(x(\alpha_{i_j})))_+ = 0 \quad \text{for} \quad x(\alpha_{i_j}) \to \bar{x}$$

Now we establish the zero duality gap. Let $\{\varepsilon_i\}$ be any decreasing sequence of positive numbers converging to 0 and let $\{\alpha_i\}$ be an unbounded increasing sequence of positive numbers chosen as follows:

$$\infty > \sup_{(x,u)\in Z} L(x,u) - x\nabla_x L(x,u) - \varepsilon_i$$

(By weak duality theorem)

$$< L(x(\varepsilon_i), u(\varepsilon_i)) - x(\varepsilon_i)\nabla_x L(x(\varepsilon_i), u(\varepsilon_i))$$

(For some $(x(\varepsilon_i), u(\varepsilon_i)) \in Z$, by definition of sup)

$$\leqq L(x(\alpha_i), u(\alpha_i)) - x(\alpha_i)\nabla_x L(x(\alpha_i), u(\alpha_i))$$

(For α_i sufficiently large s.t. $\alpha_i \geqq \|u(\varepsilon_i)\|_\infty$, because $(x(\alpha_i), u(\alpha_i))$ solves $\max L(x,u) - x\nabla_x L(x,u)$ s.t. $\nabla_x L(x,u) \geqq 0,\ \alpha_i e \geqq u \geqq 0$)

$$= f(x(\alpha_i)) + \alpha_i ez(\alpha_i)$$

(By equality of primal-dual optimal objective functions of problems (3.6') and (3.7') with $\alpha = \alpha_i$)

$$= \sup_{(x,u)} \{L(x,u) - x\nabla_x L(x,u) \,|\, \nabla_x L(x,u) \geqq 0,\ \alpha_i e \geqq u \geqq 0\}$$

$$\leqq \sup_{(x,u)\in Z} L(x,u) - x\nabla_x L(x,u)$$

Since by (3.10), $\lim_{j\to\infty} \alpha_{i_j} ez(\alpha_{i_j}) = 0$ for $x(\alpha_{i_j}) \to \bar{x}$, it follows that

$$\sup_{(x,u)\in Z} L(x,u) - x\nabla_x L(x,u) = f(\bar{x}) = \min_{x\in X} f(x) \qquad \square$$

We establish now an existence and boundedness result for the Lagrangian dual problem (3.3).

3.2 <u>Theorem</u> (Wolfe-dual feasibility & primal interior-feasibility $\Longrightarrow$ Lagrangian dual solution existence-boundedness & zero duality gap with primal) Let f and g be differentiable and convex on R^n and let $(\tilde{x},\tilde{u})$ satisfy:

(3.11) $$\tilde{x} \in X,\ (\tilde{x},\tilde{u}) \in Z,\ \tilde{x} > 0,\ g(\tilde{x}) < 0$$

There exists a dual optimal solution $(\bar{u},\bar{v})$ to the Lagrangian dual (3.3) which is bounded by

(3.12) $$\|\bar{u},\bar{v}\|_1 \leqq \frac{-\tilde{u}g(\tilde{x}) + \tilde{x}\nabla_x L(\tilde{x},\tilde{u})}{\min_{i,j} \{-g_i(\tilde{x}),\ \tilde{x}_j\}}$$

In addition there is no duality gap between the primal problem (3.1) and the Lagrangian dual (3.3), that is:

(3.13) $$\inf_{x\in X} f(x) = \max_{(u,v)\geqq 0} \inf_{x\in R^n} L(x,u) - xv$$

<u>Proof</u> For $\beta > 0$ consider the bounded version of (3.1)

