UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT

The Maratos Effect In Sequential Quadratic Programming Algorithms Using the L_1 Exact Penalty Function

I.D. Coope

CS-85-32

August, 1985

THE MARATOS EFFECT IN SEQUENTIAL QUADRATIC PROGRAMMING ALGORITHMS USING THE L_1 EXACT PENALTY FUNCTION

I.D. Coope

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

August 1985, Technical Report No. CS-85-32

ABSTRACT

It is shown that the Maratos effect can be avoided on certain convex programming problems by careful choice of the penalty parameters in the L_1 exact penalty function. Existing techniques for avoiding the effect on non-convex problems are also discussed and compared with a view to combining the best features of each approach in a single method.

Key Words: Maratos effect, sequential quadratic programming, L_1 exact penalty function.

Abbreviated title: The Maratos Effect.

THE MARATOS EFFECT IN SEQUENTIAL QUADRATIC PROGRAMMING ALGORITHMS USING THE L_1 EXACT PENALTY FUNCTION

I.D. Coope

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

1. Introduction

Sequential quadratic programming (SQP) methods for solving the nonlinear programming problem

$$\text{NLP:} \quad \text{minimize } f(x),\ x \in \mathbf{R}^n$$
$$\text{subject to } c_i(x) \leq 0 \quad i = 1,2,\ldots,m$$

construct a sequence of approximations $\{x^k;\ k = 1,2,\ldots\}$

$$x^{k+1} = x^k + \alpha^k d^k,\ 0 < \alpha^k \leq 1 \tag{1.1}$$

where d^k solves a quadratic programming problem of the form

$$\text{QP}(x^k): \quad \text{minimize } \tfrac{1}{2}\, d^T[H(x^k)]d + d^T \nabla f(x^k)$$
$$\text{subject to } d^T \nabla c_i(x^k) + c_i(x^k) \leq 0,\ i = 1,2,\ldots,m.$$

The matrix $H^k = H(x^k)$ must be chosen carefully if problem QP is to have a solution and to avoid some complications we assume that H^k is positive definite and the linearized constraints are consistent. Thus, QP is a strictly convex quadratic programming problem which has a unique solution d^k.

It is well known that if H^k is the matrix

$$H^k = \nabla^2 f(x^k) + \sum_1^m \hat{\lambda}_i^k \nabla^2 c_i(x^k) \tag{1.2}$$

(which may not be positive definite in general), then under suitable assumptions, which include second order sufficiency conditions holding at the solution x^* to problem NLP, the iteration (1.1) with $\alpha^k = 1$ converges locally at a second order rate (see [5] for example). The parameters $\hat{\lambda}_i^k$, $i = 1,2,\ldots,m$ in (1.2) are first order approximations to the optimal Lagrange multipliers $\{\lambda_i{}^*\}_1^m$ of problem

NLP. A suitable choice is

$$\hat{\lambda}^k = arg \min_{\lambda} \| \nabla f(x^k) + \sum \lambda_i \nabla c_i(x^k) \|_2 \tag{1.3}$$

except that it may be preferable to reset negative estimates to zero. Alternative choices for $\hat{\lambda}$ are suggested by Gill and Murray [9].

Many SQP algorithms have been proposed [1,7,8,10,11,13,14,15,17] including variable metric versions which do not require second derivative information. Wilson's [17] algorithm was the first SQP method to be proposed but his method required good initial estimates $(\hat{\lambda}^1, x^1)$. Therefore, we restrict attention to methods of the type advocated by Han [11] which force global convergence by choosing the steplength α^k in expression (1.1) in a way which gives "sufficient descent" of the L_1 exact penalty function

$$P(x,\mu) = f(x) + \sum \mu_i \, c_i(x)^+. \tag{1.4}$$

Thus, on every iteration the inequality

$$P(x^k + \alpha^k d^k, \mu) < P(x^k,\mu) \tag{1.5}$$

is satisfied, but usually a condition which is slightly stronger than inequality (1.5) is required (see section 3).

