ESSENTIALLY QUADRATIC PENALTY FUNCTIONS AND THE MARATOS EFFECT

by

Nikiforos A. Papadakis

and

Panos Y. Papalambros

Technical Report UM-MEAM-89-06

June 1989

DESIGN LABORATORY
THE UNIVERSITY OF MICHIGAN
ANN ARBOR

ESSENTIALLY QUADRATIC PENALTY FUNCTIONS AND THE MARATOS EFFECT

by

N. A. Papadakis
Research Fellow

and

P. Y. Papalambros
Associate Professor

Department of Mechanical Engineering
and Applied Mechanics
The University of Michigan
Ann Arbor, Michigan 48109

June 1989

ABSTRACT

An augmented Lagrangian function with essentially quadratic penalty terms is proposed as a merit function for performing line search in sequential quadratic programming methods. Convergence of the stepsize to unity is established thus eliminating occurence of the Maratos Effect. Global convergence issues are also discussed. An inexact line search routine based on the Goldstein termination criteria, together with numerical experiments are presented.

Key Words: Merit functions, sequential quadratic programming, Maratos Effect, inexact line search, superlinear convergence, global convergence

1. Introduction

Sequential quadratic programming (SQP) methods have been used extensively for solving nonlinear constrained optimization problems. An SQP method requires three basic steps. At the beginning of each iteration, a quadratic programming approximation to the original nonlinear problem, at the current estimate of the solution $\mathbf{x}_k$, is performed. The quadratic subproblem is then solved to obtain a search direction $\mathbf{d}_k$, emanating from $\mathbf{x}_k$. At the second step, a line search along $\mathbf{d}_k$ is performed to reduce the value of an appropriate merit function and thus obtain a better estimate $\mathbf{x}_{k+1}$. Finally, an approximation of the Hessian of the Lagrangian function at $\mathbf{x}_{k+1}$ is computed using first-order information and a variable metric formula, in preparation for the next iteration.

These methods have several desirable features. In particular, SQP methods with appropriate merit functions are globally convergent (see, e.g., Han [7]). Locally, close to the problem solution, if unity step sizes are accepted by the line search routine, Q-superlinear convergence is achieved (Han [8], Powel [13]). Only first order information is needed and the methods perform equally well for both equality and inequality constrained models.

An unfavorable characteristic of SQP methods, when they employ nonsmooth merit functions, is the so-called Maratos Effect. As the solution of the problem is approached, the step length calculated by the line search routine may not converge to unity and the superlinear convergence may be destroyed (see, Maratos [10], Chamberlain [2]). Several remedies have been developed to deal with the Maratos Effect. Chamberlain et al.[3], developed a watchdog technique, while Mayne & Polak [11], Gabay [6] and Fukushima [5], developed a line search strategy along a curved direction that tends to follow the constraint surface. Differentiable merit functions have also been used as merit functions, see, e.g., Di Pillo & Grippo [4]. Schittkowski [19], Powell & Yuan [16], and Rustem

[17], devised differentiable merit functions and developed line search procedures that accept unit step sizes, when approaching the solution.

Continuous penalty functions deserving more careful consideration are those with second derivatives increasing as the feasible set is approached. In a previous investigation (Papadakis & Papalambros [12]), we examined the use of the merit function

$$m(\mathbf{x}, c) = f(\mathbf{x}) + c \sum_{j} n_j(\mathbf{x})\, g_j(\mathbf{x}) \tag{1.1}$$

where $f(\mathbf{x})$ is the function to be minimized, $g_j(\mathbf{x})$ is the constraint function of the jth violated constraint, c an appropriate penalty coefficient and

$$n_j(\mathbf{x}) = \begin{cases} 1 & \text{if } |g_j(\mathbf{x})| > 1 \\ 1 - \ln^3 g_j(\mathbf{x}) & \text{otherwise.} \end{cases}$$

A line search algorithm based on this merit fiunction, local convergence analysis and numerical results are discussed in the cited reference. Those promising results motivated further study of this type of merit functions.

In the present article we consider a line search algorithm based on essentially quadratic penalty functions, first proposed by Kort & Bertsekas [9]. In the next section, we discuss the basic reasons favoring the use of essentially quadratic functions in the context of SQP methods. We then establish local convergence of the stepsize to unity, and we analyze the global convergence problem. We discuss an algorithmic implementation and illustrate the behavior of the method using two simple examples.

