THE $L_\infty$ EXACT PENALTY FUNCTION IN SEMI-INFINITE PROGRAMMING

by

C.J. PRICE and I.D. COOPE

Department of Mathematics, University of Canterbury,
Christchurch, New Zealand.

No. 51 November, 1989.

Abstract

An algorithm for semi-infinite programming using sequential quadratic programming techniques together with an $L_\infty$ exact penalty function is presented, and global convergence is shown. An important feature of the convergence proof is that it does not require an implicit function theorem to be applicable to the semi-infinite constraints; a much weaker assumption concerning the finiteness of the number of global maximizers of each semi-infinite constraint is sufficient. In contrast to proofs based on an implicit function theorem, this result is also valid for a large class of $C^1$ problems.
THE $L_\infty$ EXACT PENALTY FUNCTION IN SEMI-INFINITE PROGRAMMING

C. J. PRICE and I. D. COOPE

Abstract.

An algorithm for semi-infinite programming using sequential quadratic programming techniques together with an $L_\infty$ exact penalty function is presented, and global convergence is shown. An important feature of the convergence proof is that it does not require an implicit function theorem to be applicable to the semi-infinite constraints; a much weaker assumption concerning the finiteness of the number of global maximizers of each semi-infinite constraint is sufficient. In contrast to proofs based on an implicit function theorem, this result is also valid for a large class of $C^1$ problems.

1 Introduction.

Semi Infinite Programming (SIP) problems arise in many practical problems such as computer aided design, production planning, and the like. Many algorithms for solving such problems have been proposed. A common approach which yields global convergence is the use of Sequential Quadratic Programming (SQP) techniques in conjunction with an exact penalty function [2,4,8,9]. The methods given in [4,8,9] use an implicit function theorem on each semi-infinite constraint to demonstrate convergence. The $L_1$ exact penalty function algorithm of Conn and Gould [2] is along somewhat different lines, but makes use of similarly restrictive assumptions. The purpose of this paper is to show that provided the exact penalty function is based on the infinity norm, a much weaker condition than that required for the implicit function theorem to hold is sufficient to ensure convergence for $C^1$ problems. The algorithm presented can take second order information into account, yielding superlinear convergence on problems with the requisite degree of continuity.

The SIP considered is of the form:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ subject to } g(x,t) \leq 0 \quad \forall t \in T, \text{ where } T \subset \mathbb{R}^p. \quad (1)$$

The objective function $f$, mapping $\mathbb{R}^n$ into $\mathbb{R}$, and the constraint function $g$, mapping $\mathbb{R}^n \times T$ into $\mathbb{R}$, are both continuously differentiable in all arguments. The set $T$ is compact, connected, and defined by a finite number of continuously differentiable constraints which satisfy an appropriate constraint qualification. Frequently $T$ is a Cartesian product of intervals. For convenience the problem has been restricted to one semi-infinite constraint, and auxiliary constraints have been omitted.
2 The penalty function problem.

The approach taken is to replace the SIP with the rather more tractable problem of minimizing a non-differentiable penalty function chosen so that solutions of the SIP are also solutions of the Penalty Function Problem (PFP). The exact penalty function used is:

\[ \phi(\mu, \nu; x) = f(x) + \mu \theta + \frac{\nu}{2} \theta^2 \text{ where } \theta = \max_{t \in I} [g(x, t)]_+ . \]

Clearly \( \theta(x) \) is the infinity norm of the constraint violations. The penalty parameters \( \mu \) and \( \nu \) are restricted to the ranges \( \mu > 0 \), and \( \nu \geq 0 \).

The algorithm to be described uses only first derivatives: accordingly it is desirable that the algorithm be capable of solving \( C^1 \) problems. This precludes the use of second order optimality conditions in specifying solutions of an arbitrary problem of the form (1). Consequently stationary points of the SIP will be regarded as valid solution points. The first order optimality conditions, together with an appropriate regularity assumption, are as follows.

**Theorem 2.1** Let \( x^* \) be any optimal point of the SIP, and let the following regularity assumption hold at \( x^* \):

\[ \exists u \in \mathbb{R}^n \text{ such that } g(x^*, t) + u^T \nabla_x g(x^*, t) < 0, \forall t \in T. \quad (2) \]

Then there exists a finite number of global maximizers \( \tau_i^* \) of \( g(x^*, t) \), each with an associated Lagrange multiplier \( \lambda_i^* \), and satisfying

\[ \nabla f + \sum_{i=1}^{m} \lambda_i^* \nabla_x g(x^*, \tau_i^*) = 0 \text{ with } m \leq n, \quad (3) \]

where \( g(x^*, \tau_i^*) = 0, \text{ and } \lambda_i^* \geq 0, \forall i = 1, \ldots, m. \quad (4) \]

**Proof.** By lemmas 2 and 3 of [6]. \( \square \)

Points at which no direction of descent exists for \( \phi \) are accepted as solutions to the PFP.

**Definition 2.2** A point \( x \) is a critical point of the penalty function \( \phi \) iff,

\[ \text{for every unit vector } u \in \mathbb{R}^n, \quad \lim_{\epsilon \to 0^+} \frac{\phi(x + \epsilon u) - \phi(x)}{\epsilon} \geq 0. \quad \square \]

If solving the PFP is to yield a solution of the SIP, it is highly desirable that the solution set for the PFP be contained in (and ideally be equal to) the solution set for the SIP. This can be achieved to a limited extent by a suitable choice of \( \mu \), for any \( \nu \geq 0 \).
THEOREM 2.3 Let \( x^* \) be an optimal point of the SIP (1) at which the regularity assumption (inequality (2)) holds, and let \( \lambda^* \) be the vector of Lagrange multipliers as specified in (3,4). If \( \mu \) satisfies

\[
\mu > \|\lambda^*\|_1
\]

then \( x^* \) is a critical point of \( \phi(\mu, \nu; x) \).

Conversely, if \( x^* \) is both feasible, and a critical point of \( \phi(\mu, \nu; x) \) for some \( \mu > 0 \), and \( \nu \geq 0 \), then \( x^* \) is a solution point of the SIP.

PROOF. The first item follows from theorem 2.1 of [1], and from theorem (2.1). For the second item, if \( x^* \) is a critical point of \( \phi \) for some \( \mu \), and \( \nu \), then

\[
\forall x \text{ near } x^*, \quad \phi(\mu, \nu; x) \geq \phi(\mu, \nu; x^*) + o(||x - x^*||).
\]

Now \( \phi \equiv f \) on the SIP's feasible region, and so \( x^* \) is a solution point for the SIP.