In order for there to be a correspondence between local solutions to problem NLP and local minimizers of $P(x, \mu)$ satisfying the constraints of NLP the inequalities

$$\mu_i > \lambda_i^* \quad i = 1,2,...,m \tag{1.6}$$

must be satisfied [3]. Also if the Lagrange multipliers, λ_i^k of problem QP(x^k) satisfy

$$\lambda_i^k \le \mu_i \quad i = 1,2,...,m \tag{1.7}$$

then d^k is a direction of descent for $P(x,\mu)$ at x^k [11]. Therefore, condition (1.4) can be satisfied for some $\alpha^k \in (0,1]$.

The fast local convergence of the SQP approach, combined with the global convergence property obtained by forcing descent of the L_1 penalty function provides a simple and elegant approach to solving nonlinear programming problems. Unfortunately, as Maratos [12] has demonstrated, the choice $\alpha^k = 1$ may not be allowed by inequality (1.5) for any positive values of μ_i , $i = 1,2,...,m$ in **any** neighbourhood of the solution on some problems. Thus the technique for forcing global convergence can result in a very slow, linear, asymptotic convergence rate.

A striking example of the "Maratos effect" is described by Powell [16]. Suggestions for countering the effect have been given by several authors and these are considered in section 4. However, the proposed remedies usually require extra calculation. Therefore we ask whether it is possible to **avoid** the Maratos effect by suitable choice of the penalty parameters. We show in the next section that for the special case of **convex** programming problems, the Maratos effect may be avoided.

2. The Convex Programming Problem

In this section we assume that the objective and constraint functions $f,\{c_i\}_1^m$ are twice continuously differentiable **convex** functions and we introduce the notation

$$G_0(x) = \nabla^2 f(x)$$

$$G_i(x) = \nabla^2 c_i(x) \quad i = 1,2,\ldots,m.$$

Thus for all x, $s \in \mathbf{R}^n$ and for each $i = 0,1,\ldots,m$

$$s^T[G_i(x)]s \geq 0. \tag{2.1}$$

We also assume that the SQP algorithm of section 1 is used to calculate the search direction d^k and that the sequence $\{x^k; k = 1,2,\ldots\}$ converges to a Kuhn-Tucker point x^* of problem NLP. Because we are interested in determining if there exist values of μ which ensure satisfaction of inequality (1.5) with $\alpha^k = 1$ when x^k is close to x^*, we further assume that for k sufficiently large

$$x^{k+1} = x^k + d^k \tag{2.2}$$

and hence that $\|d^k\|$ tends to zero as $k \rightarrow \infty$ under the assumption of convergence.

Finally we make two assumptions that are usual in asymptotic analysis of nonlinear programming algorithms:

i) The active constraint normals at x^* are linearly independent.

ii) Strict complementarity holds.

Thus λ^* is uniquely defined by the equation

$$\nabla f(x^*) + \sum \lambda_i^* \nabla c_i(x^*) = 0 \tag{2.3}$$

and in a neighbourhood of the solution $c_i(x) > 0$ for all indices $i \notin I^*$ where

$$I^* = \{j : \lambda_j^* > 0\}. \tag{2.4}$$

To simplify the notation we frequently drop the "k" superscript and adopt the convention that if an argument is missing then it is assumed to be x. Also we refer to **three** types of Lagrange multiplier vector : λ^* refers to the optimal multiplier vector satisfying (2.3); $\lambda(x^k)$ or λ^k or frequently just λ, refers to the multiplier vector of problem QP(x^k) introduced in section 1; $\hat{\lambda}(x^k)$ or $\hat{\lambda}^k$ is **any** Lagrange multiplier estimate satisfying the condition