(3.14) $$\min f(x) \quad \text{s.t. } g(x) \leqq 0,\ \beta e \geqq x \geqq 0$$

and its Wolfe dual

(3.15)
$$\max_{(x,u,v,w)} L(x,u) - vx + w(x - \beta e)$$
$$\text{s.t. } \nabla_x L(x,u) - v + w = 0,\ u, v, w \geqq 0$$

or equivalently

(3.15')
$$\max_{(x,u,w)} L(x,u) - x\nabla_x L(x,u) - \beta ew$$
$$\text{s.t. } \nabla_x L(x,u) + w \geqq 0,\ u, w \geqq 0$$

which again is equivalent to

(3.15")
$$\max_{\substack{(x,u) \\ u \geqq 0}} L(x,u) - x\nabla_x L(x,u) - \beta e(-\nabla_x L(x,u))_+$$

which is nothing other than an exterior penalty function formulation for the Wolfe dual (3.2') with penalty parameter β. Thus the bound β on the ∞-norm of the primal variable x becomes a penalty parameter on the Wolfe dual. Now for any $\varepsilon > 0$, the point

$$(\tilde{x}, \tilde{u}, \tilde{w} := \varepsilon e)$$

satifies a Slater constraint qualification for the dual problems (3.14)-(3.15') for $\beta > \|\tilde{x}\|_\infty$. Hence [10, Theorem 2.3] there exists $(x(\beta), u(\beta), v(\beta), w(\beta))$ which solves the dual problems (3.14)-(3.15) with equal extrema. For any such solution, $(u(\beta), v(\beta))$ is bounded by [10, Theorem 2.2]

(3.16)
$$\|u(\beta), v(\beta)\|_1 \leqq \frac{-\tilde{u}g(\tilde{x}) + \beta\varepsilon ee + \tilde{x}\nabla_x L(\tilde{x},\tilde{u})}{\min_{i,j} \{-g_i(\tilde{x}), \tilde{x}_j\}}$$

Since the left side of (3.16) does not depend on ε, we can let $\varepsilon \to 0$ in (3.16) and we have

(3.17) $$\|u(\beta), v(\beta)\|_1 \leqq \frac{-\tilde{u}g(\tilde{x}) + \tilde{x}\nabla_x L(\tilde{x},\tilde{u})}{\min_{i,j} \{-g_i(\tilde{x}), \tilde{x}_j\}}$$

Define now

(3.18) $$\phi(u,v) := \inf_{x \in R^n} L(x,u) - vx$$

(3.19) $$\psi(u,v,w) := \inf_{x \in R^n} L(x,u) - vx + wx$$

Then

(3.20) $$\phi(u,v) = \psi(u,v,0)$$

Note now that by the weak duality theorem [5]

$$\infty > f(x) \quad \sup_{(x,u) \in Z} L(x,u) - x\nabla_x L(x,u)$$

Hence for an unbounded increasing sequence of positive numbers $\{\beta_i\}$ exceeding $\|\tilde{x}\|_\infty$, it follows [10, Theorem 2.3] that there exists a sequence of points $\{x(\beta_i), u(\beta_i), v(\beta_i), w(\beta_i)\}$ which solve the dual pair (3.14)-(3.15) for $\beta = \beta_i$, giving equal extrema and such that $\{u(\beta_i), v(\beta_i)\}$ is bounded by (3.17). Since $ew(\beta_i) = e(-\nabla_x L(x(\beta_i), u(\beta_i)))_+$ constitutes the penalty term for (3.15"), it follows by (2.7) that $\{ew(\beta_i)\}$ converges to zero and since $w(\beta_i) \geqq 0$, it follows that $\{w(\beta_i)\}$ also converges to $\bar{w} = 0$. Let $(\bar{u},\bar{v},0)$ be an accumulation point of the bounded sequence $\{u(\beta_i), v(\beta_i), w(\beta_i)\}$. Now we have

$$c := L(\tilde{x},\tilde{u}) - \tilde{x}\nabla_x L(\tilde{x},\tilde{u}) \leqq \inf_{x \in X} f(x) \qquad \text{(By weak duality)}$$

$$\leqq f(x(\beta_i)) \qquad \text{(Since } x(\beta_i) \in X)$$

$$\leqq L(x(\beta_i), u(\beta_i)) - v(\beta_i)x(\beta_i) + w(\beta_i)x(\beta_i)$$