This theorem implies the set of feasible critical points of the PFP are a subset of the set of stationary points of the SIP. The relationship between the two solution sets falls short of the ideal in two respects.

Firstly, there may be critical points which are not feasible, and therefore not solutions to the SIP. This admits the possibility that the algorithm may fail to solve the SIP by (in essence) failing to find a feasible point. This is characteristic of any algorithm attempting to attain feasibility from an arbitrary initial point by seeking a local minimum of the constraint violations. If the algorithm fails for this reason a common response is to consider other initial points.

Secondly, there may be solution points of the SIP which are not feasible critical points. This problem is circumvented by automatically adjusting \( \mu \) so that any SIP solution is a critical point provided it is sufficiently close to some iterate.

3 Existence of an approximating \( L_{\infty} \) QP.

It has been shown in the previous section that the SIP may be replaced by the problem of locating feasible solutions of the PFP. The PFP is tackled as follows. At each iterate linear approximations to all global (and some local) maximal values of the constraint function are formed. From these a local approximation to \( \theta(x) \) can be constructed. This, together with an approximation to the objective function, yields an approximation to \( \phi \), and hence an \( L_{\infty} \) Quadratic Programme locally approximating the PFP. The solution of this \( L_{\infty} \) QP yields a search direction along which the next iterate is sought, using an Armijo type line search.

In order to ensure each iteration of the algorithm listed in section 4 is a finite computational process, the following assumption is made.

ASSUMPTION 3.1 For each \( x \in R^n \), the number of global maximizers of \( g(x, t) \) over \( T \) is finite. \( \Box \)
This, together with the other usual assumptions, is also sufficient to ensure the convergence of the algorithm; use of an implicit function theorem is superfluous. Actually, it is sufficient that the number of global maximizers of \( g \) is finite at each point \( x \) at which approximations to the global maximizers are calculated explicitly, and at each cluster point of the sequence of iterates. For convenience, assumption 3.1 will be used.

The existence of an approximating \( L_{\infty} \)QP is shown by examining the behaviour of the set of global maximizers \( \Gamma(x) \) of \( g(x, t) \) at points \( x \) near some point \( x_c \) satisfying assumption 3.1. The first result states \( \Gamma(x) \) is semi-continuous with respect to \( x \).

**Proposition 3.2** Let \( C \) be a compact subset of \( T \), and let \( \Omega(x_c) \) be the set of global minimizers of \( g(x_c, t) \) on \( C \). If \( \Omega(x_c) \) is a subset of the interior of \( C \) relative to \( T \) (hereafter \( \text{int}(C) \)), then firstly

\[
\forall \epsilon > 0, \ \exists \eta(\epsilon) > 0 : \forall x \in \mathbb{R}^n, \ |x - x_c| < \eta \Rightarrow \Omega(x) \subset \mathcal{N}_\epsilon(\Omega(x_c)),
\]

where \( \mathcal{N}_\epsilon(\Omega(x_c)) = \{ t \in T : \exists \gamma \in \Omega(x_c) \text{ satisfying } |t - \gamma| < \epsilon \} \),

and secondly, each element of \( \Omega(x) \) is a local maximizer of \( g(x, t) \) over \( T \), for all \( x \) sufficiently near \( x_c \).

**Proof.** Use the topology on \( T \) induced by the standard topology on \( \mathbb{R}^n \). Let \( g_c \) be the global maximal value of \( g(x_c, t) \) on \( C \). For all small positive \( \epsilon \), as \( C - \mathcal{N}_\epsilon(\Omega(x_c)) \) is compact and non-empty, \( g(x_c, t) \) achieves its supremum on \( C - \mathcal{N}_\epsilon(\Omega(x_c)) \), which must be strictly less than \( g_c \). Define

\[
m(\epsilon) = g_c - \max_{t \in C - \mathcal{N}_\epsilon(\Omega(x_c))} g(x_c, t).
\]

Now the continuity of \( \nabla_x g \) with respect to all arguments, and the compactness of \( T \) imply the set of functions \( \{g(x, t)\}_{t \in T} \) is equicontinuous with respect to \( x \). Therefore

\[
\forall \epsilon > 0, \ \exists \eta(\epsilon) > 0, \text{ such that } \forall x, \text{ and } \forall t \in T,
\]

\[
|x - x_c| < \eta(\epsilon) \Rightarrow |g(x_c, t) - g(x, t)| < \frac{1}{4} m(\epsilon).
\]

Hence, for all these values of \( x \),

\[
\forall t \in C - \mathcal{N}_\epsilon(\Omega(x_c)), \quad g(x, t) < g_c - \frac{3}{4} m(\epsilon),
\]

and \( \forall t \in \Omega(x_c), \quad g(x, t) > g_c - \frac{1}{4} m(\epsilon). \)

Hence, \( \Omega(x) \subset \mathcal{N}_\epsilon(\Omega(x_c)) \).

Moreover, as \( g \) is continuous, and \( C \) compact, \( \Omega(x_c) \) is also compact. Whence, for all small positive \( \epsilon \), \( \mathcal{N}_\epsilon(\Omega(x_c)) \subset \text{int}(C) \), and so \( \Omega(x) \) is a subset of the local maximizers of \( g(x, t) \) over \( T \).  \( \square \)
For any $x_0 \in \mathbb{R}^n$, $\Gamma(x_0) = \{\tau_1, \ldots, \tau_j\}$ is the finite set of strict global maximizers of $g(x_0,t)$. Proposition 3.2 implies each member of $\Gamma(x_0)$ may be considered separately. Let

$$
epsilon_0 = \frac{1}{4} \min\{\|\tau_i - \tau_k\| : i, k \in 1, \ldots, j, \ i \neq k\},$$

and let $B_i(\epsilon_0) = \{t \in T : \|t - \tau_i\| \leq \epsilon_0\}, \ \forall i = 1, \ldots, j.$

The set of global maximizers of $g(x,t)$ on the set $B_i(\epsilon_0)$ is denoted by $\Xi_i(x)$. The behaviour of $\Gamma$ with respect to changes in $x$ is examined by considering each $\Xi_i$ along each ray of the form $x(\sigma) = x_0 + \sigma u$, where $\sigma \geq 0$, and $u$ is a unit vector in $\mathbb{R}^n$.