$$\lim_{k \rightarrow \infty} \hat{\lambda}(x^k) = \lambda^*. \tag{2.5}$$

Finally we make extensive use of the order notation o(·). Thus we may write (2.5) in the form

$$\hat{\lambda}^k = \lambda^* + o(1), \quad as \quad k \to \infty.$$

Moreover, we note that the definition of λ^k implies

$$\lambda^k = \lambda^* + o(1). \tag{2.6}$$

In fact if $(x^k, \lambda^k) \to (x^*, \lambda^*)$ at a superlinear rate then the stronger condition

$$\lambda^k = \lambda^* + o(\|d^k\|) \tag{2.7}$$

holds (see Gill and Murray [9], for example, on second order Lagrange multiplier estimates).

Theorem 2.1

Let $H(x^k)$ be the matrix (1.2) and let d^k solve problem QP(x^k) of section 1. If the conditions of this section are satisfied and if $\mu_i > \lambda_i^*$, $i \in I^*$ then $\exists\ \delta(\mu)$ such that $\forall\ \| x - x^* \| < \delta$

$$\begin{aligned} P(x + d,\mu) - P(x,\mu) \le & -\tfrac{1}{2}\, d^T G_0^* d \\ & -\tfrac{1}{2} \sum_{i \in I^*} (2\lambda_i^* - \mu_i)\, d^T G_i^* d \\ & + o(\|d\|^2). \end{aligned} \tag{2.8}$$

Proof:

Using Taylor series expansions

$$\begin{aligned} P(x + d,\mu) - P(x,\mu) = & \ d^T \nabla f + \tfrac{1}{2}\, d^T G_0 d + o(\|d\|^2) \\ & + \sum \mu_i \,(c_i(x + d)^+ - c_i(x)^+), \end{aligned} \tag{2.9}$$

and

$$c_i(x + d) = c_i + d^T \nabla c_i + \tfrac{1}{2}\, d^T G_i d + o(\|d\|^2). \tag{2.10}$$

But the Kuhn-Tucker conditions for (d,λ) to solve problem QP(x) are

$$Hd + \nabla f + \sum \lambda_i \nabla c_i = 0 \tag{2.11}$$

$$\left.\begin{aligned} & \lambda_i\, (c_i + d^T \nabla c_i) = 0 \\ & \lambda_i \ge 0,\ c_i + d^T \nabla c_i \le 0 \end{aligned}\right\} \quad i = 1,2,\ldots,m. \tag{2.12}$$

Hence from (2.10), (2.12), (2.1)

$$c_i(x + d)^+ \le \tfrac{1}{2}\, d^T G_i d + o(\|d\|^2). \tag{2.13}$$

Also from (2.11), (2.12)

$$d^T \nabla f = \sum \lambda_i c_i - d^T H d. \tag{2.14}$$

Substituting expressions (2.13), (2.14) in (2.9) then gives the inequality

$$P(x + d,\mu) - P(x,\mu) \leq - d^T Hd + \sum \lambda_i c_i + \tfrac{1}{2}\, d^T G_0 d + \tfrac{1}{2} \sum \mu_i\, d^T G_i d - \sum \mu_i\, c_i^+ + o(\|d\|^2), \tag{2.15}$$

and we note that the strict complementarity assumption allows the summations in (2.15) to be taken over the set (2.4).

Now $\sum \lambda_i\, c_i - \sum \mu_i\, c_i^+ \leq 0$ if $\mu_i \geq \lambda_i$ and since $\mu_i > \lambda_i^*$ we deduce from expression (2.6) that there exists a neighbourhood of x^* for which $\mu_i > \lambda_i$. Therefore, we may replace inequality (2.15) by

$$P(x + d,\mu) - P(x,\mu) \leq -d^T Hd + \tfrac{1}{2}\, d^T G_0 d + \tfrac{1}{2} \sum_{i \in I^*} \mu_i d^T G_i d + o(\|d\|^2) \tag{2.16}$$

$$= - \tfrac{1}{2}\, d^T G_0 d - \tfrac{1}{2} \sum_{i \in I^*} (2\hat{\lambda}_i - \mu_i) d^T G_i d + o(\|d\|^2) \tag{2.17}$$