2. Motivation

Consider the nonlinear constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \end{aligned} \tag{2.1}$$

where $\mathbf{x}$ is in $\boldsymbol{R}^n$ and f, $\mathbf{g} = [g_1, \ldots, g_m]^T$ are twice continuously differentiable functions. The quadratic programming approximation at $\mathbf{x}_k$ (which is the current estimate of the solution $\mathbf{x}^*$ at the beginning of the kth iteration) has the form

$$\text{minimize} \quad \frac{1}{2}\mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k + \nabla \mathbf{f}_k^T \mathbf{d}_k \tag{2.2a}$$

$$\text{subject to} \quad \mathbf{g}_k + \nabla \mathbf{g}_k^T \mathbf{d}_k \leq \mathbf{0} \tag{2.2b}$$

where $\mathbf{d}_k$ is the direction of search emanating from $\mathbf{x}_k$, and ∇f_k, $\nabla \mathbf{g}_k$ denote the gradients of f, $\mathbf{g}$ evaluated at $\mathbf{x}_k$. Matrix $\mathbf{B}_k$ is a positive definite approximation to the Hessian of the Lagrangian function

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^{*T} \mathbf{g}(\mathbf{x}) \tag{2.3}$$

at the location $\mathbf{x}_k$, $\lambda^* = [\lambda_1^*, \ldots, \lambda_m^*]^T$ being the Lagrange multiplier vector at the solution of Eq. (2.1).

Solving the problem in Eq.(2.2) gives the direction of search $\mathbf{d}_k$ and multipliers $\lambda_k = [\lambda_1, \ldots, \lambda_m]^T$ corresponding to the linearized constraints. The next iterate $\mathbf{x}_{k+1}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \tag{2.4}$$

where α_k is selected so that an appropriate merit function is decreased. Finally, the Hessian approximation $\mathbf{B}_{k+1}$ is computed with a variable metric formula (Powell [14]), using function and gradient values at $\mathbf{x}_k$, $\mathbf{x}_{k+1}$, and the values of the multiplier vector λ_k as an approximation to λ^*.

Consider now an augmented Lagrangian - type of merit function with the quadratic term replaced by an "essentially quadratic" penalty function (Bertsekas 1982). Let us define

$$m(\mathbf{x}, \lambda, c, \rho) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + c\, P(\rho, \mathbf{g}(\mathbf{x})) \tag{2.5}$$

where

$$P(\rho, \mathbf{g}(\mathbf{x})) = \sum_J \frac{1}{\rho_j} \mid g_j(\mathbf{x}) \mid^{\rho_j} + \frac{1}{2} g_j(\mathbf{x})^2 \tag{2.6}$$

with

$$J = \{ j \mid g_j(\mathbf{x}) > 0 \} \tag{2.7}$$

and $\rho = [\rho_1, ..., \rho_J]$, $\rho_j \in (1, 2]$ and $c > 0$. Function $P(\rho, \mathbf{g}(\mathbf{x}))$ is continuously differentiable and strictly convex with respect to $\mathbf{g}(\mathbf{x})$, and

$$\begin{aligned} P(\rho, 0) &= 0 \\ \nabla P(\rho, 0) &= \mathbf{0}. \end{aligned} \tag{2.8}$$

From Eq.(2.6) we may easily verify that $\nabla^2 P$ is not continuously differentiable. In fact, for any $\rho_j \in (1, 2]$, $\nabla^2 P$ tends to infinity as $g_j(\mathbf{x})$, $j \in J$, tend to zero. The use of augmented Lagrangian functions of the form in Eq.(2.5) with $\rho_j \in (1, 2]$ is discussed by Bertsekas in [1]. It is noted there, that multiplier methods based on essentially quadratic penalty functions, even with a finite penalty coefficient, are superlinearly convergent. However, a significant disadvantage of these methods is that the essentially quadratic terms may produce ill-conditioned unconstrained problems and the superlinear convergence may not be achieved.

The purpose of using the penalty function described by Eq.(2.6) as a merit function in SQP methods, is to provide strong and differentiable representation of the remaining infeasibility, as the solution of the problem is approached. It has been observed that the

Maratos Effect usually occurs in the vicinity of $\mathbf{x}^*$, when the gradients $\nabla f(\mathbf{x})$ and $\nabla \mathbf{g}(\mathbf{x})$ are almost linearly dependent and almost orthogonal to the direction of search. In that respect, this phenomenon resembles the instability of line search algorithms for unconstrained optimization emerging when the direction of search and the gradient of the objective are almost orthogonal. A feature of SQP methods which enhances significantly the appearance of the Maratos Effect is that, as $\mathbf{x}^*$ is approached, $\mathbf{d}_k$ lies almost entirely in the null space of $\nabla \mathbf{g}(\mathbf{x}_k)$. During the late stages of the process, the direction $\mathbf{d}_k$ is given approximately by a quasi-Newton step, minimizing the quadratic approximation of the Lagrangian function. Thus, the use of function $P(\rho, \mathbf{g}(\mathbf{x}))$ in Eq.(2.5) amplifies the role of the existing infeasibility in the line search, while the "orthogonality" of $\mathbf{d}_k$, $\nabla \mathbf{g}(\mathbf{x}_k)$ tends to eliminate the effect of ill-conditioning mentioned before.