**Definition 3.3** A function $t(\sigma)$ is an extension of the global maximizer $\tau_i \in \Gamma(x_0)$ along $x(\sigma) = x_0 + \sigma u$, where $\sigma \geq 0$, iff

1. $t(0) = \tau_i$.
2. $\exists \sigma_{\text{max}} > 0$ such that $t(\sigma) \in \Xi_i(x(\sigma)), \ \forall \sigma \in [0, \sigma_{\text{max}}]$. □

From proposition 3.2, each $\tau_i$ has at least one extension for each $u$. It may have several, or even an infinite number of extensions. The extensions may be discontinuous functions. Proposition 3.2 implies that, for all $x$ near $x_0$, the extensions of $\Gamma(x_0)$, evaluated at $x$, are local maximizers of $g(x,t)$ over $T$, and contain $\Gamma(x)$. The extensions in the direction $u$ of the members of $\Gamma(x_0)$ yield the following set of values of $g$ along the ray $x(\sigma)$:

$$\{g(x(\sigma), t(\sigma)) : t(\sigma) \text{ is an extension of some } \tau_i \in \Gamma(x_0)\}.$$ 

This set is finite; any two extensions of the same $\tau_i$ take the global maximal value of $g(x(\sigma),t)$ over $B_i(\epsilon_0)$, for all $\sigma$. For $i = 1, \ldots, j$ let $t_i(\sigma)$ be an extension of $\tau_i$. Each member of the set $\{g(x(\sigma), t_i(\sigma))\}_{i=1}^j$ is locally Lipschitz with respect to $\sigma$ by the $C^1$ continuity of $g$, and the compactness of $T$. In order to form a set of linear approximations to $\{g(x(\sigma), t_i(\sigma))\}$, the following result is needed.

**Proposition 3.4** Let $t_i(\sigma)$ be any extension of $\tau_i \in \Gamma(x_0)$ along the ray $x(\sigma) = x_0 + \sigma u, \ \sigma \geq 0$. Then

$$g(x(\sigma), t_i(\sigma)) = g(x_0, \tau_i) + \sigma u^T \nabla_x g(x_0, \tau_i) + o(\sigma).$$

**Proof.**

$$g(x(\sigma), t_i(\sigma)) \geq g(x(\sigma), \tau_i), \ \Rightarrow$$

$$g(x(\sigma), t_i(\sigma)) \geq g(x_0, \tau_i) + \sigma u^T \nabla_x g(x_0, \tau_i) + o(\sigma). \ (6)$$

Also,

$$g(x(\sigma), t_i(\sigma)) = g(x_0, t_i(\sigma)) + \sigma u^T \nabla_x g(x_0, t_i(\sigma)) + o(\sigma)$$

$$\leq g(x_0, \tau_i) + \sigma u^T \nabla_x g(x_0, t_i(\tau_i)) + o(\sigma).$$
Now, as $\Xi(x_0)$ is a singleton set, proposition 3.2 implies every extension of $\tau_i$ is right continuous at $\sigma = 0$. Hence
\[ g(x(\sigma), t_i(\sigma)) \leq g(x_0, \tau_i) + \sigma u^T \nabla_x g(x_0, \tau_i) + o(\sigma) \]
This, and inequality (6) yield the required result. □

Define $\psi$ to be a continuous piecewise quadratic approximation to $\phi$ near $x_0$, where $\psi$ is based on the finite subset $A_0$ of $T$, as follows
\[ \psi(x_0, A_0; \mu, \nu; s) = f(x_0) + s^T \nabla f(x_0) + \frac{1}{2} s^T H s + \mu \vartheta(s) + \frac{1}{2} \nu \vartheta^2(s), \]
where $\vartheta(s) = \max_{t \in A}[g(x_0, t) + s^T \nabla_x g(x_0, t)]^+$,
and where $H$ is positive definite. Clearly $\psi$ is strictly convex in $s$.

Let the base set $A_0$ be $\{t_i\}_{i=1}^r$. For each $i = 1, \ldots, r$ define row $i$ of the matrix $B$ as $B_i = [\nabla_x g(x_0, t_i)]^T$, and define element $i$ of the vector $b$ to be $b_i = g(x_0, t_i)$.

**Theorem 3.5** If $\Gamma(x_0) \subseteq A_0$ then, for all $s \in \mathbb{R}^n$ such that $\|s\|$ is small,
\[ \phi(\mu, \nu; x_0 + s) = \psi(x_0, A_0; \mu, \nu; s) + o(\|s\|). \]  

**Proof.** For all $s$ sufficiently small, each element of $B_s + b$ arising from some member of $A_0 - \Gamma(x_0)$ is less than every element of $B_s + b$ arising from some member of $\Gamma(x_0)$; thus $A_0 - \Gamma(x_0)$ can be disregarded for small $s$.

The set of extensions of $\Gamma(x_0)$, evaluated at $x_0 + s$, contains $\Gamma(x_0 + s)$ for $s$ small, so proposition 3.4 implies
\[ \forall \tau \in \Gamma(x_0 + s), \exists t_0 \in \Gamma(x_0) \text{ such that } g(x_0 + s, \tau) = g(x_0, t_0) + s^T \nabla_x g(x_0, t_0) + o(\|s\|). \]
Hence
\[ \theta(x_0 + s) = \max_{t \in A}[g(x_0, t_0) + s^T \nabla_x g(x_0, t_0)]^+ + o(\|s\|). \]
Using a linear approximation to the objective function, the result follows. □

The $L^\infty QP$
\[ \min_{s \in \mathbb{R}^n} \psi(x_0, A_0; \mu, \nu; s) \]  
approximates the PFP near $x_0$. If $\Gamma(x_0) \subseteq A_0$, then $x_0$ is a critical point of $\phi$ iff $s = 0$ is the global minimizer of $\psi(x_0, A_0; \mu, \nu; s)$.

### 4 An $L^\infty$-norm algorithm for SIP.

The previous section examined the $L^\infty QP$ in detail. In this section the remainder of the algorithm is discussed, and the algorithm is presented.

At each iterate $x^{(k)}$ the global (and other local) maximizers of the constraint function are found, and the approximating $L^\infty QP$ is constructed. The solution
\( s^{(k)} \) to the \( L_\infty \)QP at \( x^{(k)} \) is used to form the line (or arc) search. The algorithm either searches along the line \( x^{(k)} + \alpha s^{(k)} \), or along the arc \( x^{(k)} + \alpha s^{(k)} + \alpha^2 c^{(k)} \), where \( c^{(k)} \) is a correction vector chosen to prevent the Maratos effect [7]. In either case \( \alpha \) is chosen to be the first member of the sequence \( 1, \beta, \beta^2, \ldots \) to satisfy the sufficient descent criterion

\[
\phi(x^{(k)}) - \phi(x^{(k)} + q^{(k)}(\alpha)) \geq \rho \alpha \left[ \psi(x^{(k)}; 0) - \psi(x^{(k)}; s^{(k)}) \right],
\]

where \( 0 < \rho < \frac{1}{2}, 0 < \beta < 1 \), and \( q^{(k)}(\alpha) \) is either the line or arc step as given above. The next iterate is then \( x^{(k)} + q^{(k)}(\alpha^{(k)}) \). For convenience the line search is treated hereafter as an arc search with \( c^{(k)} = 0 \).