The required result follows from (2.5), (2.17) and the continuity of G_i, $i = 1,2,\ldots,m$. □

Corollary 2.2

If the matrix

$$G_0^* + \sum_{i \in I^*} (2\, \lambda_i^* - \mu_i) G_i^* \tag{2.18}$$

is uniformly positive definite than the descent condition (1.5) is obtained for all k sufficiently large. □

We note that under the assumptions of this section the matrix (2.18) is at least positive semidefinite if

$$\mu_i \leq 2\, \lambda_i^*. \tag{2.19}$$

Note also that if $f(x)$ is a uniformly convex function then much larger values of μ_i than (2.19) suggests will still give descent of the penalty function. However, much weaker conditions are possible since inequality (2.8) shows that descent occurs provided that there exists a constant $\gamma > 0$ such that

$$d^T [G_0^* + \sum_{i \in I^*} (2\lambda_i^* - \mu_i) G_i^*] d \geq \gamma\, d^T d. \tag{2.20}$$

3. Sufficient Descent

Condition (1.4) is not sufficiently strong to prove global convergence. Therefore, in this section we consider replacing (1.4) with the test

$$P(x^k + \alpha^k d^k,\mu) \leq P(x^k,\mu) + \rho\alpha^k \,\Delta\, (x^k,d^k), \tag{3.1}$$

where ρ is a small number in the open interval (0,1). Several choices are possible for the number Δ , for example

$$\Delta_1(x,d) = \lim_{\epsilon \to 0^+} \{P(x + \epsilon d,\mu) - P(x,\mu)\}/\epsilon, \tag{3.2}$$

$$\Delta_2(x,d) = d^T \nabla f(x) - \sum_1^m \mu_i \, c_i(x)^+, \tag{3.3}$$

$$\Delta_3(x,d) = ½ d^T \nabla f(x) + ½ \sum \lambda_i c_i(x) - \sum \mu_i \, c_i(x)^+ \tag{3.4}$$

all give rise to convergent schemes under appropriate assumptions. Expression (3.2) is just the directional derivative of $P(x,\mu)$ in the direction d. Expression (3.3) is the reduction $P(x + d,\mu) - P(x,\mu)$ that would occur if the functions $f,\{c_i\}_1^m$ were all linear and expression (3.4) is similar to (3.3) except that the term $½ \, d^T H d$ has been added to take account of some quadratic behaviour. In both (3.3) and (3.4) the expressions for Δ_2, Δ_3 have been simplified by noting that (d,λ) is the solution to problem QP(x) so that (2.4), (2.5) are applicable. Moreover, if $\mu \geq \lambda$ it is straightforward to establish the inequalities

$$\Delta_1(x,d) \leq \Delta_2(x,d) \leq \Delta_3(x,d). \tag{3.5}$$

It is important to note that each of expressions (3.2)-(3.4) is negative if d is non-zero, if H is positive definite and if $\mu > \lambda$. Because the inequality

$$P(x + \alpha d,\mu) \leq P(x,\mu) + \alpha \, \Delta(x,d) + o(\alpha), \;\; 0<\alpha \leq 1, \tag{3.6}$$

holds when Δ is any of the expressions (3.2)-(3.4) the test (3.1) can always be satisfied by some $\alpha \in (0,1]$. To see that the definitions (3.2)-(3.4) provide inequality (3.6) it is sufficient to note that (3.6) holds with equality when the definition (3.2) is used. Then inequalities (3.5) establish the result for (3.3), (3.4). A global convergence result is given by Powell [16] when the test (3.1) is used with the definition (3.4), but the choice (3.2) or (3.3) is also possible.