(Since $u(\beta_i)g(x(\beta_i)) = 0$, $v(\beta_i)x(\beta_i) = 0$ and $w(\beta_i)x(\beta_i) \geqq 0$)

$$\leqq L(x, u(\beta_i)) - v(\beta_i)x + w(\beta_i)x \quad \forall x \in R^n$$

(Since $\nabla_x L(x(\beta_i), u(\beta_i)) - v(\beta_i) + w(\beta_i) = 0$

$L(x, u(\beta_i)) - v(\beta_i)x + w(\beta_i)x$ is convex in x)

In the limit we have

$$c \leqq L(x,\bar{u}) - \bar{v}x + \bar{w}x \quad \forall x \in R^n$$

and so (since $\bar{w} = 0$)

$$c \leqq \inf_{x \in R^n} L(x,\bar{u}) - \bar{v}x + \bar{w}x = \psi(\bar{u},\bar{v},\bar{w}) = \phi(\bar{u},\bar{v})$$

Since $\psi(\bar{u},\bar{v},\bar{w})$ is finite, it follows by Theorem A.1 of the Appendix, that $\psi(u,v,w)$ is upper semicontinuous at $(\bar{u},\bar{v},\bar{w})$ with respect to R_+^{m+2n}. Now let $\{\varepsilon_j\} \downarrow 0$. It follows by the upper semicontinuity of $\psi(u,v,w)$ at $(\bar{u},\bar{v},\bar{w})$ that there exists a subsequence $\{\beta_{i_j}\} \uparrow \infty$ of the unbounded increasing sequence $\{\beta_i\}$ such that $\{u(\beta_{i_j}), v(\beta_{i_j}), w(\beta_{i_j})\}$ converges to $(\bar{u},\bar{v},\bar{w} = 0)$ and

(3.21) $$\phi(\bar{u},\bar{v}) + \varepsilon_j = \psi(\bar{u},\bar{v},\bar{w}) + \varepsilon_j$$

$$> \psi(u(\beta_{i_j}), v(\beta_{i_j}), w(\beta_{i_j}))$$

(By usc of ψ at $(\bar{u},\bar{v},\bar{w})$)

$$= \inf_x L(x, u(\beta_{i_j})) - v(\beta_{i_j})x + w(\beta_{i_j})x$$

(By definition of ψ)

$$= L(x(\beta_{i_j}), u(\beta_{i_j})) - v(\beta_{i_j})x(\beta_{i_j}) + w(\beta_{i_j})x(\beta_{i_j})$$

(Since $x(\beta_{i_j})$ minimizes $L(x, u(\beta_{i_j})) - v(\beta_{i_j})x + w(\beta_{i_j})x$)

$$\geq f(x(\beta_{i_j}))$$

(Since $u(\beta_{i_j})g(x(\beta_{i_j})) = 0,\ v(\beta_{i_j})x(\beta_{i_j}) = 0$

and $w(\beta_{i_j})x(\beta_{i_j}) \geq 0$)

$$\geq L(x(\beta_{i_j}), u(\beta_{i_j})) - vx(\beta_{i_j}) \quad \text{for} \quad (u,v) \geq 0$$

(Since $g(x(\beta_{i_j}) \leq 0$ and $x(\beta_{i_j}) \geq 0$)

$$\geq \phi(u,v) \quad \text{(By definition of } \phi)$$

Note that for $\{\beta_{i_j}\} \uparrow \infty$, the sequence $\{f(x(\beta_{i_j}))\}$ of minima of (3.14) with $\beta = \beta_{i_j}$, constitutes a nonincreasing sequence bounded below by $\inf_{x \in X} f(x)$. Hence $\{f(x(\beta_{i_j})\}$ converges and

$$\text{(3.22)} \qquad \inf_{x \in X} f(x) \leq \lim_{j \to \infty} f(x(\beta_{i_j}))$$