The penalty parameters are adjusted in order to satisfy (5), and (hopefully) to force the sequence of constraint violations \( \theta^{(k)} \) to zero. The first requirement is met by forming lower semi-continuous estimates \( \lambda_{est}^* \) of the optimal Lagrange multipliers at each iterate and adjusting the penalty parameters accordingly. Such estimates may be calculated from the \( L_\infty \)QP's solution, or by other methods [5].

**Algorithm Summary:**

1. Coarse approximations to all global maximizers, and as many local maximizers as practicable are found using a grid search, and then refined using a Quasi-Newton method. Call this set of points \( \mathcal{A}^{(k)} \).

2. The approximating \( L_\infty \)QP is formed, and its solution \( s^{(k)} \) is calculated. If necessary the penalty parameters are increased to ensure \( \theta(s^{(k)}) \leq \theta(0) \).

3. If \( x^{(k)} + s^{(k)} \) does not satisfy the sufficient descent condition, calculate \( c^{(k)} \), and perform the arc search.

4. Estimate the optimal Lagrange multipliers at the new iterate. If \( \theta \) is less than some positive parameter \( \pi_1 \), and if \( \mu \leq \pi_2 \| \lambda_{est}^* \|_1 \), then \( \mu \) is increased to \( \pi_3 \| \lambda_{est}^* \|_1 \), where \( \pi_3 > \pi_2 > 1 \) are fixed parameters. Related research [3] suggests that \( \pi_3 < 2 \) may be desirable. If \( \theta \geq \pi_1 \), and \( \mu + \nu \theta \leq \pi_4 \| \lambda_{est}^* \|_1 \), then \( \nu \) is adjusted to give \( \mu + \nu \theta = \pi_5 \| \lambda_{est}^* \|_1 \), where \( \pi_5 > \pi_4 > 1 \).

5. Update \( H \) by a Quasi-Newton (or other) scheme satisfying the following requirement.

\[
\exists \gamma_1, \gamma_2 > 0, \text{ such that } \forall x \in \mathbb{R}^n, \forall k, \gamma_1 x^T x \leq x^T H^{(k)} x \leq \gamma_2 x^T x, \quad (10)
\]

where \( \gamma_1 \), and \( \gamma_2 \) are independent of \( k \).

6. If sufficient accuracy has not been attained, another iteration is begun.

The vector \( c^{(k)} \) is essentially that of [7], and is determined as follows. The global optimization subalgorithm is applied to \( g(x^{(k)} + s^{(k)}, t) \), yielding the set \( \mathcal{A}_M^{(k)} \). Let \( Q^{(k)} \) denote the set of elements \( t \in \mathcal{A}^{(k)} \) satisfying

\[
\theta(s^{(k)}) = g(x^{(k)}, t) + (s^{(k)})^T \nabla_g g(x^{(k)}, t).
\]
Define \( t_M(w) \) to be the closest member of \( A_M^{(k)} \) to \( w \), for each \( w \in \mathcal{Q}^{(k)} \). If \( t_M(w) \) is uniquely defined for every \( w \), if \( t_M \) is a one to one mapping, and if \( \mathcal{Q}^{(k)} \) is non-empty, then \( c^{(k)} \) is chosen as the vector of minimum length satisfying

\[
[c^{(k)}]^T \nabla_x g(x^{(k)}, w) + g(x^{(k)} + s^{(k)}, t_M(w)) = 0, \quad \forall w \in \mathcal{Q}^{(k)}.
\] (11)

Otherwise \( c^{(k)} = 0 \) is used. If the system (11) has no solution, or if \( \|c^{(k)}\| > \|s^{(k)}\| \), then \( c^{(k)} \) is reset to zero.

The vector \( c^{(k)} \) is used to avoid the Maratos effect, and thereby ensure super-linear convergence on problems with the required continuity [7]. The vector \( c^{(k)} \) is not required for convergence; the algorithm will converge for any choice of \( c^{(k)} \) satisfying \( \|c^{(k)}\| \leq \|s^{(k)}\| \), including \( c^{(k)} = 0 \).

**ASSUMPTION 4.1** At each point \( x_0 \) at which the global optimization subalgorithm is used it finds every point in \( \Gamma(x_0) \). Also each \( x \in \mathbb{R}^n \) satisfying assumption 3.1 has a neighborhood \( \mathcal{N}(x) \) such that if \( x_0 \in \mathcal{N}(x) \) then the global optimization subalgorithm finds some extension (evaluated at \( x_0 \)) of each member of \( \Gamma(x) \).

This assumption is an idealization; in practice only approximations to the points referred to in assumption 4.1 will be available. The implications of this will be discussed at the end of the following section.

## 5 Convergence.

In this section the convergence properties of the algorithm are examined.

A requirement for convergence is that each arc search be a finite process. This is so if the descent condition (9) holds for all small positive \( \alpha \). If \( s^{(k)} \) is zero, then \( c^{(k)} \) is also zero, and (9) holds for all \( \alpha \). If \( s^{(k)} \) is non-zero, then from assumption 4.1 and theorem 3.5, and using \( \psi(x^{(k)}, A^{(k)}; s) = \psi^{(k)}(s) \), (9) is equivalent to

\[
\psi^{(k)}(0) - \psi^{(k)}(\alpha s^{(k)}) + o(\alpha) \geq \rho \alpha [\psi^{(k)}(0) - \psi^{(k)}(s^{(k)})]
\] (12)

where, because \( \phi \) is locally Lipschitz, the \( c^{(k)} \) part of \( q^{(k)} \) has been incorporated into the \( o(\alpha) \) term. The strict convexity of \( \psi \) ensures (12) holds for all small positive \( \alpha \).

Before proceeding to the convergence theorem, a preliminary result on the boundedness of the step direction is derived.

**PROPOSITION 5.1** Let \( X \) be a compact subset of \( \mathbb{R}^n \). Let \( A(x) \) be any set function mapping \( X \) into the class of subsets of \( T \), and let \( s(x) \) denote the global minimizer of \( \psi(x, A(x); s) \). Then the set \( \{s(x) : x \in X\} \) is bounded.