Thus we would prefer to extend theorem (2.1) to show that there exist values of $\mu > \lambda^*$ for which "sufficient descent" can be achieved by satisfying inequality (3.1) with $\alpha^k = 1$ when $||d^k||$ is sufficiently small.

Theorem 3.1

If the conditions of theorem 2.1 apply and Δ_3 is defined by expression (3.4) then

$$P(x + d,\mu) - P(x,\mu) \leq \Delta_3 + ½ \sum_{i \in I^*} (\mu_i - \lambda_i^*) d^T G_i^* d + o(||d||^2) \tag{3.7}$$

Proof

Rearranging inequality (2.15) and making use of the definition of H in (1.2) gives

$$P(x + d,\mu) - P(x,d) \leq - \tfrac{1}{2}\, d^T H d + \sum \lambda_i c_i - \sum \mu_i c_i^+$$
$$+ \tfrac{1}{2} \sum (\mu_i - \hat{\lambda}_i) d^T G_i d + o(\|d\|^2). \qquad (3.8)$$
$$= \Delta_3 + \tfrac{1}{2} \sum (\mu_i - \hat{\lambda}_i) d^T G_i d$$
$$+ o(\|d\|^2), \qquad (3.9)$$

where (3.9) is a consequence of definition (3.4) and equation (2.11). The required result follows from (2.5), strict complementarity and continuity of the matrices G_i ,i=1,2,...,m.□

Theorem 3.2

If the assumptions of theorem (2.1) apply and Δ_2 is defined by expression (3.3) then

$$P(x + d,\mu) - P(x,\mu) \leq \tfrac{1}{2}\, \Delta_2 + \tfrac{1}{2} \sum_{i \in I^*} (\mu_i - \lambda_i^*) d^T G_i^* d + o(\|d\|^2) \qquad (3.10)$$

Proof

We show that $\Delta_3 \leq \tfrac{1}{2}\, \Delta_2$ in some neighbourhood of x^*, then application of theorem 3.1 completes the proof.

From (3.3), (3.4) we deduce

$$\Delta_3 = \tfrac{1}{2}\, \Delta_2 + \tfrac{1}{2} \sum \lambda_i c_i - \tfrac{1}{2} \sum \mu_i c_i^+$$
$$= \tfrac{1}{2}\, \Delta_2 + \tfrac{1}{2} \sum \lambda_i c_i^- - \tfrac{1}{2} \sum (\mu_i - \lambda_i) c_i^+.$$

Therefore, $\Delta_3 \leq \tfrac{1}{2}\, \Delta_2$ if $\mu_i \geq \lambda_i \geq 0$, which occurs in some neighbourhood of x^* since $\mu_i > \lambda_i^*$ and $\lambda_i \rightarrow \lambda_i^*$. □

Expression (3.7) shows that there exist values of $\mu_i > \lambda_i^*$ such that the test (3.1) is satisfied for $\alpha^k = 1$ when $\|d^k\|$ is small enough, provided that $\rho < 1$ if Δ is defined by (3.4) or $\rho < \tfrac{1}{2}$ if (3.3) is used. This observation depends on the assumption that $\Delta_3 \leq \gamma d^T d$ for some $\gamma > 0$. It is clear that "sufficient" descent with a unit steplength can be achieved in a neighbourhood of the solution if the penalty parameters are not too large.

Thus, the Maratos effect can be avoided on some convex programming problems by careful choice of the penalty parameters in a neighbourhood of the solution.

4. Techniques for countering the Maratos effect

In [16] it is shown that the presence of a single non-convex constraint can give rise to the Maratos effect for all positive values of the penalty parameters. Therefore, the results of sections 2,3 cannot be extended to the non-convex case. Of course even in the convex case, the need to satisfy inequality (1.7) at points remote from the solution may result in a value of μ which is too large to avoid the Maratos effect. Therefore, we consider techniques that have been suggested to overcome the difficulties which do not depend on the choice of μ.

