A Robust Choice of the Lagrange Multiplier in the SQP Newton Method*

Debora Cores[1] *Richard A. Tapia*[2]

[1]INTEVEP, S.A. (Research Center of the Venezuelan Oil Company)
Apartado 76343, Caracas 1070-A, Venezuela
cores@intevep.pdv.com

[2]Department of Computational and Applied Mathematics and the
Center for Research on Parallel Computations
Rice University, Houston, Texas, 77251-1892
rat@caam.rice.edu

Abstract

It is known that the augmented Lagrangian SQP-Newton method depends on the penalty parameter only through the multiplier in the Hessian matrix of the Lagrangian function. This effectively reduces the augmented Lagrangian SQP-Newton method to the Lagrangian SQP-Newton method where only the multiplier estimate depends on the penalty parameter. In this work, we construct a multiplier estimate that depends strongly on the penalty parameter and derive a choice for the penalty parameter that attempts to make the Hessian matrix, restricted to the tangent space of the constraints, positive definite and well conditioned. We demonstrate that the SQP-Newton method with this choice of Lagrange multipliers is locally and q-quadratically convergent. Encouraging numerical experimentation is included and shows that our approach exploits the good global behavior of the augmented Lagrangian SQP-Newton method.

Keywords: Augmented Lagrangian, successive quadratic programming, nonlinear programming, quadratic convergence, Lagrange multiplier estimate.

1 Introduction

In this work we consider the equality constrained nonlinear optimization problem, where the objective function and the equality constraints are nonlinear functions. The solution of this nonlinear optimization problem appears in many real applications.

*The second author was supported by DOE DEFG05-86ER25017.
Received december 1996, revised version accepted june 1997.

For example: chemical equilibrium and process control; oil extraction, blending, and distribution; pipe network design for reliable water distribution systems; location of facilities (i.e. factories, fire or police stations, etc); least squares estimation of statistical parameters and data fitting; among many others.

One of the most used and well known methods for solving the equality constrained optimization problem is the successive quadratic programming method (SQP-Newton method). Here in this work we study the SQP-Newton approach to equality constrained nonlinear optimization. The SQP-Newton method requires an approximation of the Lagrange multipliers and the solution to an equality constrained quadratic programming problem at each iteration. The objective function of this quadratic problem is a quadratic approximation of the Lagrangian function (Lagrangian SQP-Newton method) and, the equality constraints are linear approximations of the constraints of the equality constrained optimization problem. The solution of this quadratic problem may not be unique or may not exist if the Hessian matrix of the Lagrangian function is not positive definite on the tangent space of the constraints. Moreover, even when the solution of this quadratic problem exists, it may not be reliable if the Hessian matrix is not well conditioned. Several decades ago it became fashionable, promoted by the work of Hestenes in 1969, to work with the augmented Lagrangian instead of the Lagrangian. Many researchers considered the augmented Lagrangian SQP-Newton method instead of the Lagrangian SQP-Newton method. The objective function of the quadratic problem in the augmented Lagrangian SQP-Newton method is a quadratic approximation of the augmented Lagrangian function. Clearly, in this case, we not only require an approximation to the Lagrange multipliers but a value for the penalty parameter. However, the choice of the penalty parameter turned out to be an extremely delicate issue since the augmented Lagrangian SQP-Newton method is quite sensitive to this choice.

It is known that the augmented Lagrangian SQP-Newton method depends on the penalty parameter only through the multiplier in the Hessian matrix of the Lagrangian function. This effectively reduces the augmented Lagrangian SQP-Newton method to the Lagrangian SQP-Newton method where only the multiplier estimate depends on the penalty parameter. The objective of the current work is to derive a choice for the penalty parameter so that the Hessian matrix, restricted to the tangent space of the constraints, is positive definite and well conditioned. Moreover, we desire that this choice of the penalty parameter does not destroy the local and q-quadratic convergence of the Lagrangian SQP-Newton method. This work has similarities to Tapia [13] where he used the penalty parameter to obtain effective BFGS and DFP secant updates for equality constrained optimization in the SQP framework.