**PROOF.** The definition of \( \psi \) implies

\[
\forall x \in X, \quad \psi(x, T; s) \geq \psi(x, A(x); s) \geq \psi(x, \emptyset; s).
\]
Let $s_T(x)$ be the minimizer of $\psi(x, T; s)$. Define

$$\psi_{\text{max}} = \max \{ \psi(x, T; s_T(x)) : x \in X \}.$$ 

Now

$$\psi(x, 0; s) \geq s^T \nabla f(x) + \frac{1}{2} \gamma_1 s^T s$$

by (10). Also $\nabla f(x)$ is bounded on $X$, and so for all $s$ sufficiently large $\psi(x, 0; s) > \psi_{\text{max}}$. Whence $\{s(x) : x \in X\}$ must be bounded. □

**Theorem 5.2** Given:

1. All iterates generated by the algorithm lie in a bounded region of $\mathbb{R}^n$.
2. Assumptions 3.1, 4.1, and the condition (10) hold.
3. The parameters $\mu$ and $\nu$ are only altered a finite number of times.

Then every cluster point of the sequence of iterates $\{x(k)\}$ generated by the algorithm is a critical point of $\phi(\mu, \nu; x)$, where $\mu$ and $\nu$ are the final values of these parameters.

**Proof.** The proof is by contradiction. This is obtained by assuming some cluster point $(x_0, \text{say})$ of the sequence of iterates is not a critical point, and so deducing the existence of an iterate satisfying

$$\phi \left( x_0 + q(\alpha(\kappa)) \right) < \phi(x_0).$$

As the sequence $\{\phi(\kappa)\}$ is monotonically decreasing, and $\phi$ is continuous, a contradiction results.

Let $x_0$ be an arbitrary cluster point of $\{x(k)\}$. Select a subsequence $\{x_0(\kappa)\}$ of $\{x(k)\}$, generated after $\mu$ and $\nu$ assume their final values, and where the subsequences $\{x_0(\kappa)\}$, $\{H_0(\kappa)\}$, and $\{s_0(\kappa)\}$ converge to $x_0$, $H_0$, and $s_0$ respectively. Such a subsequence exists by item 1, requirement (10) and proposition 5.1.

Let $\psi_0(\kappa)$ and $\phi_0(\kappa)$ denote $\psi(x_0(\kappa), \mathcal{A}_0(\kappa); s)$ and $\phi(x_0(\kappa))$ respectively. Let $\mathcal{A}_0$ be the base set for the $L_0$QP at the $x_0$ iterate. Define $\mathcal{A}_0$ as the set of all cluster points of sequences of the form $\{x_i\}_{i=1}^\infty$, where $x_i \in \mathcal{A}_i$, for all $i$. Clearly $\mathcal{A}_0$ is compact. Also

$$\forall \varepsilon > 0, \exists K, \forall k > K, \max_{t \in \mathcal{A}_0} \min_{\tau \in \mathcal{A}_0} \|t - \tau\| < \varepsilon, \quad (14)$$

by the definition of $\mathcal{A}_0$. Let $s_0$ be the global minimizer of $\psi(x_0, \mathcal{A}_0; s)$. So

$$\forall s \neq s_0, \exists t(s) \in \mathcal{A}_0, \text{ such that } \psi(x_0, \{t(s)\}; s) > \psi_0(s_0).$$

Fixing $s$, let $\{x(k)_1\}$ be a subsequence of $\{x(k)\}$, with $\mathcal{A}_0$ containing an approximation to $t(s)$ for all $s$. Then by (14)

$$\forall k \text{ sufficiently large, } \psi(x(k)_1, \mathcal{A}_0; s) > \psi(x(k), \mathcal{A}_0; s_0).$$
Hence \( \{s_{ik}^{(k)}\} \) does not converge to \( s \), and so \( s_{ik}^{(\infty)} = s_{\infty} \). Clearly \( \psi_{\ast}^{(\infty)}(s_{ik}^{(\infty)}) \) is a cluster point of the sequence \( \{\psi_{\ast}^{(k)}(s_{ik}^{(k)})\} \). By replacing \( \{s_{ik}^{(k)}\} \) with a subsequence of itself if necessary, let \( \{\psi_{\ast}^{(k)}(s_{ik}^{(k)})\} \) converge to \( \psi_{\ast}^{(\infty)}(s_{\infty}) \).

Now \( \alpha_k \) is chosen as the first member of the sequence \( 1, \beta, \beta^2, \ldots \) which satisfies the sufficient descent condition
\[
\phi_{\ast}^{(k)} - \phi\left(x_{ik}^{(k)} + q_{ik}^{(k)}(\alpha)\right) \geq \rho \alpha \left[\psi_{\ast}^{(k)}(0) - \psi_{\ast}^{(k)}(s_{ik}^{(k)})\right].
\]
As \( \{x_{ik}^{(k)}\} \), and \( \{H_{ik}^{(k)}\} \) converge to \( x_{\ast}^{(\infty)} \), and \( H_{\ast}^{(\infty)} \), by assumption 4.1, and by the definition of \( \mathcal{A}_{\ast}^{(\infty)} \), it follows that \( \{\psi_{\ast}^{(k)}(0)\} \) converges to \( \psi_{\ast}^{(\infty)}(0) \). Whence, denoting terms which tend to zero as \( k \to \infty \) by \( o(1) \), \( \phi_{\ast}^{(k)} - \phi\left(x_{ik}^{(k)} + q_{ik}^{(k)}(\alpha)\right) \geq \rho \alpha \left[\psi_{\ast}^{(k)}(0) - \psi_{\ast}^{(k)}(s_{ik}^{(k)})\right] + o(1), \) is equivalent to
\[
\phi_{\ast}^{(\infty)} - \phi\left(x_{ik}^{(\infty)} + \alpha s_{ik}^{(\infty)}\right) + o(\alpha) \geq \rho \alpha \left[\psi_{\ast}^{(\infty)}(0) - \psi_{\ast}^{(\infty)}(s_{ik}^{(\infty)})\right] + o(1),
\]
where the \( c_{ik}^{(\infty)} \) part of \( q_{ik}^{(\infty)} \) gives rise to the \( o(\alpha) \) term. Assumption 4.1, and the definition of \( \mathcal{A}_{\ast}^{(\infty)} \), imply \( \Gamma(x_{\ast}^{(\infty)}) \subseteq \mathcal{A}_{\ast}^{(\infty)} \), and so applying equation (7) to the left hand side of (17) implies if \( \alpha_k \) satisfies
\[
\psi_{\ast}^{(\infty)}(0) - \psi_{\ast}^{(\infty)}(\alpha s_{ik}^{(\infty)}) \geq \rho \alpha \left[\psi_{\ast}^{(\infty)}(0) - \psi_{\ast}^{(\infty)}(s_{ik}^{(\infty)})\right] + o(\alpha) + o(1),
\]
then it also satisfies (15). Whence, by the convexity of \( \psi \), \( \{\alpha_k\} \) has a strictly positive lower bound \( (\alpha_L \text{ say}) \).