3. Convergence Analysis

To insure the existence of a point $\mathbf{x}_{k+1}$ satisfying the descent condition

$$m_k(0) > m_k(\alpha), \tag{3.1}$$

where $m_k(\alpha)$ denotes the merit function defined by Eqs.(2.5) to (2.7) evaluated at $(\mathbf{x}_k + \alpha\mathbf{d}_k, \lambda_k, c_k, \rho_k)$, $\alpha \in (0, 1]$, we impose the sign restriction

$$m_k'(0) < 0. \tag{3.2}$$

From Eq.(2.5) to (2.7) we have

$$m_k'(0) = \nabla f_k^T \mathbf{d}_k + \lambda_k^T \nabla g_k \mathbf{d}_k + c_k \nabla P_k \mathbf{d}_k \tag{3.3}$$

where

$$\nabla P_k = \sum_J \beta_j \nabla g_i(\mathbf{x}_k), \tag{3.4}$$

$$\beta_j = g_j(\mathbf{x}_k)^{\rho_j - 1} + g_j(\mathbf{x}_k) > 0. \tag{3.5}$$

Also, from the first order necessary conditions for optimality of problem Eq.(2.2) we have

$$\nabla f_k + \nabla \mathbf{g}_k \lambda_k = - \mathbf{B}_k \mathbf{d}_k \tag{3.6}$$

From Eqs.(3.4) and (2.2b) we obtain

$$\nabla P_k^T \mathbf{d}_k \leq - \sum_J \beta_j g_j(\mathbf{x}_k) \leq 0, \tag{3.7}$$

Finally, from Eqs.(3.3) and (3.7), and given $\mathbf{B}_k > 0$, we have

$$m_k'(0) \leq - \mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k + c_k \sum_J \beta_j g_j(\mathbf{x}_k) < 0, \tag{3.8}$$

which is satisfied for any $c_k > 0$.

Next we examine the behavior of function $m_k(\alpha)$ as $\mathbf{x}_k$ tends to be a stationary point of problem (2.1). As it will be shown in this section, a particularly good choice of the components of vector ρ_k is

$$\rho_j = \begin{cases} 1 - [1 + g_j + \nabla g_j \mathbf{d}_k] g_j / [g_j ln g_j + \nabla g_j \mathbf{d}_k] & \text{if } g_j(\mathbf{x}_k) \in (0,1) \\ 2 & \text{otherwise} \end{cases} \tag{3.9}$$

Let us also define

$$\gamma_j = (\rho_j - 1)\, g_j^{\rho_j - 2} + 1 > 0. \tag{3.10}$$

An important consequence of the selection described by Eq. (3.9) is that the merit function $m_k(\alpha)$ can be very easily convexified. In fact we have:

<u>Lemma 3.1.</u> If we keep the penalty coefficient fixed at any positive value c, there exists a certain iteration K such that the function $m_k(\alpha)$, $k > K$, is convex.

Proof: We may assume without loss of generality that as $\mathbf{d}_k \to \mathbf{0}$ there exists a sufficiently large k such that the activity of the constraint set remains invariant (the set J defined by Eq. (2.7) becomes independent of k), and it makes no difference if the violated constraints are treated as equalities, see, e. g. Powell [13]. Furthermore, $m_k(\alpha)$, $\alpha \in (0, 1]$, can be approximated by its second order Taylor expansion

$$m_k(\alpha) = m_k(0) + \alpha\, m_k(0) + \frac{1}{2}\alpha^2 m_k''(0), \tag{3.11}$$

with

$$m_k''(0) = \mathbf{d}_k^T \left[\nabla^2 L_k + c\sum_J [\gamma_j \nabla g_j \nabla g_j^T + \beta_j \nabla^2 g_j]\right] \mathbf{d}_k, \tag{3.12}$$

$$\nabla^2 L_k = \nabla^2 f_k + \sum_J \lambda_j^k \nabla^2 g_j. \tag{3.13}$$

For k large and $\mathbf{x}_k$ close enough to $\mathbf{x}^*$, we know that there exists a positive number τ such that for $c' > \tau$ the matrix

$$\nabla^2 L_k + c' \sum_J \nabla g_j \nabla g_j^T$$

becomes positive definite [1]. Therefore, if c satisfies

$$c > \tau / \min\{\gamma_j, j \in J\}\ , \tag{3.14}$$

then

$$\nabla^2 L_k + c\sum_J \gamma_j \nabla g_j \nabla g_j^T \tag{3.15}$$

is positive definite. Additionally, as $\mathbf{d}_k \to \mathbf{0}$, Eq. (3.9) combined with Eqs. (3.5) and (3.10) result in

