

Convergence Properties of an Algorithm for Non-Linear Programmes which uses a Two Parameter Exact Penalty Function

by

C. J. Price

*Department of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand*

No. 97

September, 1993

Abstract – A sequential quadratic programming algorithm for nonlinear programmes is described. The algorithm uses a two parameter ℓ_∞ exact penalty function. Convergence is shown under mild conditions, and superlinear convergence is also obtainable on problems with the requisite properties. Some numerical results are given which show that the algorithm is effective in practice.

Convergence Properties of an Algorithm for Non-Linear Programmes which uses a Two Parameter Exact Penalty Function

C. J. Price,

Department of Mathematics and Statistics,

University of Canterbury, Private Bag 4800,

Christchurch, New Zealand.

Abstract

A sequential quadratic programming algorithm for nonlinear programmes is described. The algorithm uses a two parameter ℓ_∞ exact penalty function. Convergence is shown under mild conditions, and superlinear convergence is also obtainable on problems with the requisite properties. Some numerical results are given which show that the algorithm is effective in practice.

1 Introduction

This report examines the convergence properties of a Sequential Quadratic Programming (SQP) algorithm for Non-Linear Programmes (NLP). The algorithm uses an Exact Penalty Function (EPF) which is a hybrid of the one parameter penalty function of [4, 5, 10] and a quadratic penalty function.

The Non-Linear Programme under discussion is written in the form:

$$\min_{x \in R^n} f(x) \text{ subject to } c(x) \leq 0. \quad (1)$$

The objective function f and constraint vector function c are assumed to be continuously differentiable, where f maps R^n into R and c maps R^n into R^m . Equality

constraints could easily be included in a more general expression of a NLP along with simple bounds and other linear constraints. However, for clarity these have been omitted from (1).

Assumption 1.1 *At each local minimizer of of the NLP (1) an appropriate, but unspecified, constraint qualification is assumed to hold, thereby ensuring that any optimal point x^* of the NLP (1) satisfies the following Karush–Kuhn–Tucker (KKT) conditions: there exists a vector of Lagrange multipliers λ^* in R^m such that*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0 \quad (2)$$

$$\text{and } \lambda_i^* c_i(x^*) = 0, \quad \lambda_i^* \geq 0, \quad c_i(x^*) \leq 0, \quad \forall i = 1, \dots, m. \quad (3)$$

It has not been assumed that the functions defining the NLP are twice continuously differentiable, and so second order optimality conditions are not directly applicable to the NLP. Therefore feasible points at which the KKT conditions (2,3) hold will be regarded as solutions of the NLP.

2 The Penalty Function Problem

The NLP is not solved directly — instead a non-differentiable exact penalty function Φ is minimized, where the exact penalty function is constructed so that local minimizers of the NLP are also local minimizers of the penalty function Φ . The penalty function is

$$\Phi(x) = f(x) + \mu\theta(x) + \frac{1}{2}\nu\theta^2(x), \quad (4)$$

where the degree of infeasibility $\theta(x)$ is defined as

$$\theta(x) = \max_{1 \leq i \leq m} [c_i(x)]_+. \quad (5)$$

The penalty parameters μ and ν are restricted to $\mu > 0$ and $\nu \geq 0$. The penalty function Φ may be viewed as a hybrid of a quadratic penalty function based on the infinity norm and the single parameter exact penalty function of [4, 5, 10].

Clearly θ is continuous $\forall x \in R^n$, but it is usually not differentiable for some x . However, the directional derivative

$$D_p\theta(x) = \lim_{\alpha \rightarrow 0^+} \frac{\theta(x + \alpha p) - \theta(x)}{\alpha}$$

exists for all $x, p \in R^n$. The definition (5) of θ , and the C^1 continuity of each c_i imply that $\forall x, p \in R^n$

$$D_p\theta(x) = \begin{cases} \max_{i \in I(x)} p^T \nabla c_i(x) & \text{if } \theta(x) > 0 \\ \max_{i \in I(x)} [p^T \nabla c_i(x)]_+ & \text{if } \theta(x) = 0 \text{ and } I(x) \neq \emptyset \\ 0 & I(x) = \emptyset \end{cases}$$

where $I(x) = \{i : c_i(x) = \theta(x)\}$.