The watchdog technique of Chamberlain et al [2] preserves a local superlinear rate of convergence by allowing some iterations to violate the standard line search criterion based on the test (3.1) provided that a "relaxed" line search condition is satisfied. If a lower value for the penalty function (1.4) has not been calculated in a fixed number of iterations (five is suggested) because the relaxed criterion has been used then the technique discards the work of the last few iterations and essentially restarts from the point giving the current lowest value of the penalty function. The watchdog technique is highly efficient close to the solution because it requires very little extra work. However, we note that there may be much wasted calculation when a restart is necessary.

A different approach that depends on projecting infeasible iterates onto the active constraint manifold has been used by several authors [4,6,8,13]. If $x^k + d^k$ does not satisfy the test (3.1) then in [6,8] a quadratic arc is defined as $x^k + q^k(\alpha)$ where

$$q^k(\alpha) = \alpha d^k + \alpha^2 \delta^k, \tag{4.1}$$

δ^k is usually the vector

$$\delta^k = -C[C^TC]^{-1}c(x^k + d^k) \tag{4.2}$$

and C is the matrix of active constraint normals $[\nabla c_i^k : \lambda_i^k > 0]$. However, Gabay [8] resets $\delta^k = 0$ if

$$\|c(x^k + d^k + \delta^k)\| > \| c(x^k + d^k)\| \tag{4.3}$$

whereas Mayne and Polak [13] reset $\delta^k = 0$ if

$$\|\delta^k\| > \|d^k\|. \tag{4.4}$$

Then x^{k+1} is chosen as

$$x^{k+1} = x^k + q^k(\alpha^k), \;\; 0 < \alpha^k \leq 1 \tag{4.5}$$

with α^k the first member of the sequence $\{1,\beta,\beta^2,...\}$, $0< \beta< 1$, that satisfies the sufficient descent test (3.1) with $q^k(\alpha^k)$ replacing $\alpha^k d^k$. It is shown [8,13] that the Maratos effect cannot prevent a local superlinear rate of convergence with this modification. The Mayne and Polak [13] technique is preferred because it does not require the extra constraint evaluations in (4.3) when $\delta^k = 0$ is used.

A variation on the projection technique that has just been described is to let δ^k solve the equality quadratic programming problem

$$\underset{\delta}{\text{minimize}} \; \tfrac{1}{2}\, \delta^T H^k \delta + \delta^T(\nabla f^k + H^k d^k), \tag{4.6}$$

$$\text{subject to} \;\; c_i(x^k + d^k) + \delta^T \nabla c_i^k = 0, \quad i : \lambda_i^k > 0,$$

where λ^k is still the Lagrange multiplier vector at the solution to problem QP(x^k). This quadratic programming subproblem is easily solved in $0(n^2)$ arithmetic operations from the factorizations available at the solution to problem QP(x^k) which has already been solved to obtain d^k. In fact the solution to (4.6) can be written

$$\delta^k = -H^{-1}C[C^TH^{-1}C]^{-1}c(x + d), \tag{4.7}$$

where the superscripts have been omitted for economy of notation. This form emphasizes the projection nature of the step δ. Note also that if H = I we recover the projection step (4.2) used in [4,8,13].

There is a close correspondence with problem (4.6) and Fletcher's [6] approach. The essential difference is that there are no simple bounds on the variables and only the current active set of constraints is imposed (as equalities) instead of all of the constraints. As in [8,13] we do not recommend defining δ^k by (4.6) if $x^k + d^k$ is far from x^*. However, we prefer not to use either of the tests (4.3) or (4.4) in deciding whether or not to reset δ^k to zero, because they are dependent on constraint scaling. Instead we recommend the following procedure for preventing the occurrence of the Maratos effect when the step $x^k + d^k$ is unacceptable according to the test (3.1). If $x^k + d^k$ gives descent of the Lagrangian function, that is, if