$$\beta_j \to \frac{1}{e}, \tag{3.16}$$

$$\gamma_j \approx \frac{-1}{g_j \ln g_j} \to \infty. \tag{3.17}$$

Therefore, if c_k satisfies Eq. (3.14) and for k large, we have

$$m''(0) = \mathbf{d}_k^T \left[\nabla^2 L_k + c \sum_J \gamma_j \nabla g_j \nabla g_j^T\right] \mathbf{d}_k + O(\|\mathbf{d}_k\|^2) > 0, \tag{3.18}$$

Given that $\gamma_j(\nabla g_j\, \mathbf{d}_k)^2 = O(\|\mathbf{d}_k\| / |\ln g_j|)$, $m_k(\alpha)$ is convex. ●

Lemma 3.1 is based on the fact that the representation of the existing infeasibility by $P(\rho, \mathbf{g}(\mathbf{x}))$ is convex with respect to $\mathbf{g}(\mathbf{x})$ and strong enough to insure convexity of $m_k(\alpha)$, even if c_k tends to zero, provided it satisfies Eq. (3.14). Therefore, if c is fixed to a certain positive value, $m_k(\alpha)$ becomes convex and we do not need to guess the threshold value τ to achieve convexity.

Next we examine the convergence of the stepsize α_k to unity as $\mathbf{x}_k \to \mathbf{x}^*$. We assume that the sequences $\{\mathbf{x}_k\}$, $\{\mathbf{B}_k\}$, $\{\mathbf{d}_k\}$ are bounded, $\nabla \mathbf{g}(\mathbf{x}_k)$ has full rank and that the matrices $\mathbf{B}_k$ are positive definite and uniformly bounded.

Lemma 3.2 . Let the above assumptions be satisfied and let the components of vector ρ_k be selected according to Eq. (3.9). Then, for k large enough,

$$m_k(1) < m_k(0). \tag{3.19}$$

Proof: From Eqs. (3.3) to (3.6) we have

$$m_k'(0) = \mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k - c \sum_J \beta_j \nabla g_j \mathbf{d}_k \tag{3.20}$$

Since matrices $\mathbf{B}_k$ are bounded, Eq. (3.20), may be written as

$$m_k'(0) = -c \sum_J \beta_j \nabla g_j d_k + O(\|\mathbf{d}_k\|^2). \tag{3.21}$$

To show that inequality (3.19) holds, it is sufficient to show that

$$m_k'(0) + m_k''(0)/2 \leq -m_k''(0)/2. \tag{3.22}$$

Furthermore, because of Eqs. (3.12), (3.20), and Lemma 3.1, it is enough to have

$$\mathbf{d}_k^T\left[\nabla^2 L_k - \mathbf{B}_k + c \sum_J \beta_j \nabla^2 g_j\right] \mathbf{d}_k + c \sum_J\left[g_j(\nabla g_j \mathbf{d}_k)^2 + \beta_j \nabla g_j \mathbf{d}_k\right] < 0 \tag{3.23}$$

Taking into account Eqs. (3.15) and (3.21), Eq. (3.23) becomes

$$\sum_J \left[\gamma_j(\nabla g_j \mathbf{d}_k)^2 + \beta_j \nabla g_j \mathbf{d}_k\right] + O(\|\mathbf{d}_k\|^2) = 0. \tag{3.24}$$

We examine next the inequality

$$h(\rho_j) = \gamma_j \nabla g_j \mathbf{d}_k + \beta_j \geq 0 \ , \ j \in J, \tag{3.25}$$

where γ_j, β_j are given by Eqs. (3.5) and (3.10) respectively. We have

$$h''(\rho_j) = g_j^{\rho_j - 2}\, ln g_j \left[g_j\, ln g_j + \nabla g_j \mathbf{d}_k \,(2 + (\rho_j - 1)\, ln g_j) \right] \tag{3.26}$$

Therefore, if ρ_j satisfies the inequality

$$1 < \rho_j < \zeta_j\,, \quad \zeta_j = 1 - g_j / \nabla g_j \mathbf{d}_k - 2 / ln g_j\,, \tag{3.27}$$

function $h(\rho_j)$ is convex. Moreover,

$$h'(1) < 0,$$
$$h(2) = 2(g_j + \nabla g_j \mathbf{d}_k) < 0, \tag{3.28}$$

and for $\mathbf{d}_k$ small enough,

$$h(1) = 1 + g_j + \nabla \mathbf{g}_j \mathbf{d}_k > 0. \tag{3.29}$$

From the above we conclude that Eq. (3.25) treated as equality, has only one solution in the interval $[1, \zeta_j]$. The first Newton iteration for solving Eq. (3.25) results in

$$\rho_j = 1 - \left[1 + g_j + \nabla g_j \mathbf{d}_k\right] g_j \Big/ \left[g_j ln g_j + \nabla g_j \mathbf{d}_k\right], \tag{3.30}$$