These properties of $\theta(x)$ imply that Φ is continuous for all x , and the directional derivative $D_p\Phi(x)$ also exists for all $x, p \in R^n$. This allows solutions to the problem of minimizing Φ to be defined as follows.

Definition 2.1 For fixed values of μ and ν , a point \bar{x} is a **critical point** of $\Phi(x)$ iff for all $p \in R^n$ the directional derivative $D_p\Phi(\bar{x})$ is non-negative.

The set of points satisfying definition (2.1) with fixed values for μ and ν is referred to as the set of critical points of $\Phi(x)$.

It is useful to approximate $\Phi(\mu, \nu; x + p)$ by a continuous piecewise quadratic $\Psi(x, \mu, \nu, H; p)$ which is defined as follows:

$$\Psi(x, \mu, \nu, H; p) = f(x) + p^T \nabla f(x) + \frac{1}{2} p^T H p + \mu \xi(p) + \frac{1}{2} \nu \xi^2(p),$$

$$\text{where } \xi(p) = \max_{i=1, \dots, m} [c_i(x) + p^T \nabla c_i(x)]_+,$$

and where H is a positive definite matrix used to include second order information about Φ in Ψ . The order of the approximation is given by

$$\Phi(\mu, \nu; x + p) = \Psi(x, \mu, \nu, H; p) + o(\|p\|) \text{ for } p \text{ small.} \quad (6)$$

The problem of minimising Φ can be studied by considering the locally approximating ℓ_∞ Quadratic Programme (ℓ_∞ QP)

$$\min_{p \in \mathbb{R}^n} \Psi(x, \mu, \nu, H; p). \quad (7)$$

The ℓ_∞ QP (7) can be rewritten as the quadratic programme

$$\min_{p, \zeta} \left\{ p^T \nabla f(x) + \frac{1}{2} p^T H p + \mu \zeta + \frac{1}{2} \nu \zeta^2 \right\} \quad (8)$$

$$\text{subject to } c_i(x) + p^T \nabla c_i(x) \leq \zeta \quad \forall i = 1, \dots, m, \text{ and } \zeta \geq 0. \quad (9)$$

Problems (7) and (8,9) are essentially equivalent, and both will be referred to as the ℓ_∞ QP. At the ℓ_∞ QP's solution p the equation $\xi(p) = \zeta$ holds. This approximating ℓ_∞ QP can be used to establish a link between solutions of the NLP and feasible critical points of Φ as the following theorem shows.

Theorem 2.2 *Let x^* be an optimal solution of the NLP (1) at which conditions (2,3) hold, and let $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$ be a vector of Lagrange multipliers satisfying these conditions for which $\|\lambda^*\|_1$ is minimal. If μ satisfies*

$$\mu > \|\lambda^*\|_1 \quad (10)$$

then x^ is a critical point of $\Phi(x)$.*

Conversely, if x^ is both feasible and a critical point of Φ for some $\mu > 0$ and $\nu \geq 0$, then x^* is a KKT point of the NLP.*

Proof: The KKT conditions (2), (3) for the QP formulation of the ℓ_∞ QP constructed about the point x^* are

$$[Hp + \nabla f + \sum_{i=1}^m \eta_i \nabla c_i]^* = 0 \quad (11)$$

$$[p^T \nabla c_i + c_i - \zeta]^* \leq 0, [\eta_i(p^T \nabla c_i + c_i - \zeta)]^* = 0, \text{ and } \eta_i^* \geq 0, \forall i = 1, \dots, m, \quad (12)$$

$$\zeta^* \geq 0, \zeta^* \lambda_\zeta^* = 0, \lambda_\zeta^* \leq 0, \text{ and } \mu + \nu \zeta^* + \lambda_\zeta^* = \|\eta^*\|_1, \quad (13)$$

where H^* can be selected as any positive semi-definite symmetric matrix. Here λ_ζ^* is the Lagrange multiplier for the constraint $\zeta \geq 0$, and η_i^* is the Lagrange multiplier for the constraint involving c_i in (9).