Letting $\varepsilon_j \to 0$ in the string of inequalities of (3.21) gives

$$\phi(\bar{u},\bar{v}) \geq \lim_{j \to \infty} f(x(\beta_{i_j})) \geq \phi(u,v) \qquad \forall (u,v) \geq 0$$

Hence

$$\text{(3.23)} \quad \phi(\bar{u},\bar{v}) = \lim_{i \to \infty} f(x(\beta_{i_j})) = \max_{(u,v) \geq 0} \phi(u,v) = \max_{(u,v) \geq 0} \inf_{x \in R^n} L(x,u) - vx$$

and $(\bar{u},\bar{v})$ solves the Lagrangian dual problem (3.3). The bound (3.12) on $(\bar{u},\bar{v})$ follows from (3.17). To show a zero duality gap, just note that

$$\inf_{x \in X} f(x) \leq \lim_{j \to \infty} f(x(\beta_{i_j})) = \max_{(u,v) \geq 0} \phi(u,v) \leq \inf_{x \in X} f(x)$$

where the first inequality follows from (3.22), the equality from (3.23) and the last inequality from the weak duality theorem for the Lagrangian dual [4,1]. Hence

$$\inf_{x \in X} f(x) = \max_{(u,v) \geqq 0} \phi(u,v) \qquad \square$$

We remark that the existence part of this theorem and the zero duality gap result can also be derived as a consequence of the strong duality theorem of Lagrangian duality (e.g. [4, Theorem 3]) which is based on the entirely different argument of a separating hyperplane. Our explicit bound on the dual optimal variables (3.12) however does not follow from Lagrangian duality and is based on the recent boundedness results of [10].

4. Penalty Functions in Linear Programming

In this final section we show how to use penalty function results to determine precisely the value of the parameter in the quadratic perturbation to a linear program [6,7,8] in order to obtain a solution to the perturbed problem which is dual feasible to within any preassigned tolerance. This is a practical and important issue which has not been completely resolved before in the iterative successive overrelaxation (SOR) methods for solving huge sparse linear programs [8].

We consider the primal linear program

$$(4.1) \qquad \max_{x}\ cx \quad \text{s.t. } Ax \leqq b,\ x \geqq 0$$

where A is given $m \times n$ real matrix, $c \in R^n$ and $b \in R^m$, and its dual

$$(4.2) \qquad \min_{(u,v)}\ bu \quad \text{s.t. } v = A^T u - c,\ u, v \geqq 0$$

In [8] it has been shown that perturbed primal program

$$(4.3) \qquad \max_{x}\ cx - \frac{\varepsilon}{2}xx \quad \text{s.t. } Ax \leqq b,\ x \geqq 0$$

is solvable for all $\varepsilon \in (0,\bar{\varepsilon}]$ for some $\bar{\varepsilon}$ if and only if (4.1) is solvable, in which case the unique solution $\bar{x}$ of (4.3) for $\varepsilon \in (0,\bar{\varepsilon}]$ is independent of ε and is the point in the solution set of (4.1) with least 2-norm. If we consider the Wolfe dual to (4.3) we obtain

$$(4.4) \qquad \min_{(x,u,v)}\ bu + \frac{\varepsilon}{2}xx \quad \text{s.t. } c - \varepsilon x - A^T u + v = 0,\ u, v \geqq 0$$