$$L(x^k + d^k,\lambda^k) < L(x^k,\lambda^k) \tag{4.8}$$

where $L(x,\lambda)$ is the Lagrangian function

$$L(x,\lambda) = f(x) + \sum \lambda_i c_i(x) \tag{4.9}$$

then δ^k is defined as the solution to problem (4.6). Otherwise $\delta^k = 0$ is used. Then x^{k+1} is calculated as described in the paragraph that includes equation (4.5). Thus the recommended technique incorporates what we consider to be the best features of each of the methods [2,4,6,8,13]. It is shown in [2] that the test (4.8) is always satisfied in a neighbourhood of the solution. Moreover, if the matrices $\{H^k ; k = 1,2,...\}$ are positive definite with bounded condition numbers then the projection step (4.7) has similar properties to the projection (4.2). We note that even if $\nabla^2 L(x^*,\lambda^*)$ is indefinite then suitable positive definite matrices $\{H^k\}$ can be defined which give superlinear convergence [16].

Tables 4.1 and 4.2 display numerical results for the recommended technique on the problems described in [12,16] since these were designed to exhibit the Maratos effect. For these simple examples the matrix H^k was chosen to be $H(x^*)$ because this is positive definite in each case. The results show that the recommended technique is successful even though the values of μ are much larger than would normally be used.

k	x_1	x_2	c	P	μ
0	0.6000	1.2000	0.0000	1.8000	100
1	0.7620	0.9465	0.0005	1.5271	
2	0.9366	0.5031	0.0034	1.4697	
3	1.0010	0.0003	0.0041	1.4122	
4	1.0000	0.0000	0.0000	1.0001	
0	0.6000	1.2000	0.0000	1.8000	1000
1	0.6855	1.0766	0.0000	1.6607	
2	0.7541	0.9608	0.0000	1.5400	
3	0.8569	0.7432	0.0002	1.5308	
4	0.9179	0.5675	0.0003	1.4224	
5	0.9533	0.4300	0.0002	1.3126	
6	0.9882	0.2176	0.0002	1.2675	
7	1.0000	0.0000	0.0001	1.1415	

Table 4.1 Results for Maratos' [12] example

If the second order correction δ^k is always set to zero then the number of iterations required to obtain similar accuracy to that given in the tables is 142 and and 1244 for $\mu = 100$ and $\mu = 1000$ respectively on the Maratos problem; 148 and 1249 on the Powell problem. In each case very small steps had to be taken in order to achieve a reduction in the penalty function causing a dramatic reduction in the rate of convergence. However, it is interesting to note that the unit steplength was always allowed in the last few iterations.

k	x_1	x_2	c	P	μ
0	0.8000	0.6000	0.0000	-0.8000	100
1	0.8810	0.4733	0.0001	-0.8671	
2	0.9683	0.2515	0.0008	-0.8749	
3	1.0005	0.0001	0.0010	-0.8877	
4	1.0000	0.0000	0.0000	-1.0000	
0	0.8000	0.6000	0.0000	-0.8000	1000
1	0.8428	0.5383	0.0000	-0.8348	
2	0.8771	0.4804	0.0000	-0.8649	
3	0.9284	0.3716	0.0001	-0.8668	
4	0.9589	0.2837	0.0001	-0.8938	
5	0.9766	0.2150	0.0001	-0.9214	
6	0.9941	0.1088	0.0001	-0.9326	
7	1.0000	0.0000	0.0000	-0.9643	

Table 4.2 Results for Powell's [16] example

5. Concluding Remarks

The results of sections 2,3 help to explain why the Maratos effect is not often observed in practice. Many test problems for nonlinear programming problems are convex because such problems are easy to construct with a guaranteed solution. Moreover, we note that non-convex inequality constraints may be present in the problem, if they are not active at the solution, without affecting the results of sections 2,3. This is not to say that the Maratos effect should be ignored but we do believe that more attention should be focussed on the problem of choosing suitable values for the penalty parameters since this can go a long way towards alleviating the difficulties.