Now let x^* is an optimal point of the NLP. Equations (11,12) hold with $p^* = \zeta^* = 0$ and $\eta_i^* = \lambda_i^*$ for all $i = 1, \dots, m$ because x^* satisfies (2,3). The last equation in (13) with $\zeta^* = 0$ gives the correct sign for λ_ζ provided $\mu > \|\lambda^*\|_1$. Let $\Psi^*(p)$ denote $\Psi(x^*, \mu^*, \nu^*, H^*; p)$. Now Ψ^* is convex, and so $p^* = 0$ is a global minimiser of $\Psi^*(p)$. Equation (6) implies that x^* is a critical point of Φ given (10) holds.

Conversely, if x^* is a critical point of Φ for some μ and ν , then

$$\forall x \text{ near } x^*, \Phi(x) = \Phi(x^*) + D_{x-x^*}\Phi(x^*) + o(\|x - x^*\|) \geq \Phi(x^*) + o(\|x - x^*\|).$$

Now $\Phi \equiv f$ on the NLP's feasible region, and so the absence of a feasible direction of descent for f at x^* is established.

Given a suitable choice of the penalty parameters, the NLP may be replaced by the problem of finding critical points of Φ which satisfy $c(x) \leq 0$. The penalty function (4) is minimised by an iterative process which alternates between calculating a descent direction for Φ at the current iterate $x^{(k)}$, and calculating the next iterate $x^{(k+1)}$ using an arc search. The descent direction $p^{(k)}$ at the k^{th} iterate $x^{(k)}$ is calculated by approximating Φ by a continuous piecewise quadratic function

$\Psi^{(k)}(p) = \Psi(x^{(k)}, \mu^{(k)}, \nu^{(k)}, H^{(k)}; p)$, and solving the ℓ_∞ QP

$$\min_{p \in R^n} \Psi^{(k)}(p)$$

for $p^{(k)}$. The bound

$$\|p\|_\infty \leq M_{\text{bound}}, \quad \text{where } M_{\text{bound}} \gg 0 \quad (14)$$

is imposed on each ℓ_∞ QP in order to ensure convergence.

Clearly $\Psi^{(k)}$ is strictly convex in p , and the level set $\{p \in R^n : \Psi^{(k)}(p) \leq \Psi^{(k)}(0)\}$ is bounded for all $\mu > 0$ and all $\nu \geq 0$. Thus $\Psi^{(k)}$ has a unique global minimizer $p^{(k)}$.

Theorem 2.3 *If $p^{(k)} \neq 0$ then $p^{(k)}$ is a direction of strict descent for $\Phi(x)$ at $x^{(k)}$, otherwise $x^{(k)}$ is a critical point of Φ .*

Proof: If $p^{(k)} = 0$, then for all non-zero p , $D_p \Psi^{(k)}(0) \geq 0$ because 0 is the unique global minimiser of $\Psi^{(k)}(p)$. Hence, equation (6) implies $D_p \Phi(x^{(k)}) \geq 0$ for all non-zero p ; that is to say $x^{(k)}$ is an critical point of the penalty function.

When $p^{(k)} \neq 0$, equation (6) yields

$$D_{p^{(k)}} \Phi(\mu^{(k)}, \nu^{(k)}; x^{(k)}) = D_{p^{(k)}} \Psi^{(k)}(0).$$

Now $p^{(k)}$ is the strict global minimiser of $\Psi^{(k)}$, so $\Psi^{(k)}(p^{(k)}) < \Psi^{(k)}(0)$. The convexity of $\Psi^{(k)}$ implies $D_{p^{(k)}} \Psi^{(k)}(0) < 0$, yielding the required result. ♣

The following algorithm is one of many that can be constructed using the hybrid penalty function.

Algorithm A.