If \( x_{\ast}^{(\infty)} \) is not a critical point, then for some \( u \in \mathbb{R}^n \), the directional derivative of \( \phi \) at \( x_{\ast}^{(\infty)} \) in the direction \( u \) is strictly negative. This, and equation (7) imply
\[
\psi_{\ast}^{(\infty)}(s_{\ast}^{(\infty)}) - \psi_{\ast}^{(\infty)}(0) = \kappa_\ast < 0.
\]
Once again from equation (16), \( \phi_{\ast}^{(k)} \to \phi_{\ast}^{(\infty)} \) implies
\[
\phi\left(x_{ik}^{(k)} + q_{ik}^{(k)}(\alpha_k^{(k)})\right) \leq \phi_{\ast}^{(\infty)} + \rho \alpha_k^{(k)} \kappa_\ast + o(1).
\]
Thus as \( \alpha_k \geq \alpha_L \) for all \( k \), and as \( \kappa_\ast < 0 \), the existence of an iterate \( x_{ik}^{(k)} \) satisfying equation (13) is clear. \( \Box \)

The implications of only approximations to the global maximizers being available is now discussed. This has two aspects: ensuring the sequence of iterates converges only to critical points, and ensuring the arc search still guarantees sufficient descent whilst remaining a finite process. For the former case it suffices that the maximum of the distances between each member of \( \mathcal{A}^{(k)} \), and the corresponding point it approximates tends to zero as \( k \to \infty \). Before discussing the latter case the following result on the \( L_{\infty} \) QP's solution is given.
**Proposition 5.3** The solution to the L_0 QP

\[
\min_{s \in \mathbb{R}^n} h^T s + \frac{1}{2} s^T H s + \mu \|[B s + b]_+\|_\infty + \frac{1}{2} \nu \|[B s + b]\|_\infty^2
\]

where \(H\) satisfies (10), is continuous with respect to \(B\), and \(b\).

**Proof.** Let \(s_0\) be the solution of the unperturbed \(L_\infty\) QP defined by \(B_0\), and \(b_0\), and similarly for \(s_1, B_1, \) and \(b_1\). Call the \(L_\infty\) QP objective functions \(\psi_0\), and \(\psi_1\) respectively. Without loss of generality, let \(s_0 = 0\). Using \(\Delta B = B_1 - B_0\), and etc,

\[
\psi_0(s_1) - \psi_0(s_0) \geq \frac{1}{2} \gamma_1 \|s_1\|_2^2 \geq \frac{1}{2} \gamma_1 \|s_1\|_\infty^2
\]

and

\[
|\Delta \psi(s)| \leq \mu \left[\left\|\Delta B\right\|_\infty \|s\|_\infty + \|\Delta b\|_\infty\right] + \frac{\nu}{2} \left[\left(\|\Delta B\|_\infty \|s\|_\infty + \|\Delta b\|_\infty\right)^2\right], \quad \forall s \in \mathbb{R}^n.
\]

Hence, as \(s_1\) minimizes \(\psi_1\),

\[
\frac{1}{2} \gamma_1 \|s_1\|_\infty^2 \leq \frac{\nu}{2} \left[\left(\|\Delta B\|_\infty \|s_1\|_\infty + \|\Delta b\|_\infty\right)^2 + \|\Delta b\|_\infty^2\right]
\]

thus as \(\|\Delta B\|_\infty, \|\Delta b\|_\infty \to 0\), \(s_1\) must also tend to \(s_0 = 0\), and the result follows.

To show the arc search is still a finite process, the effect of numerical errors on (9) must be considered in detail.

Let \(A^{(k)}\) be the set of approximations located by the global optimization subalgorithm at \(x^{(k)}\), and let the distance between each element of \(A^{(k)}\) and the point it approximates be less than \(\epsilon\). Let \(\mathcal{A}_n\) be the set of points approximated by the elements of \(A^{(k)}\). Now, to approximate \(\theta\),

\[
\forall \tau \in \Gamma(x^{(k)}), \quad g(x^{(k)}, t) = g(x^{(k)}, \tau) + o(||t - \tau||),
\]

and so \(\max_{t \in A^{(k)}} g(x^{(k)}, t) = \theta(x^{(k)}) + o(\epsilon)\).

Similarly, for \(\varphi(s)\),

\[
\max_{t \in A^{(k)}} \left[g(x^{(k)}, t) + (s^{(k)})^T \nabla g(x^{(k)}, t)\right]_+
\]

\[
= \max_{t \in A^{(k)}} \left[g(x^{(k)}, t) + (s^{(k)})^T \nabla g(x^{(k)}, t)\right]_+ + \frac{o(\epsilon)}{\epsilon}
\]

by the \(C^1\) continuity of \(g\). Now for each trial value of \(\alpha\), \(A^{(k)}\) is used in approximating all except the \(\phi(x^{(k)} + q^{(k)}(\alpha))\) term of (9). If the local method used in the global optimization subalgorithm is superlinear, and the number of iterations of the local method used to refine the members of \(A^{(k)}\) is proportional to the number
of trial $\alpha$-values considered, then because $\alpha \to 0$ linearly, $\epsilon = o(\alpha)$. As the global optimization problem of approximating $\theta(x^{(k)} + q^{(k)}(\alpha))$ is different for different values of $\alpha$, rather than the above approach, it is easier to note that the exact value of $\phi(x^{(k)} + q^{(k)}(\alpha))$ is an upper bound for its calculated approximation, and thus may be used in place of its calculated value in proving the arc search is a finite process.

Inserting the above error estimates into (9), if $\alpha$ satisfies

$$\phi(x^{(k)}) - \phi(x^{(k)} + q^{(k)}(\alpha)) + o(\epsilon) \geq \rho \alpha \left[ \psi^{(k)}(0) - \psi^{(k)}(s^{(k)}) \right] + \alpha o(\epsilon) \epsilon^{-1}$$  \hspace{1cm} (18)

where all terms are exact, then $\alpha$ will satisfy (9). As $\Gamma(x^{(k)}) \subseteq A_{\phi}^{(k)}$, theorem 3.5 can be invoked, and (18) is equivalent to

$$\psi^{(k)}(0) - \psi^{(k)}(\alpha s^{(k)}) + o(\alpha) \geq \rho \alpha \left[ \psi^{(k)}(0) - \psi^{(k)}(s^{(k)}) \right]$$  \hspace{1cm} (19)

where, because $\phi$ is locally Lipschitz, the correction term for the Maratos effect has been incorporated into the $o(\alpha)$ term. The strict convexity of $\psi$ implies (19) holds for all small positive $\alpha$ provided $s^{(k)}$ is a descent direction for $\phi$ at $x^{(k)}$.