Because of the geometry of function $h(\rho_j)$ discussed before, ρ_j given by Eq. (3.30) satisfies Eq. (3.25) as strict inequality and for k large enough Eq. (3.23) and thus Eq. (3.19) becomes valid ●

Lemma 3.2 guarantees the acceptance of unity stepsize as $\mathbf{d}_k \to 0$. Moreover, from Eq. (3.22) we have

$$m_k(1) - m_k(0) < -c \sum \gamma_j(\nabla g_j \, \mathbf{d}_k)^2 + O(\|\mathbf{d}_k\|)^2 \qquad (3.31)$$

and because of Eq. (3.17)

$$m_k(1) - m_k(0) < -c \, O(\|\mathbf{d}_k\| / |\ln g|) + O(\|\mathbf{d}_k\|^2) \qquad (3.32)$$

Therefore, although the terms $O(\|\mathbf{d}_k\|^2)$ depend on c (see also Eq. (3.23)), small values of c are enough to insure validity of Eq. (3.32). In fact, from Eqs. (3.23), (3.31) and (3.32) we may conclude that small values of the penalty parameter c are enough to compensate for the effect of the Hessian approximations ($\mathbf{B}_k$ converges to $\nabla^2 L(\mathbf{x}^*, \lambda^*)$ on the tangent plane only) and eliminate the effect of the constraint curvature. Finally, from Eq. (3.30) we may also conclude that as $\mathbf{d}_k \rightarrow \mathbf{0}$, $\rho_j \rightarrow 1$, and the penalty function $P(\rho_k, \mathbf{g}_k)$ tends uniformly to the absolute value penalty function while differentiability is maintained.

4. Global Convergence

The exponent vector ρ_k and the form of the merit function defined in the last section depend on the location $\mathbf{x}_k$. Although the analysis we performed so far ensures the existence of $\mathbf{x}_{k+1}$, such that Eq. (3.1) is valid and unity stepsize becomes acceptable at the end of the process, it may arise questions concerning the global convergence of the algorithm. The penalty coefficient may be kept constant but the global descent condition

$$m_k(\mathbf{x}_k) > m_k(\mathbf{x}_{k+1}) > m_{k+1}(\mathbf{x}_{k+1}) > m_{k+1}(\mathbf{x}_{k+2}) > \ldots \qquad (4.1)$$

may not be valid. Note that in Eq. (4.1), function $m_k(.)$ is evaluated with $\lambda = \lambda_k$, $\rho = \rho_k$, $k = 0,1, \ldots.$

One means to achieve a consistent measure of the progress of the algorithm, is to establish a correspondence between the sequence in Eq. (4.1) and the associated sequence of values of the *absolute value merit function* defined by

$$m_a(\mathbf{x}) = f(\mathbf{x}) + c_a P_a(\mathbf{g}(\mathbf{x})), \tag{4.2}$$

where the set J is defined by Eq. (2.7), and

$$P_a(\mathbf{g}(\mathbf{x})) = \sum_J g_j(\mathbf{x}) , \tag{4.3}$$

$$c_a > \max_J (\lambda_j^k, \lambda_j^*). \tag{4.4}$$

It is known that if c_a satisfies Eq. (4.4), and $m_a(.)$ is used as the objective of the line search algorithm so that

$$m_a(\mathbf{x}_k) > m_a(\mathbf{x}_{k+1}) > \ldots, \tag{4.5}$$

then the SQP method is globally convergent (Han [7]).

Next we define the sets

$$\begin{aligned} L &= \{j \mid g_j(\mathbf{x}_k) > 0, g_j(\mathbf{x}_{k+1}) > 0\} \\ M &= \{j \mid g_j(\mathbf{x}_k) > 0, g_j(\mathbf{x}_{k+1}) \le 0\} \\ N &= \{j \mid g_j(\mathbf{x}_k) \le 0, g_j(\mathbf{x}_{k+1}) > 0\} \end{aligned} \tag{4.6}$$

Let us first assume that $L \Upsilon M \neq \emptyset$. Inequality (3.1) implies (4.5), if

$$c_a\big[P_a(\mathbf{g}_k) - P_a(\mathbf{g}_{k+1})\big] \ge c\,\big[P(\rho_k, \mathbf{g}_k) - P(\rho_k, \mathbf{g}_{k+1})\big] + \lambda_k^T\big[\mathbf{g}_k - \mathbf{g}_{k+1}\big]. \tag{4.7}$$