1. INITIALIZATION.

$$\mu^{(1)} = 1, \quad \nu^{(1)} = 1, \quad k = 1,$$

$$\rho = 0.02, \quad H^{(1)} = I_n, \quad \theta_{\text{crossover}} = 1, \quad \theta_{\text{cap}} = 100,$$

$$\kappa_1 = 1.2, \quad \kappa_2 = 1.5, \quad \kappa_3 = 1.2, \quad \kappa_4 = 4.$$

2. SOLVE THE ℓ_∞ QP. If $\theta^{(k)} > \theta_{\text{cap}}$ then the capping constraint $\zeta \leq \theta^{(k)}$ is also imposed on the ℓ_∞ QP (8,9). This ℓ_∞ QP is then solved. If the capping constraint is not active at the ℓ_∞ QP's solution then the algorithm proceeds directly to step 3. Otherwise the penalty parameters are updated as described in step 7, except that $\|\lambda^{(k)}\|_1$ is replaced by $\mu^{(k)} + \nu^{(k)}\theta^{(k)} + |\lambda_{\text{cap}}|$, where λ_{cap} is the Lagrange multiplier of the capping constraint. The ℓ_∞ QP (8,9) is then re-solved.

3. ATTEMPT THE PROPOSED STEP. If both of the following conditions hold:

$$\text{first } \Phi(x^{(k)}) - \Phi(x^{(k)} + p^{(k)}) \geq \rho\alpha \left[\Psi^{(k)}(0) - \Psi(p^{(k)}) \right],$$

and second, either the penalty parameters were not altered in step 2 or the inequality $\theta(x^{(k)} + p^{(k)}) \leq \theta(x^{(k)})$ is satisfied, then the proposed step $p^{(k)}$ is accepted and the algorithm proceeds to step 6. Otherwise execution continues at the next step.

4. CALCULATE THE SECOND ORDER CORRECTION. Solve the following QP for the second order correction $t^{(k)}$:

$$\min_{t \in \mathbb{R}^n} \|t\|_2^2,$$

$$\text{such that } t^T \nabla c_i(x^{(k)}) + c_i(x^{(k)} + p^{(k)}) \geq 0 \quad \forall i \in T,$$

where T is the set of indices of the constraints active at the QP's solution in step 2. If $\|t^{(k)}\|_2 \geq \|p^{(k)}\|_2$, then set $t^{(k)} = 0$. This vector is essentially that of Mayne and Polak [4].

5. DO THE ARC SEARCH. Consider successive values of the sequence $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$ as trial values of α . If $t^{(k)} = 0$ then omit the first member of the sequence, otherwise start with $\alpha = 1$. Accept the first trial value of α which satisfies the following two conditions: first

$$\Phi(x^{(k)}) - \Phi(x^{(k)} + q^{(k)}(\alpha)) \geq \rho\alpha [\Psi^{(k)}(0) - \Psi(p^{(k)})], \quad (15)$$

and second, if the penalty parameters were altered in step 2, then the step $q^{(k)}(\alpha)$ is also required to satisfy the condition $\theta(x^{(k)} + q^{(k)}(\alpha)) \leq \theta^{(k)}$. After a satisfactory value of α has been found, set $x^{(k+1)} = x^{(k)} + q^{(k)}(\alpha)$.

6. ESTIMATE THE LAGRANGE MULTIPLIERS. Form lower semi-continuous estimates $\lambda^{(k)}$ of the optimal Lagrange multipliers.
7. UPDATE H , AND THE PENALTY PARAMETERS. H is updated by, for example, Powell's modified BFGS formula [8, 9] or some other means [4]. The penalty parameters are updated as follows:

$$\text{if } \theta^{(k)} \leq \theta_{\text{crossover}} \text{ and } \mu^{(k)} < \kappa_1 \|\lambda^{(k)}\|_1$$

$$\text{then set } \mu^{(k+1)} = \kappa_2 \|\lambda^{(k)}\|_1 \text{ and } \nu^{(k+1)} = \nu^{(k)}.$$

$$\text{If } \theta^{(k)} > \theta_{\text{crossover}} \text{ and } \mu^{(k)} + \nu^{(k)}\theta^{(k)} < \kappa_3 \|\lambda^{(k)}\|_1$$

$$\text{then set } \nu^{(k+1)} = \frac{\kappa_4 \|\lambda^{(k)}\|_1 - \mu^{(k)}}{\theta^{(k)}} \text{ and } \mu^{(k+1)} = \mu^{(k)}.$$

Otherwise the penalty parameters are not altered.

8. CHECK THE STOPPING CONDITIONS. If sufficient accuracy has been attained then the algorithm halts, otherwise increment k and go to step 2.