The projection techniques described in section 4 have been shown to be effective in giving vast reductions in the number of iterations that would otherwise be required when a poor choice for μ is made. However, even greater reductions are possible if extra derivative information is made available to the projection technique. Fukushima [7] in a recent and interesting approach has shown that if the constraint gradients are evaluated at $x + d$ then a new quadratic programming subproblem can be formulated which gives very high accuracy. Of course the amount of work required for an iteration when the technique is applied is almost doubled and it may be preferable to simply use another full iteration of the SQP method from the point $x^k + d^k$ (using the same H^k matrix) to define the projection step δ^k when needed.

Acknowledgements

This work was undertaken whilst the author was a Visiting Professor at the Department of Computer Science, University of Waterloo and has benefited greatly from many discussions with N.I.M. Gould. The excellent typesetting was achieved by Zoë Kaszas using the Imagen typesetter and Troff.

REFERENCES

[1] M.C. Bartholomew-Biggs, *Recursive quadratic programming methods for nonlinear constraints*, in Nonlinear Optimization 1981, ed. M.J.D. Powell, Academic Press, New York, 1982, pp. 213-221.

[2] R.M. Chamberlain, M.J.D. Powell, C. Lemarechal and H.C. Pederson, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Mathematical Programming Study 16 (1982) 1-17.

[3] C. Charalambous, *A lower bound for the controlling parameter of the exact penalty function*, Mathematical Programming 17 (1978) 278-290.

[4] T.F. Coleman and A.R. Conn, *Nonlinear programming via an exact penalty function: Asymptotic analysis*, Mathematical Programming 24 (1982) 123-136.

[5] R. Fletcher, *Practical methods of optimization, Volume 2, Constrained Optimization*, John Wiley, Chichester, 1981.

[6] R. Fletcher, *Second order corrections for nondifferentiable optimization*, in *Numerical Analysis, Dundee 1981*, ed. G.A. Watson, Springer-Verlag, Berlin, 1982, pp. 85-114.

[7] M. Fukushima, *A successive quadratic programming algorithm with global and superlinear convergence properties*, Technical Report #84022, Department of Applied Mathematics and Physics, Kyoto University, Japan (1985).

[8] D. Gabay, *Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization*, Mathematical Programming Study 16 (1982) 18-44.

[9] P.E. Gill and W. Murray, *The computation of Lagrange multiplier estimates for constrained minimization*, Mathematical Programming 17, (1979) 32-60.

[10] S.P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Mathematical Programming 11 (1976) 263-282.

[11] S.P. Han, *A globally convergent method for nonlinear programming*, Journal of Optimization Theory and Applications 22 (1977) 297-309.

[12] N. Maratos, *Exact penalty function algorithms for finite dimensional and control problems*, Ph.D. Dissertation, Imperial College, University of London, 1978.

[13] D.Q. Mayne and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Mathematical Programming Study 16 (1982) 45-61.

[14] W. Murray, *An algorithm for constrained minimization*, in *Optimization*, ed. R. Fletcher, Academic Press, London and New York, 1969, pp. 247-258.

[15] M.J.D. Powell, *A fast algorithm for nonlinearly constrained optimization calculations*, in *Numerical Analysis, Dundee 1977*, ed. G.A. Watson, Springer-Verlag, Berlin, 1978, pp. 144-157.

[16] M.J.D. Powell, *Variable metric methods for constrained optimization*, in *Mathematical Programming: The State of the Art, Bonn 1982*, eds. A Bachem, M. Gröschel and B. Korte, Springer-Verlag, Berlin, 1983, pp. 288-311.

[17] R.B. Wilson, *A simplicial algorithm for concave programming*, Ph.D. Dissertation, Graduate School of Business Administration, Harvard University, 1963.