Elimination of x through the constraint relation

$$(4.5) \qquad x = \frac{1}{\varepsilon}(-A^T u + v + c)$$

gives

$$(4.6)\qquad \min_{(u,v)\geqq 0} \; bu + \frac{1}{2\varepsilon}\|-A^T u + v + c\|_2^2$$

which is precisely the exterior penalty function associated with the dual linear program (4.2) with penalty parameter $\frac{1}{\varepsilon}$. Using standard exterior penalty function results, one needs that $\varepsilon \to 0$ in order for solutions $(u(\varepsilon), v(\varepsilon))$ of (4.6) to approach a solution of the dual linear problem (4.2). However by computing $\bar{x}$ from $(u(\varepsilon), v(\varepsilon))$ through the relation (4.5), it turns out [8] that for $\varepsilon \in (0,\bar{\varepsilon}]$, $\bar{x}$ is independent of ε and is the unique point in the solution set of (4.1) with least 2-norm. In [8] SOR methods were prescribed for solving (4.6) for ε sufficiently small and then computing x from (4.5). Very large sparse problems ($n = 20{,}000$, $m = 5{,}000$) were solved by this technique, without knowing what $\bar{\varepsilon}$ is, but merely by decreasing ε until certain approximate optimality criteria were met. We would like to show here that by solving the penalty problem (4.6) for only <u>two</u> values of ε, we can satisfy the Karush-Kuhn-Tucker optimality conditions for the linear program to any preassigned tolerance. In fact such a solution will be primal feasible, satisfy the complementarity conditions between primal and dual linear programs, and satisfy dual feasibility to any required tolerance. More specifically we have the following.

4.1 <u>Theorem</u> Let $\delta > 0$, $\varepsilon_1 > 0$, let $(\hat{u},\hat{v})$ be dual feasible, that is $\hat{v} = A^T\hat{u} - c \geqq 0$, $\hat{u} \geqq 0$, and let $(u(\varepsilon_1), v(\varepsilon_1))$ be a solution of (4.6) with $\varepsilon = \varepsilon_1$. If $b\hat{u} \leqq bu(\varepsilon_1)$ then $(\hat{u},\hat{v})$ solves the dual problem (4.2), else for

$$(4.7)\qquad \varepsilon_2 < \varepsilon_1 \quad \text{and} \quad \varepsilon_2 \leqq \frac{\delta}{b\hat{u} - bu(\varepsilon_1)}$$

it follows that

$$(4.8)\qquad \frac{1}{2}\|-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c\|_2^2 \leqq \delta,\ bu(\varepsilon_2) \leqq \min_u \{bu \mid A^Tu \geqq c, u \geqq 0\}$$

where $(u(\varepsilon_2), v(\varepsilon_2))$ is a solution of (4.6) with $\varepsilon = \varepsilon_2$. Furthermore for $x(\varepsilon_2)$ defined by

$$(4.9)\qquad x(\varepsilon_2) := \frac{1}{\varepsilon_2}(-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c)$$

we have that the Karush-Kuhn-Tucker conditions for the linear program (4.1) are satisfied to within a tolerance δ as follows

$$(4.10)\qquad \left\langle \begin{array}{l} x(\varepsilon_2) \geqq 0,\ Ax(\varepsilon_2) \leqq b,\ u(\varepsilon_2) \geqq 0,\ v(\varepsilon_2) \geqq 0 \\ u(\varepsilon_2)(b - Ax(\varepsilon_2)) = 0,\ v(\varepsilon_2)x(\varepsilon_2) = 0 \\ \|-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c\|_2 \leqq (2\delta)^{\frac{1}{2}} \end{array} \right.$$

<u>Proof</u> The first part of the theorem, (4.7)-(4.8), follows directly from Theorem 2.3. The last part of the theorem (4.10) follows from (4.8) and from the Karush-Kuhn-Tucker optimality conditions for (4.6) with $\varepsilon = \varepsilon_2$, that is

$$(4.11)\qquad \left\langle \begin{array}{l} b - \frac{1}{\varepsilon_2} A(-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c) \geqq 0,\ u(\varepsilon_2) \geqq 0 \\ u(\varepsilon)(b - \frac{1}{\varepsilon_2} A(-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c) = 0 \\ \frac{1}{\varepsilon_2}(-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c) \geqq 0,\ v(\varepsilon_2) \geqq 0 \\ \frac{v(\varepsilon_2)}{\varepsilon_2}(-A^Tu(\varepsilon_2)+v(\varepsilon_2)+c) = 0 \end{array} \right.$$

These conditions together with (4.8) and the definition (4.9) imply (4.10). □

Appendix

A.1 Theorem Let $\psi(s) := \inf_{t \in T} h(s,t)$ where $h: S \times T \to R$, $\phi \neq S \subset R^k$, $\phi \neq T \subset R^n$ and h is upper semicontinuous on S with respect to S for each fixed $t \in T$. Then ψ is upper semicontinuous with respect to S at each $\bar{s} \in S$ for which $\psi(\bar{s}) > -\infty$.

