BIT 30 (1990), 723–734

AN EXACT PENALTY FUNCTION ALGORITHM FOR SEMI-INFINITE PROGRAMMES

C. J. PRICE and I. D. COOPE

Department of Mathematics, University of Canterbury, Christchurch 1, New Zealand.

Abstract.

An algorithm for semi-inifinite programming using sequential quadratic programming techniques together with an L_∞ 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.

AMS classifications: 65K05, 90C30.

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, 10, 11]. The methods given in [4, 10, 11] 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:

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

Received January 1990. Revised April 1990.

The objective function f, mapping R^n into R, and the constraint function g, mapping $R^n \times T$ into 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:

$$\phi(\mu, \nu; x) = f(x) + \mu\theta + \tfrac{1}{2}\nu\theta^2 \quad \text{where} \quad \theta = \max_{t \in T} [g(x, t)]_+.$$

The penalty parameters μ and ν are restricted to $\mu > 0$, and $\nu \geq 0$. Clearly $\theta(x)$ is the infinity norm of the constraint violations, so ϕ is continuous $\forall\, x \in R^n$. In contrast, the L_1 penalty function of [4, 11] may be discontinuous at infeasible points [10]. For example, if $n = p = 1$, $T = [-10, 10]$, $f \equiv 0$, $\mu = 1$, $\nu = 0$, and

$$g(x, t) = \frac{25x^2 - 1}{25x^2 + 1} + t^2(\tfrac{1}{4} - x^2) - 8t^4,$$

then the L_1 and L_∞ penalty functions are as in figures (1a) and (1b) respectively. The potential for discontinuities in the L_1 penalty function to prevent convergence to a solution of the SIP is obvious.

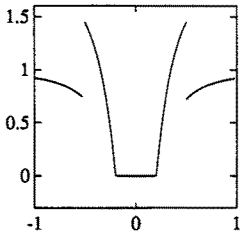

Fig. 1a. The L_1 penalty function.

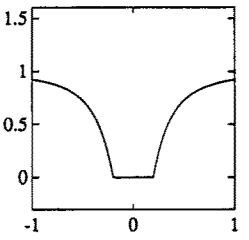

Fig. 1b. The L_∞ penalty function.

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^*:*

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

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

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

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

PROOF. By lemmas 2 and 3 of [6]. ■

DEFINITION 2.2 *For fixed values μ_0 and v_0 of μ and v, a point x_0 is a critical point of $\phi(\mu_0, v_0; x)$ iff at x_0 the directional derivative of $\phi(\mu_0, v_0; x)$ with respect to x along every direction is non-negative.*

The solution set of the PFP for $\mu = \mu_0$, and $v = v_0$ is defined as the set of critical points of $\phi(\mu_0, v_0; x)$.

If solving the PFP is to yield a solution of the SIP, it is highly desirable that the PFP's solution set be contained in (and ideally be equal to) the SIP's solution set. This can be achieved to a limited extent by a suitable choice of μ, for any v.

THEOREM 2.3 *Let x^* be an optimal point of the SIP (1) at which the regularity assumption (inequality (2)) holds, and let λ^* be the vector of Lagrange multipliers as specified in (3,4). If μ satisfies*

$$\mu > \|\lambda^*\|_1 \tag{5}$$

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

Conversely, if x^* is both feasible, and a critical point of $\phi(\mu, v; x)$ for some $\mu > 0$, and $v \geq 0$, 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 ϕ for some μ, and v, then

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

Now $\phi \equiv f$ on the SIP's feasible region, and so x^* is a solution of the SIP. ■

This theorem implies the set of feasible critical points of the PFP is 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.

First, 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.

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

3 Existence of an approximating L_∞ 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 ϕ, and hence an L_∞ Quadratic Programme locally approximating the PFP. The solution of this L_∞ 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.*

This, together with the other usual assumptions, ensures 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 for each 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 is used.

The existence of an approximating L_∞ 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 maximizers of $g(x_c,t)$ on C. If $\Omega(x_c)$ is a subset of the interior of C relative to T (hereafter* int (C)*), then first*

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

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

and second, 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 R^p. Let g_c be the global maximal value of $g(x_c, t)$ on C. For all small positive ε, as $C - \mathcal{N}_\varepsilon(\Omega(x_c))$ is compact and non-empty, $g(x_c, t)$ achieves its supremum on $C - \mathcal{N}_\varepsilon(\Omega(x_c))$, which must be strictly less than g_c. Define