The $L_\infty$ QP solution $s^{(k)}$ is also calculated using inexact information from $A^{(k)}$. If $s^{(k)}$ is recalculated every time a fixed number of $\alpha$ values are tested in (9), and if $x^{(k)}$ is non-critical, the sequence of search directions so generated will converge to a direction of strict descent for $\phi$ at $x^{(k)}$, by proposition 5.3. As the set of strict descent directions for $\phi$ at $x^{(k)}$ is open, the calculated values for $s^{(k)}$ will eventually be descent directions for $\phi$ at $x^{(k)}$. If $x^{(k)}$ is critical, the algorithm has found a solution, and terminates at this point.

If the satisfaction of (9) using approximate values is to ensure sufficient descent is obtained, the accuracy to which the global maximizers at $x^{(k)}$, and $x^{(k)} + q^{(k)}(\alpha)$ are found is such that the errors in the calculated values of $\rho \alpha \psi^{(k)}(0)$, $\rho \alpha \psi^{(k)}(s^{(k)})$, and the two left hand side terms in (9) are small compared to the right hand side of (9).

6 Conclusion.

It has been shown that any cluster point of the sequence of iterates generated by the algorithm is a critical point of the penalty function. Each such critical point, if feasible, is also a stationary point of the SIP. Additionally, it can be shown that the set $\mathcal{K}$ of all cluster points of the sequence of iterates is connected. Consequently, if one point in $\mathcal{K}$ is feasible, then $\mathcal{K}$ is a subset of the feasible region, and the objective function is constant on $\mathcal{K}$.

In contrast to algorithms based on the exact $L_1$ penalty function [4,9], this algorithm does not depend upon the applicability of an implicit function theorem; assumption 3.1 is sufficient. Consequently, the minimum necessary degree of continuity of the semi-infinite constraint function is reduced from $C^2$ to $C^1$; this widens the class of problems which may be solved by this type of algorithm.
ACKNOWLEDGEMENT.

The authors wish to thank Professor M. J. D. Powell for many useful comments on an earlier version of this paper, and for furnishing a much simpler proof of proposition 3.4, which appears herein.

References


DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND
Appendix A.

Theorem A-1.

Let \( x^* \) be an optimal point of the SIP (1) at which the regularity assumption (inequality (2)) holds. Let \( t_i^* \) be the global maximizers of \( g(x^*, t) \) satisfying \( g(x^*, t_i^*) = 0 \), and let \( \lambda^* \) be the vector of Lagrange multipliers \( \lambda_i^* \) as specified in (3,4). If \( \mu \) satisfies \( \mu \geq \|\lambda^*\|_1 \), then \( x^* \) is a (feasible) critical point of \( \phi(\mu, \nu; x) \).

Conversely, if \( x^* \) is both feasible and a critical point of \( \phi(\mu, \nu; x) \) for some \( \mu > 0 \), and \( \nu \geq 0 \), then \( x^* \) is a solution point of the SIP.

Proof: Assume \( x^* \) is an optimal point of the SIP. It is shown that \( x^* \) is a critical point of \( \phi \) by showing the directional derivative of \( \phi \) along an arbitrary unit vector \( u \in \mathbb{R}^n \) is non-negative, with the penalty parameters as above. Let \( \Delta x = \alpha u \) be a perturbation of \( x^* \), where \( \alpha > 0 \). Define \( T^* = \{ t \in T : g(x^*, t) = 0 \} \), and let \( w(t) = u^T \nabla_x g(x^*, t) \) for \( t \in T^* \), and \( w_+ = \max_{i=1,...,m} [w_i^*]_+ \), where \( w_i^* = w(t_i^*) \). The optimality conditions, and the definition of \( \theta \) give

\[
\Delta f + \alpha \sum_{i=1}^m w_i^* \lambda_i^* = o(\alpha), \quad \text{and} \\
\Delta \theta = \alpha \max_{t \in T^*} [w(t)]_+ + o(\alpha) \geq \alpha w_+ + o(\alpha).
\]

Whence, from the definition of \( \phi \)

\[
\Delta \phi \geq -\alpha \sum_{i=1}^m w_i^* \lambda_i^* + \mu w_+ + \frac{1}{2} \nu \alpha^2 w_+^2 + o(\alpha)
\]

which, as \( \alpha \) tends to zero gives

\[
- \sum_{i=1}^m w_i^* \lambda_i^* + \mu w_+ \geq 0 \quad \Rightarrow \quad \lim_{\alpha \to 0^+} \frac{\Delta \phi}{\alpha} \geq 0.
\]

For any fixed value of \( w_+ \) the left hand side of the left hand inequality is minimized by \( w_1^* = w_2^* = ... = w_m^* = w_+ \), as \( \lambda_i^* \geq 0 \) for all \( i \). Whence, if \( \mu \geq \sum \lambda_i^* = \|\lambda^*\|_1 \), and \( x^* \) is a stationary point of the SIP, then \( x^* \) is also a critical point of \( \phi \).
Secondly, if $x^*$ is both feasible, and a critical point of $\phi(\mu, \nu; x)$ for some values of the penalty parameters, then

$$\forall x \text{ near } x^*, \quad \phi(\mu, \nu; x) \geq \phi(\mu, \nu; x^*) + o(\|x - x^*\|).$$

As $\phi = f$ on the feasible region, $x^*$ is a solution point of the SIP.

This agrees with existing results concerning penalty functions of the above form. □

**Corollary:** If $\mu > \|\lambda^*\|$ then:

1. If $x^*$ is an optimal point of the SIP, then there is a neighbourhood $\mathcal{N}_1$ of $x^*$ such that $\phi(x) > \phi(x^*)$ for all infeasible points $x$ in $\mathcal{N}_1$.
2. A feasible point $x^*$ is a local minimizer of the SIP iff it is a local minimizer of $\phi$.

**Proof:** If $x$ is in the interior of the feasible region there is nothing to prove for item 1. Otherwise there is at least one active constraint. Using $\alpha$ and $u$ as before, the condition

$$\lim_{\alpha \to 0^+} \frac{\Delta \phi}{\alpha} > 0 \quad \forall \text{ infeasible directions } u : \|u\| = 1,$$

is sufficient to ensure item 1 holds. The same line of reasoning used in theorem A-1 gives

$$-\sum_{i=1}^{m} w_i^* \lambda_i^* + \mu w_+ > 0 \quad \Rightarrow \quad \lim_{\alpha \to 0^+} \frac{\Delta \phi}{\alpha} > 0.$$

Whence item 1 follows.