Proof Suppose ψ is not usc at $\bar{s}$ with respect to S. Then

$$(A.1) \qquad \exists \varepsilon > 0: \forall \delta > 0 \;\; \exists s(\delta) \in S: \; \|s(\delta) - \bar{s}\| < \delta, \; \psi(s(\delta)) - \psi(\bar{s}) \geqq \varepsilon$$

Let ε be fixed. Since $-\infty < \psi(\bar{s}) = \inf_{t \in T} h(\bar{s},t)$, there exists $t(\varepsilon) \in T$ such that

$$(A.2) \qquad h(\bar{s}, t(\varepsilon)) < \psi(\bar{s}) + \varepsilon$$

Combining (A.1), (A.2) and the definition of ψ gives

$$(A.3) \left\langle \begin{array}{l} h(\bar{s}, t(\varepsilon)) < \psi(\bar{s}) + \varepsilon \leqq \psi(s(\delta)) \leqq h(s(\delta), t(\varepsilon)) \\ \qquad\qquad \forall \delta > 0, \text{ for some } s(\delta) \in S \text{ such that } \|s(\delta) - \bar{s}\| < \delta \end{array} \right.$$

Since $h(s, t(\varepsilon))$ is usc with respect to S at $\bar{s} \in S$ we have

$$(A.4) \qquad \forall \gamma > 0, \exists \delta(\gamma) > 0: \forall s \in S \;\; \|s - \bar{s}\| < \delta(\gamma), \; h(s, t(\varepsilon)) < h(\bar{s}, t(\varepsilon)) + \gamma$$

Combining (A.3) and (A.4) gives

$$(A.5) \qquad h(\bar{s}, t(\varepsilon)) < \psi(\bar{s}) + \varepsilon < h(\bar{s}, t(\varepsilon)) + \gamma \qquad \forall \gamma > 0$$

Since $\bar{s}$ and ε do not depend on γ, (A.5) gives a contradiction by letting γ approach zero. Hence ψ is usc at $\bar{s}$ with respect to S. $\square$

Acknowledgement

I am indebted to my colleague Lynn McLinden for helpful discussions on duality theory.

References

1. M. S. Bazaraa and C. M. Shetty: "Nonlinear programming theory and algorithms", Wiley, New York 1979.

2. E. J. Beltrami: "An algorithmic approach to nonlinear analysis and optimization", Academic Press, New York 1970.

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

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

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

6. O. L. Mangasarian: "Iterative solution of linear programs", SIAM Journal on Numerical Analysis 18, 1981, 606-614.

7. O. L. Mangasarian: "Sparsity-preserving SOR algorithms for separable quadratic and linear programming problems", Computers and Operations Research 11, 1984, 105-112.

8. O. L. Mangasarian: "Normal solutions of linear programs", Mathematical Programming Study 22, 1984, 206-216.

9. O. L. Mangasarian: "A condition number for differentiable convex inequalities", Mathematics of Operations Research 10, 1985, 175-179.

10. O. L. Mangasarian and L. McLinden: "Simple bounds for solutions of monotone complementarity problems and convex programs", Mathematical Programming 32, 1985, 32-40.

11. G. P. McCormick: "Nonlinear programming theory, algorithms and applications", Wiley, New York 1983.

12. P. Wolfe: "A duality theorem for nonlinear programming", Quarterly of Applied Mathematics, 19, 1961, 239-244.