3 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 (15) holds for all small positive α . If $p^{(k)}$ is zero, then $t^{(k)}$ is also zero, and (15) holds for all α . If $p^{(k)}$ is non-zero, then the sufficient descent condition (15) is equivalent to

$$\Psi^{(k)}(0) - \Psi^{(k)}(\alpha p^{(k)} + \alpha^2 t^{(k)}) + o(\alpha) \geq \rho \alpha [\Psi^{(k)}(0) - \Psi^{(k)}(p^{(k)})].$$

Now, because Ψ is locally Lipschitz, the $t^{(k)}$ term can be included in the $o(\alpha)$ term to yield:

$$\Psi^{(k)}(0) - \Psi^{(k)}(\alpha p^{(k)}) + o(\alpha) \geq \rho \alpha [\Psi^{(k)}(0) - \Psi^{(k)}(p^{(k)})]. \quad (16)$$

The strict convexity of Ψ ensures (16) holds for all small positive α .

Theorem 3.1 *Given:*

1. *The sequence of iterates $\{x^{(k)}\}$ generated by the algorithm is bounded.*
2. *The sequence of matrices $\{H^{(k)}\}$ is bounded in norm.*
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 at their final values.

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

$$\Phi(x^{(k)} + \alpha^{(k)} p^{(k)}) < \Phi(x_*^{(\infty)}). \quad (17)$$

As the sequence $\{\Phi^{(k)}\}$ is monotonically decreasing, and Φ is continuous, a contradiction results. The existence of an iterate satisfying (17) is shown by using the following (loosely outlined) argument. If $x_*^{(\infty)}$ is not a critical point then any solution $p_*^{(\infty)}$ of some approximating ℓ_∞ QP at $x_*^{(\infty)}$ will be non-zero, and thus will be a direction of strict descent for Φ at $x_*^{(\infty)}$. It is shown that the sequence of prospective steps $\{p^{(k)}\}$ converges to $p_*^{(\infty)}$, and the sequence $\{\alpha^{(k)}\}$ is bounded away from zero. Using continuity arguments (17) is then established.

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 $\{p_*^{(k)}\}$ converge to the unique limits $x_*^{(\infty)}$, $H_*^{(\infty)}$, and $p_*^{(\infty)}$ respectively. Such a subsequence exists by items 1 and 2, and the bound on p in (14).

First it is shown that $p_*^{(\infty)}$ is a solution of a locally approximating ℓ_∞ QP at the point $x_*^{(\infty)}$. Let $\Phi_*^{(k)}$ denote $\Phi(x_*^{(k)})$. The convergence of the sequences $\{x_*^{(k)}\}$ and $\{H_*^{(k)}\}$ implies each term in

$$\begin{aligned} \Psi_*^{(k)}(p) &= p^T \left[\nabla f_*^{(k)} + \frac{1}{2} H_*^{(k)} p \right] + \mu \max_{i \in 1, \dots, m} \left[c_i(x_*^{(k)}) + p \nabla c_i(x_*^{(k)}) \right]_+ \\ &\quad + \frac{1}{2} \nu \left(\max_{i \in 1, \dots, m} \left[c_i(x_*^{(k)}) + p^T \nabla c_i(x_*^{(k)}) \right]_+ \right)^2 \end{aligned}$$

converges uniformly to the corresponding term in the expression for $\Psi_*^{(\infty)}(p)$ on the set $\{p : \|p\|_\infty \leq M_{\text{bound}}\}$. If $p_*^{(\infty)}$ is not a global minimiser of $\Psi_*^{(\infty)}$ then there exists a p such that $\Psi^{(k)}(p) < \Psi^{(k)}(p_*^{(\infty)})$ for all k sufficiently large, by the uniform convergence of $\{\Psi^{(k)}\}$ to $\Psi^{(\infty)}$ on $\{p : \|p\|_\infty \leq M_{\text{bound}}\}$. Hence $\Psi^{(k)}(p_*^{(k)}) > \Psi^{(k)}(p)$ for k sufficiently large — a contradiction. Therefore $p_*^{(\infty)}$ must be a global minimum of $\Psi_*^{(\infty)}$.