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

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

$$\forall \varepsilon > 0, \ \exists \eta(\varepsilon) > 0, \text{ such that } \forall x, \text{ and } \forall t \in T,$$

$$\|x - x_c\| < \eta(\varepsilon) \Rightarrow |g(x_c, t) - g(x, t)| < \tfrac{1}{4} m(\varepsilon).$$

Hence, for all these values of x,

$$\forall t \in C - \mathcal{N}_\varepsilon(\Omega(x_c)), \quad g(x, t) < g_c - \tfrac{3}{4} m(\varepsilon),$$

$$\text{and } \forall t \in \Omega(x_c), \quad g(x, t) > g_c - \tfrac{1}{4} m(\varepsilon).$$

Hence, $\Omega(x) \subset \mathcal{N}_\varepsilon(\Omega(x_c))$. Moreover, as g is continuous, and C compact, $\Omega(x_c)$ is also compact. Whence, for all small positive ε, $\mathcal{N}_\varepsilon(\Omega(x_c)) \subset \text{int}(C)$, and so $\Omega(x)$ is a subset of the local maximizers of $g(x, t)$ over T. ■

For any $x_0 \in R^n$, $\Gamma(x_0) = \{\tau_1, \ldots, \tau_j\}$ 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

$$\varepsilon_0 = \tfrac{1}{4} \min \{\|\tau_i - \tau_k\| : i, k \in 1, \ldots, j, \ i \neq k\},$$

$$\text{and let } B_i(\varepsilon_0) = \{t \in T : \|t - \tau_i\| \leq \varepsilon_0\}, \ \forall i = 1, \ldots, j.$$

The set of global maximizers of $g(x, t)$ on the set $B_i(\varepsilon_0)$ is denoted by $\Xi_i(x)$. The behaviour of Γ with respect to changes in x is examined by considering each Ξ_i along each ray of the form $x(\sigma) = x_0 + \sigma u$, where $\sigma \geq 0$, and u is a unit vector in 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. $\exists \sigma_{max} > 0$ such that $t(\sigma) \in \Xi_i(x(\sigma)), \ \forall \sigma \in [0, \sigma_{max}]$.

From proposition 3.2, each τ_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 τ_i take the global maximal value of $g(x(\sigma), t)$ over $B_i(\varepsilon_0)$, for all σ. For $i = 1, \ldots, j$ let $t_i(\sigma)$ be an extension of τ_i. Each member of the set $\{g(x(\sigma), t_i(\sigma))\}_{i=1}^{j}$ is locally Lipschitz with respect to σ 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.

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

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

Also,

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

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

Now, as $\Xi(x_0)$ is a singleton set, proposition 3.2 implies every extension of τ_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 ψ to be a continuous piecewise quadratic approximation to ϕ near x_0, where ψ is based on the finite subset $\mathscr{A}_0$ of T, as follows

$$\psi(x_0, \mathscr{A}_0; \mu, \nu; s) = f(x_0) + s^T \nabla f(x_0) + \tfrac{1}{2} s^T H s + \mu \vartheta(s) + \tfrac{1}{2} \nu \vartheta^2(s),$$

$$\text{where } \vartheta(s) = \max_{t \in \mathscr{A}_0} [g(x_0, t) + s^T \nabla_x g(x_0, t)]_+,$$

and where H is positive definite. Clearly ψ is strictly convex in s.

Let the base set $\mathscr{A}_0$ 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 $\Gamma(x_0) \subseteq \mathscr{A}_0$ then, for all $s \in R^n$ such that $\|s\|$ is small,*

$$\phi(\mu, \nu; x_0 + s) = \psi(x_0, \mathscr{A}_0; \mu, \nu; s) + o(\|s\|). \tag{7}$$

PROOF. For all s sufficiently small, each element of $Bs + b$ arising from some member of $\mathscr{A}_0 - \Gamma(x_0)$ is less than every element of $Bs + b$ arising from some member of $\Gamma(x_0)$; thus $\mathscr{A}_0 - \Gamma(x_0)$ 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

$$\forall \tau \in \Gamma(x_0 + s),\ \exists t_0 \in \Gamma(x_0) \text{ such that}$$

$$g(x_0 + s, \tau) = g(x_0, t_0) + s^T \nabla_x g(x_0, t_0) + o(\|s\|). \bullet$$

Hence

$$\theta(x_0 + s) = \max_{t \in \mathscr{A}_0} [g(x_0, t_0) + s^T \nabla_x g(x_0, t_0)]_+ + o(\|s\|).$$

Using a linear approximation to the objective function, the result follows. ■

The convergence proof requires that $\|s\|_\infty$ be subject to an upper bound, specifically $S_b \gg 0$. The $L_\infty QP$

$$\min_{s \in R^n} \psi(x_0, \mathscr{A}_0; \mu, v; s) \text{ subject to } \|s\|_\infty \leq S_b, \tag{8}$$

approximates the PFP near x_0. If $\Gamma(x_0) \subseteq \mathscr{A}_0$, then x_0 is a critical point of ϕ iff $s = 0$ is the global minimizer of $\psi(x_0, \mathscr{A}_0; \mu, v; s)$.