This paper is organized as follows. In Section 2 we give background material on the augmented Lagrangian SQP method for equality constrained optimization and motivate our approach for picking the penalty parameter. In Section 3 we pose an ideal constrained optimization problem whose solution gives the desired penalty parameter. In Section 4 this ideal barrier function problem is studied closely for the special case of equality constrained optimization with only one constraint. The procedure for obtaining the penalty parameter for the special case of one constraint is extended to many constraints in Section 5. Our idea is to use, as the basis of our extension, the one-constraint approach developed in Section 4. We do this in two distinct ways; the first we call the parallel approach and the second we call the sequential approach. In Section 6 we establish the local and q-quadratic convergence of the SQP-Newton method with this new choice of Lagrange multipliers. Finally, in Section 7 we present numerical results using our new choice of Lagrange multipliers in the SQP-Newton method. The numerical results obtained by using this new approach for computing the Lagrange multipliers are encouraging. It was possible to achieve convergence in many examples, where the SQP-Newton method with well-known choices for the Lagrange multipliers (the least-squares multiplier estimate [10], the Miele-Cragg-Levy multiplier [12], and the multiplier associated with the solution of the quadratic program (see Tapia [14])) did not produce iteration sequences which converged. For a complete discussion on the standard choices of the Lagrange multipliers see [1], [2], [14] and [8].

A significant part of this work is the study of the effectiveness and the robustness of this new choice of Lagrange multiplier estimate in the SQP-Newton framework. Hence, in our numerical comparisons we did not embed the SQP-Newton method in a globalization strategy. Our reason for not doing this is that we feel that good global behavior of the local method speaks strongly to the effectiveness of our multiplier choice. The problems used to test this new choice of the Lagrange multipliers are the classical test problems from Hock and Schittkowski [11]. The size of these test problems oscillates from 1 to 10 unknowns and from 1 to 5 constraints. These are relative small problems since we are interested in showing the robustness and performance of the SQP-Newton method with this new choice of the lagrange multipliers when compared with the standard choices of the Lagrange multipliers. We do not intent to show the performance of the SQP-Newton method for large scale equality constrained optimization problems.

2 Preliminaries

We are concerned with the nonlinear equality constrained optimization problem:

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & h(x) = 0 \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be smooth nonlinear functions and $m < n$. The Lagrangian function associated with problem (1) is

$$\ell(x, \lambda) = f(x) + h(x)^T\lambda; \quad \text{and} \tag{2}$$

the augmented Lagrangian function associated with problem (1) is

$$L(x, \lambda, \rho) = f(x) + h(x)^T\lambda + \frac{\rho}{2}h(x)^T h(x). \tag{3}$$

In the above $\lambda = (\lambda_1, \lambda_2, ..., \lambda_m)^T$ is called the Lagrange multiplier and the parameter $\rho \in \mathbb{R}$, $\rho \geq 0$ is called the penalty parameter.

Throughout this work we will assume that problem (1) has a solution x_* with associated Lagrange multiplier λ_* and we also assume the standard assumptions for the analysis of Newton's method:

(A1) $f, h_i \in C^2(D)$, where D is an open convex neighborhood of the local solution x_* of problem (1), and $\nabla^2 f$ and $\nabla^2 h_i$ are Lipschitz continuous at x_*.

(A2) $\nabla h(x_*)$ has full rank.

(A3) $z^T \nabla_x^2 \ell(x_*, \lambda_*) z > 0$ for all $z \neq 0$ satisfying $\nabla h(x_*)^T z = 0$.

Consider the augmented Lagrangian SQP-Newton method (see [1], [3], [8], and [14]). For a given iterate (x_k, λ_k) and penalty parameter ρ_k we let

$$\begin{aligned} x_{k+1} &= x_k + \Delta x_k \\ \lambda_{k+1} &= \lambda_k + \Delta \lambda_k \end{aligned} \tag{4}$$

where Δx_k is the solution, and $\Delta \lambda_k$ is the multiplier associated with the solution, of the quadratic program

$$\begin{array}{ll} \textit{minimize} & q(\Delta x) = \nabla_x L(x_k, \lambda_k, \rho_k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_x^2 L(x_k, \lambda_k, \rho_k) \Delta x \\ \textit{subject to} & \nabla h(x_k)^T \Delta x + h(x_k) = 0. \end{array} \tag{5}$$

The second-order sufficiency conditions for problem (5) are the associated first order necessary conditions and

$$z^T \nabla_x^2 L(x_k, \lambda_k, \rho_k) z > 0 \quad \text{for all} \quad z \neq 0 \quad \text{such that} \tag{6}$$

$$\nabla h(x_k)^T z = 0. \tag{7}$$

Clearly, if $\nabla_x^2 L(x_k, \lambda_k, \rho_k)$ does not satisfy these sufficiency conditions, then problem (5) may not have a solution; however, even if $\nabla_x^2 L(x_k, \lambda_k, \rho_k)$ satisfies these conditions, the solution may not be reliable.