Now, for some $\beta \in (0, 1)$, $\alpha_*^{(k)}$ is chosen as the first member of the sequence

$1, \beta, \beta^2, \dots$ which satisfies the sufficient descent condition

$$\Xi^{(k)}(\alpha) = \Phi_*^{(k)} - \Phi \left(x_*^{(k)} + q_*^{(k)}(\alpha) \right) - \rho\alpha \left[\Psi_*^{(k)}(0) - \Psi_*^{(k)}(p_*^{(k)}) \right] \geq 0.$$

Denoting terms which tend to zero as $k \rightarrow \infty$ by $o_k(1)$,

$$\Xi^{(k)} = \Phi_*^{(k)} - \Phi \left(x_*^{(k)} + q_*^{(k)}(\alpha) \right) - \rho\alpha \left[\Psi_*^{(\infty)}(0) - \Psi_*^{(\infty)}(p_*^{(\infty)}) \right] + o_k(1), \quad (18)$$

by the convergence of $\{\Psi_*^{(k)}(p_*^{(k)})\}$, and $\{\Psi_*^{(k)}(0)\}$ to $\Psi_*^{(\infty)}(p_*^{(\infty)})$, and $\Psi_*^{(\infty)}(0)$. Now, as $q_*^{(k)}(\alpha) = \alpha p_*^{(k)} + \alpha^2 t_*^{(k)}$, and as Φ is locally Lipschitz,

$$\Xi^{(k)} = \Phi_*^{(k)} - \Phi \left(x_*^{(k)} + \alpha p_*^{(k)} \right) + O(\|t_*^{(k)}\|) \cdot o(\alpha) - \rho\alpha \left[\Psi_*^{(\infty)}(0) - \Psi_*^{(\infty)}(p_*^{(\infty)}) \right] + o_k(1),$$

where the $t_*^{(k)}$ part of $q_*^{(k)}$ gives rise to the $o(\alpha)$ term. Now, because $\|t_*^{(k)}\| < \|p_*^{(k)}\|$, by the convergence of $\{p_*^{(k)}\}$ to $p_*^{(\infty)}$, and by the convergence of $\{x_*^{(k)}\}$ to $x_*^{(\infty)}$,

$$\Xi^{(k)} = \Phi_*^{(\infty)} - \Phi \left(x_*^{(\infty)} + \alpha p_*^{(\infty)} \right) + o(\alpha) - \rho\alpha \left[\Psi_*^{(\infty)}(0) - \Psi_*^{(\infty)}(p_*^{(\infty)}) \right] + o_k(1),$$

where the $o(\alpha)$ term is now independent of k . Equation (6) implies

$$\Xi^{(k)} = \Psi_*^{(\infty)}(0) - \Psi_*^{(\infty)}(\alpha p_*^{(\infty)}) - \rho\alpha \left[\Psi_*^{(\infty)}(0) - \Psi_*^{(\infty)}(p_*^{(\infty)}) \right] + o(\alpha) + o_k(1). \quad (19)$$

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 (6) imply

$$\Psi_*^{(\infty)}(p_*^{(\infty)}) - \Psi_*^{(\infty)}(0) = \kappa_* < 0.$$

Whence, by the convexity of Ψ , equation (19), and the method of choosing $\alpha_*^{(k)}$ to satisfy $\Xi^{(k)} \geq 0$ imply there exists an $\alpha_{\text{lower}} > 0$ such that $\alpha_*^{(k)} \geq \alpha_{\text{lower}}$ for all k .

Once again from equation (18), $\Phi_*^{(k)} = \Phi_*^{(\infty)} + o_k(1)$ implies

$$\Xi^{(k)} \geq 0 \iff \Phi \left(x_*^{(k)} + q_*^{(k)}(\alpha_*^{(k)}) \right) \leq \Phi_*^{(\infty)} + \rho\alpha_*^{(k)}\kappa_* + o_k(1).$$

Thus as $\alpha_*^{(k)} \geq \alpha_{\text{lower}}$ for all k , and as $\kappa_* < 0$, the existence of an iterate $x_*^{(k)}$ satisfying equation (17) is clear. ♣

The above theorem shows that, under mild assumptions, convergence of the sequence of iterates generated by the algorithm to one or more critical points occurs. The issue of the rate at which convergence occurs is now addressed. Subject to additional assumptions required for superlinear convergence using the single parameter exact penalty function, it is shown that, for the purposes of superlinear convergence, the hybrid and single parameter exact penalty functions are equivalent.