4 An L_∞-norm algorithm for SIP.

The previous section examined the L_∞ 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_∞ QP is constructed. The solution $s^{(k)}$ to the L_∞ 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 α is chosen to be the first member of the sequence $1, \beta, \beta^2, \ldots$ to satisfy the sufficient descent criterion

$$\phi(x^{(k)}) - \phi(x^{(k)} + q^{(k)}(\alpha)) \geq \rho\alpha[\psi(x^{(k)}; 0) - \psi(x^{(k)}; s^{(k)})], \tag{9}$$

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 λ^*_{est} of the optimal Lagrange multipliers at

each iterate and adjusting the penalty parameters accordingly. Such estimates may be calculated from the L_∞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 $\mathscr{A}^{(k)}$.
2. The approximating L_∞QP is formed, and its solution $s^{(k)}$ is calculated. If necessary the penalty parameters are increased to ensure $\vartheta(s^{(k)}) \leq \vartheta(0)$.
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 θ is less than some positive parameter π_1, and if $\mu \leq \pi_2 \|\lambda^*_{est}\|_1$, then μ 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 ν is adjusted to give $\mu + \nu\theta = \pi_5 \|\lambda^*_{est}\|_1$, where $\pi_5 > \pi_4 > 1$.
5. Update H using a quasi-Newton scheme whilst ensuring,

$$\exists \gamma > 0, \text{ such that } \forall x \in R^n - \{0\}, \quad \forall k, 0 < x^T H^{(k)} x \leq \gamma x^T x; \tag{10}$$

   for example Powell's modified BFGS update [8] could be used.
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 $\mathscr{A}^{(k)}_M$. Let $\mathscr{Q}^{(k)}$ denote the set of elements $t \in \mathscr{A}^{(k)}$ satisfying

$$\vartheta(s^{(k)}) = g(x^{(k)}, t) + (s^{(k)})^T \nabla_x g(x^{(k)}, t).$$

Define $t_M(w)$ to be the closest member of $\mathscr{A}^{(k)}_M$ to w, for each $w \in \mathscr{Q}^{(k)}$. If $t_M(w)$ is uniquely defined for every w, if t_M is a one to one mapping, and if $\mathscr{Q}^{(k)}$ is non-empty, then $c^{(k)}$ is chosen as the vector of minimum length satisfying

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

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 superlinear convergence on problems with the required continuity. Mayne and Polak [7] show if f and g are C^3, if x^* is a solution of the SIP at which strict complementarity, second order sufficiency conditions, and an implicit function theorem hold, and if the vectors $\{\nabla_x g(x^*, t) : t \in \Gamma(x^*)\}$ are linearly independent, then $x^{(k)} \to x^*$ implies

$x^{(k)} \to x^*$ superlinearly. 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 $c^{(k)} = 0$. However, for problems which are sufficiently continuous, choosing $c^{(k)}$ as above ensures superlinear convergence will be obtained.

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 $x \in R^n$ 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 in [9].

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 α. If $s^{(k)}$ is zero, then $c^{(k)}$ is also zero, and (9) holds for all α. If $s^{(k)}$ is non-zero, then from assumption 4.1 and theorem 3.5, and using $\psi(x^{(k)}, \mathcal{A}^{(k)}; s) = \psi^{(k)}(s)$, (9) is equivalent to

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

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

THEOREM 5.1 *Given:*

1. *All iterates generated by the algorithm lie in a bounded region of R^n.*
2. *Assumptions 3.1, 4.1, and the condition (10) hold.*
3. *The parameters μ and ν 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 μ and ν are the final values of these parameters.

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

$$\phi(x^{(k)} + q^{(k)}(\alpha^{(k)})) < \phi(x_*^{(\infty)}). \tag{13}$$

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

Let $x_*^{(\infty)}$ be an arbitrary cluster point of $\{x^{(k)}\}$. Select a subsequence $\{x_*^{(k)}\}$ of $\{x^{(k)}\}$, generated after μ and ν assume their final values, and where the subsequences $\{x_*^{(k)}\}$, $\{H_*^{(k)}\}$, and $\{s_*^{(k)}\}$ converge to $x_*^{(\infty)}$, $H_*^{(\infty)}$, and $s_*^{(\infty)}$ respectively. Such a subsequence exists by item 1, requirement (10) and the bound on s in (8).