If $\mathbf{x}_{k+1}$ satisfies

$$P_a(\mathbf{g}_k) > P_a(\mathbf{g}_{k+1}) \tag{4.8}$$

then c_a can be selected so that both Eqs.(4.7) and (4.4) are valid for any iteration. Let α_k be a stepsize satisfying the local descent condition (3.1) and small enough so that Eq. (4.7) may be approximated by

$$\mathbf{d}_k^T \sum_L \left[\alpha_k \nabla g_j(\mathbf{x}_k) + \frac{\alpha_k{}^2}{2} \nabla^2 g_j(\mathbf{x}_k)\mathbf{d}_k \right] + \sum_M g_j(\mathbf{x}_k) +$$

$$\mathbf{d}_k^T \sum_N \left[(\alpha_k - \alpha_j) \nabla g_j(\mathbf{x}_k + \alpha_j \mathbf{d}_k) + \frac{(\alpha_k - \alpha_j)^2}{2} \nabla^2 g_j(\mathbf{x}_k + \alpha_j \mathbf{d}_k)\mathbf{d}_k \right] < 0, \tag{4.9}$$

where α_j, $j \in N$, is the steplength with

$$g_j(\mathbf{x}_k + \alpha_j \mathbf{d}_k) = 0. \tag{4.10}$$

From the definition Eq. (4.6) we also have

$$\nabla g_j(\mathbf{x}_k)\mathbf{d}_k \leq - g_j(\mathbf{x}_k) < 0, j \in L, \tag{4.11}$$

and

$$\nabla g_j(\mathbf{x}_k + \alpha_j \mathbf{d}_k)\mathbf{d}_k > 0, j \in N. \tag{4.12}$$

Furthermore, α_k satisfies Eq. (3.1) and since $m_k(\alpha_k)$ behaves like a quadratic form, for any $\alpha \in (0, \alpha_k]$, we have

$$m_k(\alpha) \leq m_k(\alpha_k). \tag{4.13}$$

The cardinality of set N is an increasing function of α_k. Because of Eq. (4.13) we may select a stepsize $\alpha'_k \in (0, \alpha_k]$ such that the local descent condition (3.1) is valid and the contribution of set N in Eq. (4.9) becomes small. Moreover, from Eq. (4.13) the terms contributed by the sets L, M in (4.9) are negative and bounded away from zero. Because of the definition of set M and Eq. (4.11) we may always select a steplength such that Eq. (4.9) and consequently Eq. (4.8) is valid.

The case $L \Upsilon M = \emptyset, N \neq \emptyset$ may be treated in a similar way. We only note that in this case we may select α_k such that

$$f_{k+1} - f_k < 0, \tag{4.14}$$

and Eq. (4.5) becomes valid. Moreover, as $\mathbf{x}_k$ approaches $\mathbf{x}^*$, the sets M, N become empty and from Eq. (4.9) we may immediately conclude that Eq. (4.5) does not prevent unity stepsize to be accepted by the line search routine, and so superlinear convergence may be achieved.

Finally, we note that to check if $\{\mathbf{x}_k, k = 0, 1,...\}$ is globally convergent, we only need additional constraint and objective function evaluations. The stepsize α_k may be reduced until inequalities (4.8) or (4.14) become valid, while local descent with respect to $m_k(\alpha)$ is always satisfied. Since the line search algorithm can be considered as a closed map on a compact set, global convergence can be achieved.

5. The Algorithmic Implementation and Numerical Results

An inexact line search algorithm was developed using the Goldstein's termination criterion. At the ith iteration of the line search algorithm, a stepsize $a_{k,i}$ is accepted if

$$B_k(\alpha_{k,i}) \geq m_k(\alpha_{k,i}) \geq \Gamma_k(\alpha_{k,i}), \tag{5.1}$$

where

$$B_k(\alpha_{k,i}) = m_k(0) + \varepsilon_k m'(0)\, \alpha_{k,i}, \tag{5.2}$$

$$\Gamma_k(\alpha_{k,i}) = m_k(0) + (1 - \varepsilon_k)\, m'(0)\, \alpha_{k,i} \tag{5.3}$$

with $\varepsilon_k \in (0, \frac{1}{2})$ and $\alpha_{k,1} = 1$.