Assume that the sequence $\{x^{(k)}\}$ converges to the feasible point x^* , and let \mathcal{A}^* be the set of indices of the active constraints at x^* . In addition to the applicability of theorem (3.1), the following assumption is made.

Assumption 3.2

- (a) f and c are three times continuously differentiable.
- (b) The set of active constraint normals $\{\nabla c_i(x^*) : i \in \mathcal{A}^*\}$ is linearly independent.
- (c) Strict complementarity and second order sufficient conditions for optimality hold at x^* .
- (d) The method of generating the sequence of Lagrange multiplier estimates $\{\lambda^{(k)}\}$ ensures that $\{\lambda^{(k)}\}$ converges to λ^* .

Theorem 3.1 implies x^* is a critical point of Φ . The lower semi-continuity of the Lagrange multiplier estimates $\lambda^{(k)}$ ensures $\mu > \kappa_1 \|\lambda^{(k)}\|_1$ for all k sufficiently large. The KKT conditions for $\ell_\infty \text{QP}^{(k)}$ then give $\lambda_\zeta^{(k)} < 0$ for all k sufficiently large. Therefore $\zeta^{(k)} = 0$ and so $p^{(k)}$ and $t^{(k)}$ are independent of $\nu^{(k)}$ and thus have the same values for both the hybrid and single parameter exact penalty functions.

The second order sufficiency conditions, together with an appropriate choice of each matrix $H^{(k)}$ [4, 9] imply that $p^{(k)} \rightarrow 0$ as $k \rightarrow \infty$; an appropriate choice of H matrices here would be, for example, any sequence $\{H^{(k)}\}$ for which the cluster point(s) were positive definite on the subspace orthogonal to the active constraint normals at x^* . The fact that $p^{(k)} \rightarrow 0$ as $k \rightarrow \infty$ means that (see eg. [9]), for large k , the solution to the ℓ_∞ QP $^{(k)}$ is determined by the system of equations

$$p^T \nabla c_i(x^{(k)}) + c_i(x^{(k)}) = 0 \quad \forall i \in \mathcal{A}^* \quad (20)$$

$$\text{and} \quad H^{(k)}p + \nabla f(x^{(k)}) + \sum_{i \in \mathcal{A}^*} \lambda_i^{(k)} \nabla c_i(x^{(k)}) = 0.$$

The constraints which are inactive at x^* are irrelevant, and so are ignored for the remainder of this section.

The equations in (20) imply that $c_i(x+p) = O(\|p\|^2)$. The second order correction is calculated by solving

$$\min_{t \in \mathbb{R}^n} \|t\|^2 \quad \text{such that} \quad t^T \nabla c_i + c_i(x+p) = 0 \quad \forall i \in \mathcal{A}^*.$$

The linear independence of the active constraint normals at x^* and the C^3 continuity of c imply that

$$t = O(\|p\|^2) \quad (21)$$

$$\text{and} \quad c_i(x+p+t) = O(\|p\|^3). \quad (22)$$

Now $p^{(k)} \rightarrow 0$ as $k \rightarrow \infty$, and (21) imply that $\|t^{(k)}\| < \|p^{(k)}\|$ holds for all sufficiently large k , hence the second order correction can always be used in the later iterations.

It is now shown that, for sufficiently large k , the sufficient descent condition (15) is satisfied by $\alpha = 1$ whenever the corresponding condition for the single parameter exact penalty function is satisfied by $\alpha = 1$. The criterion for sufficient descent with $\alpha = 1$ is

$$\Xi_H^{(k)} = \Phi^{(k)} - \Phi(x^{(k)} + p^{(k)} + t^{(k)}) - \rho [\Psi^{(k)}(0) - \Psi^{(k)}(p^{(k)})] \geq 0,$$

where the subscript H refers to the hybrid penalty function. The subscript S will be used to refer to the single parameter exact penalty function (ie. Φ with $\nu \equiv 0$). Omitting the (k) superscripts,

$$\begin{aligned} \Xi_H = f(x) + \mu\theta + \frac{1}{2}\nu\theta^2 - & \left[f(x+p+t) + \mu\theta(x+p+t) + \frac{1}{2}\nu\theta^2(x+p+t) \right] \\ & - \rho \left[\mu\theta + \frac{1}{2}\nu\theta^2 - p^T \nabla f(x) - \frac{1}{2}p^T H p - \mu\zeta - \frac{1}{2}\nu\zeta^2 \right]. \end{aligned}$$