As $x^*$ is a local minimizer of the SIP, there is a neighbourhood $\mathcal{N}_2$ of $x^*$ satisfying $f(x) \geq f(x^*)$ for all feasible $x$ in $\mathcal{N}_2$. As $\phi(x) = f(x)$ for all feasible $x$, and by part 1,

$$\phi(x) \geq \phi(x^*) \quad \forall x \in \mathcal{N}_1 \cap \mathcal{N}_2,$$

and so $x^*$ is a local minimizer of $\phi$. The converse follows immediately on noting again that $\phi(x) = f(x)$ for all feasible $x$ near $x^*$. □

**Theorem A-2.**

Let $\mathcal{K}$ denote the set of limit points of the iterates generated by the algorithm. Then:

1. The penalty function $\phi$ is constant on the set $\mathcal{K}$.
2. The set $\mathcal{K}$ is connected.
3. If there is a feasible point in $\mathcal{K}$ then $\mathcal{K}$ is a subset of the feasible region.
Proof: The sequence \( \{ \phi(k) \} = \{ \phi(x^{(k)}) \} \) is monotonically decreasing. As there exists a subsequence \( \{ x^{(k)}_* \} \) which converges to \( x^{(\infty)}_* \), the sequence \( \{ \phi(k) \} \) is bounded below, and thus \( \{ \phi(k) \} \) converges to a unique value \( \phi(x^{(\infty)}_*) \). Whence \( \phi \) must be constant on \( \mathcal{K} \).

To show that statements 2 and 3 are true it is first necessary to show that the step length tends to zero as \( k \to \infty \). Consider an iterate \( x^{(k)} \) with the actual step \( q^{(k)}(\alpha) \), which satisfies the sufficient descent condition (9):

\[
\phi(x^{(k)}) - \phi(x^{(k)} + q^{(k)}(\alpha)) \geq \rho \alpha \left[ \psi(x^{(k)}; 0) - \psi(x^{(k)}; s^{(k)}) \right].
\]

By the definition of \( \psi \), and requirement (10),

\[
(\Delta \psi)_k = \psi(x^{(k)}; 0) - \psi(x^{(k)}; s^{(k)}) \geq \gamma_1 \| s^{(k)} \|^2.
\]

As \( 0 \leq \alpha \leq 1 \),

\[
(\Delta \phi)_k = \phi(x^{(k)}) - \phi(x^{(k)} + q^{(k)}(\alpha)) \geq \rho \alpha^2 (\Delta \psi)_k,
\]

which gives

\[
(\Delta \phi)_k \geq \rho \gamma_1 \alpha^2 \| s^{(k)} \|^2.
\]

Now

\[
\| s^{(k)} \| \leq \| s^{(k)} \| \quad \Rightarrow \quad \| q^{(k)}(\alpha) \| \leq 2\alpha \| s^{(k)} \|
\]

so

\[
\| q^{(k)}(\alpha) \| \leq 2 \sqrt{\frac{(\Delta \phi)_k}{\rho \gamma_1}}
\]

and because \( \{ \phi(k) \} \) converges, \( (\Delta \phi)_k \to 0 \) and thus the step length also tends to zero.

Assume \( \mathcal{K} \) is not a connected set. The set of iterates is bounded, therefore \( \mathcal{K} \) is compact. Let \( \mathcal{K}_1 \) and \( \mathcal{K}_2 \) be two disjoint non-empty subsets of \( \mathcal{K} \), with \( \mathcal{K} \) the union of \( \mathcal{K}_1 \) and \( \mathcal{K}_2 \), and

\[
\inf_{c_1, c_2} \| c_2 - c_1 \| = \xi > 0, \quad c_1 \in \mathcal{K}_1, \; c_2 \in \mathcal{K}_2.
\]

Because

\[
\| q^{(k)}(\alpha) \| \to 0, \quad \exists N_1, \text{ such that } \forall k > N_1, \quad \| q^{(k)}(\alpha) \| < \frac{\xi}{4}.
\]

Also, as all limit points of the set of iterates are in \( \mathcal{K} \)

\[
\exists N_2, \; \forall k > N_2, \quad \inf_{c} \| c - x^{(k)} \| < \frac{\xi}{4}, \quad c \in \mathcal{K}.
\]

Whence either \( \mathcal{K}_1 \) or \( \mathcal{K}_2 \) is void — a contradiction. So the set \( \mathcal{K} \) is connected.
Let $x_0$ be a feasible point in $\mathcal{K}$, and let $\lambda_0^*$ be the associated Lagrange multiplier vector. By theorem 5.2 $x_0$ is a critical point of $\phi$. As $\mu > 0$, by the method used to update $\mu$, and by the lower semi-continuity of the Lagrange multiplier estimates, in the limit $k \to \infty$, the inequality $\mu > \| \lambda_0^* \|$ clearly holds. Accordingly the first part of the corollary to theorem A-1 may be invoked. Thus all infeasible points in some neighborhood of $x_0$ have strictly greater values for $\phi$ than $\phi(x_0)$. Whence by the connectedness of $\mathcal{K}$, if there exists a feasible point in $\mathcal{K}$ then all of $\mathcal{K}$ is feasible. By the compactness of $\mathcal{K}$, if $\mathcal{K}$ is not a subset of the feasible region then the minimum distance between $\mathcal{K}$ and the feasible region is strictly positive. □

**Corollary:** If the set of points to which the algorithm converges is feasible, the penalty parameters satisfy the conditions $\mu > \| \lambda^* \|$ and $\nu \geq 0$, and the stationary points of the SIP are isolated, then the sequence of iterates converges to a single point, and that convergence point is a stationary point of the SIP (1).

**Proof:** By theorems 5.2, A-1, and A-2. □
Appendix B.

One purpose of this report is to support the paper entitled "The $L_\infty$ exact penalty function in semi-infinite programming" appearing in the proceedings of the Conference on Theoretical and Applied Computation, held at Griffith University, Brisbane, Queensland, Australia, in 1989. Accordingly this is a convenient place to list the following three corrections which should be made to that paper.

1. Item 3 of theorem 11 should read "The sequence of matrices $H^{(k)}$ has positive upper and lower bounds in the 2-norm."

2. The term "Gateaux derivative" should be replaced by "directional derivative" wherever it appears.

3. The term NLPP is a misnomer for the intermediate problem stated in theorem 6. It is possible that this intermediate problem may have an infinite number of constraints. This does not affect any of the results stemming from theorem 6. In particular, all constraints $c_l(x)$ arising from the same local maximizer of $g(x, t)$ have the same linearization. Hence the vector $\ell(\delta)$ appearing in equation (3) is still of finite dimension, and each iteration is a finite process.