If $m_k(\alpha_{k,i}) < \Gamma_k(\alpha_{k,i})$, $\alpha_{k,i+1}$ is calculated as the intersection of the second order approximation of $m_k(\alpha)$ (based on $m_k(0)$, $m_k'(0)$, $m_k(\alpha_{k,i})$) with $B_k(\alpha)$. Similarly, if $m_k(\alpha_{k,i}) > B_k(\alpha_{k,i})$, $\alpha_{k,i+1}$ is selected as the intersection of the same approximation with $\Gamma_k(\alpha)$. The algorithm also employs a bracketing procedure. An interval $[\gamma_i, \beta_i]$, $\gamma_1 = 0$, $\beta_1 > 1$, which contains a subset of acceptable values of α_k is used. If the value of $\alpha_{k,i}$ is such that $m(\alpha_{k,i}) < \Gamma(\alpha_{k,i})$, the lower bound γ_j is increased by the amount $\delta_i = \sigma(\beta_i - \gamma_i)$, σ small. Similarly, if $m(\alpha_{k,i}) > B(\alpha_{k,i})$, then β_i is decreased by δ_i.

Let us discuss the performance of the above algorithm on two well known examples, devised to test line search routines. We have chosen $\varepsilon_k = 10^{-4}$, $\gamma_o = 0$, $b_o = 1$, $\sigma = 0.1$. In both cases the Hessian of the Lagrangian has been approximated by using the Broyden-Fletcher-Goldfarb-Shanno update and the value of penalty coefficient was set equal to one. Finally, we have used the termination criteria $\|\mathbf{d}_k\| \leq 10^{-5}$, $\|g(x)\| \leq 10^{-5}$.

The first problem (see also Maratos [10]) is

$$\text{minimize} \quad x_1^2 + x_2^2$$

$$\text{subject to} \quad (x_1 + 1)^2 + x_2^2 - 4 = 0,$$

with solution $(1, 0)^T$. Four different initial points and two different initial Hessian approximations, shown in Table 1, have been examined. In all cases, fast convergence of the stepsize to unity has been achieved.

The second problem devised by Powel [1982] is

$$\text{minimize} \quad 10(x_1^2 + x_2^2 - 1) - x_1$$

$$\text{subject to} \quad x_1^2 + x_2^2 - 1 = 0$$

with solution $(1,0)^T$. Again, the numerical results summarized in Table 2 confirm rapid convergence of the stepsize to unity and superlinear convergence of the SQP algorithm.

Direct comparison of the results described in Tables 1 and 2 indicates that the convergence of the stepsize to unity is nearly unaffected by the guess of $\mathbf{B}_0$ and by the way the sequence $\{\mathbf{B}_k, k = 0, 1, ...\}$ converges to $\nabla^2 L(\mathbf{x}^*, \lambda^*)$ on the tangent plane. It appears that this is due to the fact that the use of the merit function described in Section 2 emphasizes the part of the Newton iteration concerning feasibility, thus reducing the effect of $\mathbf{B}_k$ on the line search algorithm. Besides the results presented in Tables 1 and 2, additional tests have been carried out with penalty coefficients varying from 0.5 to 2.0. In most of the cases, the number of iterations needed to achieve unity stepsizes was identical to those given in Tables 1 and 2.

6. Conclusion

The purpose of this work was to examine the use of augmented Lagrangian merit functions, with the quadratic term replaced by an essentially quadratic penalty function. The use of essentially quadratic penalty terms emphasized the effect of the existing infeasibility in the line search routine and increased the reliability of the method. The analysis of Section 3 established convergence of the stepsize to unity and gave a rational way for selecting the exponents of the essentially quadratic functions. As the solution of the problem is approached, the essentially quadratic terms converge to the absolute value penalty function providing a strong differentiable representation of the remaining infeasibility. The use of this augmented Lagrangian merit function, enhancing the role of the constraint linearization, arrests the growth of the penalty parameter, restricts the effect of the Hessian approximation and limits the effect of the constraint curvature on the line search routine.

In Section 4 we described a procedure that insures that the sequence$\{\mathbf{x}_k, k = 0, 1, \ldots\}$, generated by decreasing the merit function along the direction of search, is globally convergent. This procedure is based on a parallelism between the values of the augmented Lagrangian function and the absolute penalty value merit function and requires only function evaluations. Numerical results on the two classical test problems were very encouranging. Further testing on larger problems is necessary to confirm general utility of this merit function.

Acknowledgement

This research has been partially supported by the National Science Foundation Grant No. DMC85-14721 at The University of Michigan. This support is gratefully acknowledged.