The quantity corresponding to Ξ_H for the single parameter exact penalty function is Ξ_S , and may be obtained from Ξ_H by setting $\nu \equiv 0$. Now (20) implies $\zeta = 0$, and so (22) yields

$$\Xi_H = \Xi_S + \frac{1}{2}(1 - \rho)\nu\theta^2 + O(\|p\|^3).$$

Hence any updating process for H which guarantees unit steps using the single parameter penalty function by ensuring

$$\exists K, \exists \gamma > 0, \text{ and } \exists r < 3 \text{ such that } \forall k > K, \Xi_S^{(k)} \geq \gamma (p^{(k)})^r$$

will also ensure that unit steps are eventually taken when the single parameter penalty function is replaced by the hybrid penalty function. Under the standard conditions required for superlinear convergence, the single parameter exact penalty function and hybrid penalty function are equivalent.

4 Numerical Results and Concluding Remarks.

Full numerical results are presented in [2], however in order to show that the algorithm is effective in practice some results are given here. The significance of the second penalty parameter is discussed at length in [2] and is also discussed in [6, 7] in the context of semi-infinite programming. The algorithm was tested on the seven problems listed in [3] which are solved in [1]. On six of these problems algorithm A

problem	7	27	39	46	52	56	78
nr itns	7	22	13	14	8	9	7
nr fcn evals	10	26	14	20	13	13	10

Table 1: Results for the problems solved by Bartholomew-Biggs.

performed as well or better than the algorithm described in [1]. On the other problem algorithm A was slower by 4 iterations and 6 function evaluations. These results are listed in Table 1. The legend for this table is as follows: ‘nr itns’ denotes the number of iterations taken to solve the problem, and ‘nr fcn evals’ denotes the number of objective and constraint function evaluations performed in the process.

In conclusion, the algorithm has been shown to generate convergent sequences under mild conditions. Superlinear convergence is obtainable under the usual conditions on problems for which f and c are sufficiently continuous. The numerical results show that algorithm A is effective in practice.

References

- [1] Bartholomew-Biggs M. C., *Recursive quadratic programming methods based on the augmented Lagrangian*, Math. Prog. Study 31, pp 21–41 (1987).
- [2] Coope, I. D., and C. J. Price, *A two parameter exact penalty function for nonlinear programming*, Research Report number 96, Department of Mathematics and Statistics, University of Canterbury, New Zealand (1993).
- [3] Hock, W. and K. Schittkowski, *Test examples for nonlinear programming codes*, Springer-Verlag, Berlin, ©1981.

- [4] Mayne, D. Q. and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Math. Prog. Study 16, 45-61 (1982).
- [5] Pantoja, J. F. E. De O., and D. Q. Mayne, *Exact penalty function algorithm with simple updating of the penalty parameter*, J. Opt. Th. Appl. 69, pp 441–467.
- [6] Coope, I. D., and C. J. Price, *Numerical experiments in semi-infinite programming*, Research Report number 91, Department of Mathematics and Statistics, University of Canterbury, New Zealand (1993).
- [7] Price, C. J., *Nonlinear semi-infinite programming*, PhD thesis, Department of Mathematics, University of Canterbury (1992).
- [8] Powell, M. J. D., *A fast algorithm for nonlinearly constrained optimization calculations*, in *Numerical Analysis, Dundee 1977*, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer-Verlag, Berlin (1978).
- [9] Powell, M. J. D., *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in *Nonlinear Programming 3*, eds. O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, (1978).
- [10] Sahba, M., *A globally convergent algorithm for nonlinearly constrained optimization problems*, J. Opt. Th. Appl. 52, pp 291–309 (1987).