Let $\psi_*^{(k)}(s)$ and $\phi_*^{(k)}$ denote $\psi(x_*^{(k)}, \mathscr{A}(x_*^{(k)}); s)$ and $\phi(x_*^{(k)})$ respectively. Also let $\mathscr{A}_*^{(k)}$ denote $\mathscr{A}(x_*^{(k)})$. Define $\mathscr{A}_*^{(\infty)}$ as the set of all cluster points of sequences of the form $\{\xi_i\}_{i=1}^{\infty}$, where $\xi_i \in \mathscr{A}_*^{(i)}$, for all i. Clearly $\mathscr{A}_*^{(\infty)}$ is compact. Also

$$\forall \varepsilon > 0,\ \exists K,\ \forall k > K,\ \max_{t \in \mathscr{A}_*^{(k)}} \left[\min_{\tau \in \mathscr{A}_*^{(\infty)}} \|t - \tau\| \right] < \varepsilon, \tag{14}$$

by the definition of $\mathscr{A}_*^{(\infty)}$. Let $\mathscr{S}_{GM}$ be the set of global minimizers of $\psi(x_*^{(\infty)}, \mathscr{A}_*^{(\infty)}; s)$, and let s_∞ be any element of $\mathscr{S}_{GM}$. Then

$$\forall s \notin \mathscr{S}_{GM},\ \exists t(s) \in \mathscr{A}_*^{(\infty)}, \text{ such that } \psi(x_*^{(\infty)}, \{t(s)\}; s) > \psi_*^{(\infty)}(s_\infty).$$

Fixing s, let $\{x_\#^{(k)}\}$ be a subsequence of $\{x_*^{(k)}\}$, with $\mathscr{A}_\#^{(k)}$ containing an approximation $t_\#^{(k)}$ to $t(s)$ such that $t_\#^{(k)} \to t(s)$ as $k \to \infty$. Then by (14)

$$\forall k \text{ sufficiently large, } \psi(x_\#^{(k)}, \mathscr{A}_\#^{(k)}; s) > \psi(x_\#^{(k)}, \mathscr{A}_\#^{(k)}; s_\infty).$$

Hence $\{s_\#^{(k)}\}$ does not converge to s, and so $s_*^{(\infty)} \in \mathscr{S}_{GM}$. Clearly $\psi_*^{(\infty)}(s_*^{(\infty)})$ is a cluster point of the sequence $\{\psi_*^{(k)}(s_*^{(k)})\}$. By replacing $\{x_*^{(k)}\}$ with a subsequence of itself if necessary, let $\{\psi_*^{(k)}(s_*^{(k)})\}$ converge to $\psi_*^{(\infty)}(s_*^{(\infty)})$.

Now $\alpha_*^{(k)}$ is chosen as the first member of the sequence $1, \beta, \beta^2, \ldots$ which satisfies the sufficient descent condition

$$\phi_*^{(k)} - \phi(x_*^{(k)} + q_*^{(k)}(\alpha)) \geq \rho\alpha[\psi_*^{(k)}(0) - \psi_*^{(k)}(s_*^{(k)})]. \tag{15}$$

As $\{x_*^{(k)}\}$ converges to $x_*^{(\infty)}$, by assumption 4.1, and by the definition of $\mathscr{A}_*^{(\infty)}$, it follows that $\{\psi_*^{(k)}(0)\}$ converges to $\psi_*^{(\infty)}(0)$. Whence, denoting terms which tend to zero as $k \to \infty$ by $o(1)$, (15) is equivalent to

$$\phi_*^{(k)} - \phi(x_*^{(k)} + q_*^{(k)}(\alpha)) \geq \rho\alpha[\psi_*^{(\infty)}(0) - \psi_*^{(\infty)}(s_*^{(\infty)})] + o(1), \tag{16}$$

by the convergence of $\{\psi_*^{(k)}(s_*^{(\infty)})\}$, and $\{\psi_*^{(k)}(0)\}$ to $\psi_*^{(\infty)}(s_*^{(\infty)})$, and $\psi_*^{(\infty)}(0)$. Now, as $q_*^{(k)}(\alpha) = \alpha s_*^{(k)} + \alpha^2 c_*^{(k)}$, as $\|c_*^{(k)}\| \leq \|s_*^{(k)}\|$, as ϕ is locally Lipschitz, and by the convergence of $\{x_*^{(k)}\}$, (16) is equivalent to