TABLE 1. Performance of the algorithm on Maratos' problem.

x_0	(0.985, 0.2)		(1.002, 0.1)		(0.99999, 0.2)		(0, $\sqrt{3}$)	
B_0	I	2I	I	2I	I	2I	I	2I
MI	4	5	3	4	4	5	8	7
I*	1	1	1	1	1	1	4	3*
FE	5	6	4	5	5	6	12	9
GE	4	5	3	4	4	5	8	7

*For iterations 1, and 3 to 7, $\alpha_k = 1$ while $\alpha_2 < 1$

TABLE 2. Perforance of the algotithm on Powell's problem.

x_0	(0.8, 0.6)		(0.1, 0)		(50,50)	
B_0	I	20I	I	20I	I	20I
MI	6	8	7	7	13	14
I*	1	3*	1	1	1	11**
FE	7	10	8	8	14	15
GE	5	7	6	6	12	13

* For iterations 1, and 3 to 8, $\alpha_k = 1$, while $\alpha_2 < 1$

** For iterations 1 to 9, and 11 to 14, $\alpha_k = 1$, while $\alpha_{10} < 1$

Abbreviations:

MI: Major Iterations to Achieve Convergence

I*: Iteration at which stepsize converged to unity

FE: Objective and Constraint Function Evaluation

GE: Objective and Constraint Gradient Evaluation

References

[1] Bertsekas, D. P., "Constrained Optimization and Lagrange Multiplier Methods," Academic Press, New York, 1982.

[2] Chamberlain, R. M., "Some Examples of Cylcing in Variable Metric Methods for Constrained Optimization," Mathematical Programming Study, Vol. 16, pp. 18-44, 1982.

[3] Chamberlain, R. M., Powell, M. J. D., Lemarechal C., and Pedensen H. C., "The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization," Mathematical Programming Study, Vol. 16, pp. 1-17, 1982.

[4] Di Pillo, G. and Grippo, L., "A New Class of Augmented Lagrangians in Nonlinear Programming," SIAM J. Control Optim. 17, pp 618-628, 1979.

[5] Fukushima, M., "A Successive Quadratic Programming Algorithm with Global and Superlinear Convergence Properties," Mathematical Programming 35, pp. 253-264, 1986.

[6] Gabay, D., "Reduced Quasi-Newton Methods with Feasibility Improvement for Nonlinear Constrained Optimization," Mathematical Programming Study, Vol. 16, pp. 18-44, 1982.

[7] Han, S. P., "A Globally Convergent Method for Nonlinear Programming," Journal of Optimization Theory and Applications, Vol 22, pp. 297-309, 1977.

[8] Han, S.P., "Superlinearly Convergent Variable-Metric Algorithms for General Nonlinear Programming Problems," Mathematical Programming, Vol. 11, pp. 243-283, 1976.

[9] Kort, B. W. and Bertsekas, D. P., "A New Penalty Function Method For Constrained Minimization," Proc. IEEE Conf. Desition & Control, New Orleans, 1972.

[10] Maratos, N., "Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems," University of London, Ph.D. Thesis, 1978.

[11] Mayne, D. Q., and Polak, E., "A superlinearly Convergent Algorithm for Constrained Optimization Problems," Mathematical Programming Study, Vol 16, pp. 45-61, 1982.

[12] Papadakis, N. A., and Papalambros, P. Y., "A New Merit Function For Variable Metric Constrained Optimization," Technical Report # UM-MEAM-89-02, Dept of Mechanical Engineering & Applied Mechanics, The University of Michigan, Ann Arbor, 1989.

[13] Powell, M. J. D., "The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations," in Nonlinear Programming 3, (O. L. Manasarian, R. R. Meyer and S. M. Robinson, eds.), Academic Press, 1976, pp 27-65.

[14] Powell, M. J. D., "Variable Metric Methods for Constrained Optimization," in Mathematical Programming, the State of the Art, (A. Bachem, M. Grötschel, and B. Korte, eds.), Springer-Verlag, 1982, pp. 288-311.

[15] Powell, M. J. D., "Variable Metric Methods for Constrained Optimization," Nonlinear Optimization, Theory and Algorithms,(L. C. W. Dixon, E. Spedicato and G. P. Szegö, eds.), Birkhauser, Boston 1980, pp. 279-294.

[16] Powell, M. J. D. and Yuan Y., "A Recursive Quadratic Programming Algorithm that uses Differentiable Exact Penalty Functions," Mathematical Programming 35, pp. 265-278, 1986.

[17] Rustem B., "Convergent Stepsizes for Constrained Optimization Algorithms," Journal of Optimization Theory and Applications, Vol. 49, No. 1, pp. 135-160, 1986.

[18] Schittkowski K., "A Numerical Comparison of Optimization Software Using Randomly Generated Test Problems," in Numerical Optimization of Dynamic Systems, (L. C. W. Dixon and G. P. Szegö, eds.), North-Holland Publishing Co., Amsterdam, 1980.

[19] Schittkowski K., "On the Convergence of a Sequential Quadratic Programming Method with an Augmented Langrangian Line Search Function," Technical Report Vol 82-4,Systems Optimization Laboratory, Department of Operations Research, Stanford University, 1982.