$$\phi_*^{(\infty)} - \phi(x_*^{(\infty)} + \alpha s_*^{(\infty)}) + o(\alpha) \geq \rho\alpha[\psi_*^{(\infty)}(0) - \psi_*^{(\infty)}(s_*^{(\infty)})] + o(1), \tag{17}$$

where the $c_*^{(\infty)}$ part of $q_*^{(\infty)}$ gives rise to the $o(\alpha)$ term. Assumption 4.1, and the definition of $\mathscr{A}_*^{(\infty)}$, imply $\Gamma(x_*^{(\infty)}) \subseteq \mathscr{A}_*^{(\infty)}$, and so applying equation (7) to the left hand side of (17) yields the following statement: If $\alpha_*^{(k)}$ satisfies

$$\psi_*^{(\infty)}(0) - \psi_*^{(\infty)}(\alpha s_*^{(\infty)}) \geq \rho\alpha[\psi_*^{(\infty)}(0) - \psi_*^{(\infty)}(s_*^{(\infty)})] + o(\alpha) + o(1), \tag{18}$$

then it also satisfies (15). If $x_*^{(\infty)}$ is not a critical point, then for some u in R^n, the directional derivative of ϕ at $x_*^{(\infty)}$ in the direction u is strictly negative. This, and equation (7) imply

$$\psi_*^{(\infty)}(s_*^{(\infty)}) - \psi_*^{(\infty)}(0) = \kappa_* < 0.$$

Whence, by the convexity of ψ, and from (18) $\{\alpha_*^{(k)}\}$ has a strictly positive lower bound (α_L say). Once again from equation (16), $\phi_*^{(k)} \to \phi_*^{(\infty)}$ implies

$$\phi(x_*^{(k)} + q_*^{(k)}(\alpha_*^{(k)})) \leq \phi_*^{(\infty)} + \rho\alpha_*^{(k)}\kappa_* + o(1).$$

Thus as $\alpha_*^{(k)} \geq \alpha_L$ for all k, and as $\kappa_* < 0$, the existence of an iterate $x_*^{(k)}$ satisfying equation (13) is clear. ■

6 Conclusion.

Under fairly mild assumptions convergence to a set of critical points of the PFP has been shown. Each such critical point, if feasible, is also a solution of the SIP.

In contrast to algorithms based on the exact L_1 penalty function [4, 11], 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. Superlinear convergence is obtainable on problems with the requisite degree of continuity.

The L_∞ penalty function, unlike its L_1 counterpart in [4, 11], is continuous at all points in R^n; hence one potential method of failure for algorithms based on the L_1 norm does not occur with the algorithm presented herein.

Numerical results will be presented in [9].

Acknowledgements.

The authors wish to thank Professor M. J. D. Powell for furnishing a much simpler proof of proposition 3.4 (appearing herein), and to thank Professor Powell and an anonymous referee for many helpful comments.

REFERENCES

[1] Bandler, J. W. and C. Charalambous, *Nonlinear programming using minimax techniques*, J. of Opt. Th. and Appl. 13, 607–618 (1974).

[2] Conn, A. R. and N. I. M. Gould, *An exact penalty function for semi-infinite programming*, Math. Prog. 37, 19–40 (1987).

[3] Coope, I. D., *The Maratos effect in sequential quadratic programming algorithms using the L_1 exact penalty function*, Technical Report CS85-32, Computer Science Department, University of Waterloo, August 1985.

[4] Coope, I. D. and G. A. Watson, *A projected Lagrangian algorithm for semi-infinite programming*, Math. Prog. 32, 337–356 (1985).
[5] Gill, P. E. and W. Murray, *The computation of Lagrange multiplier estimates for constrained optimization*, Math. Prog. 17, 32–60 (1976).
[6] Gustafson, S.-Å., *A three phase algorithm for semi-infinite programmes*, in *Semi-Infinite Programming, and Applications*, eds. A. V. Fiacco, and K. O. Kortanek, Springer-Verlag, Berlin (1983).
[7] Mayne, D. Q. and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Math. Prog. Study 17, 45–61 (1982).
[8] Powell, M. J. D., *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in *Nonlinear Programming 3*, eds. O. L. Mangasasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, (1978).
[9] Price, C. J., PhD thesis, forthcoming.
[10] Tanaka, Y., M. Fukushima, and T. Ibaraki, *A globally convergent SQP method for semi-infinite non-linear optimization*, Journal of Computational and Applied Mathematics 23, 141–153 (1988).
[11] Watson, G. A., *Globally convergent methods for semi-infinite programming*, BIT 21, 362–373 (1981).