Consider the augmented Lagrangian SQP-Newton method. The first-order necessary conditions associated with the quadratic program (5) are

$$\begin{pmatrix} \nabla_x^2 L(x_k, \lambda_k, \rho_k) & \nabla h(x_k) \\ \nabla h(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x_k, \lambda_k, \rho_k) \\ -h(x_k) \end{pmatrix}. \tag{8}$$

It is not difficult to see that (8) reduces to

$$\begin{pmatrix} \nabla_x^2 \ell(x_k, \lambda_k + \rho_k h(x_k)) & \nabla h(x_k) \\ \nabla h(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = \begin{pmatrix} -\nabla_x \ell(x_k, \lambda_k) \\ -h(x_k) \end{pmatrix}. \quad (9)$$

From equation (8) and the equivalence between (8) and (9) we observe that:

- The augmented Lagrangian SQP iterate x_{k+1} depends on the penalty constant ρ_k only through the multiplier estimate in the Hessian matrix $\nabla_x^2 \ell(x_k, \lambda_k + \rho_k h(x_k))$.

Hence, we consider using, as a Lagrange multiplier estimate in the Hessian matrix of the Lagrangian function, a quantity of the form

$$\hat{\lambda}_k = U(x_k) + C_k h(x_k) \quad (10)$$

where $C_k = diag(c_k)$, $c_k \in \mathbb{R}^m$ is considered as penalty parameter vector instead of a scalar, and $U(x_k)$ is a Lagrange multiplier formula which does not depend on c_k. Ideally we propose to choose the penalty parameter in (10) so that:

(C1) the Hessian matrix of the Lagrangian functional $\nabla_x^2 \ell(x_k, \hat{\lambda}_k)$ is positive definite on the subspace $Z(x_k) = \{z \in \mathbb{R}^n : \nabla h(x_k)^T z = 0\}$.

(C2) the Hessian matrix $\nabla_x^2 \ell(x_k, \hat{\lambda}_k)$, restricted to the tangent space of the constraints, is well conditioned.

(C3) the local convergence properties of the Lagrangian SQP-Newton method are maintained.

Moreover, we expect that as a by-product we will obtain improved global behavior of the augmented Lagrangian SQP-Newton method. It is worth mentioning that none of the Lagrange multiplier formulas in the literature guarantee that conditions (C1), (C2) and (C3) are satisfied.

3 A Robust Choice of The Lagrange Multipliers

Consider (x_k, λ_k) the k^{th} iterate of the Lagrangian SQP-Newton method for problem (1). Let $U(x)$ be a Lagrange multiplier formula, i.e., $\lambda_* = U(x_*)$ when (x_*, λ_*) is a stationary point of problem (1). Also, let $B(x_k)$ be a matrix whose columns form an orthonormal basis for the null space of $\nabla h(x_k)^T$. For $c \in \mathbb{R}^m$ we denote the Hessian of the Lagrangian function on the tangent space of the constraints by $H(c)$ and

$$H(c) := B(x_k)^T \nabla_x^2 \ell(x_k, U(x_k) + C h(x_k)) B(x_k) \quad (11)$$

As before, C denotes the diagonal matrix $diag(c)$.

The task at hand is to determine $c_k \in \mathbb{R}^m$ so that the objectives (C1)-(C3) of Section 2 hold with $\hat{\lambda}_k$ given by (10). Hence, we look for c_k such that $H(c_k)$ is positive definite and well conditioned.

The above considerations lead us to an ideal optimization problem of the form

$$\begin{array}{ll} minimize & f(c) = \|Ch(x_k)\|_2^2 + \mu\|H(c) - I\|_F^2 \\ subject\ to & H(c) \in S \end{array} \tag{12}$$

for some choice of $\mu > 0$, and where S is an appropriate set contained in the set of symmetric and positive definite matrices (SPD).

It is important to mention that the barrier term on the optimization problem (12) tends to find a well conditioned Hessian matrix, $H(c)$, since it is related to the Byrd and Nocedal measure function [4], also studied by Fletcher [9], which has the flavor of the condition number of a matrix in SPD. For more details see [5]. We now make an observation that will facilitate our presentation. Clearly, if a particular component of the constraint vector $h(x_k)$ is zero, then the corresponding component of the penalty parameter c_k does not enter into the multiplier estimate (10); and hence does not enter into problem (12). For this reason we will always choose this component of c_k to be zero. Therefore, problem (12) will be only to select components of the penalty parameter corresponding to nonzero components of $h(x_k)$. Hence, without loss of generality in the analysis of problems of the form (12) we will assume that the vector $h(x_k)$ has no zero components.

In problem (12) we expect to define the set S in terms of the properties of the eigenvalues of $H(c)$. Hence it is most satisfying that the eigenvalues of $H(c)$ are invariant with respect to the choice of the matrix $B(x_k)$ used in defining the regularized reduced Hessian $H(c)$. This fact is the topic of the following theorem.

Theorem 3.1: Let $A \in \mathbb{R}^{(n\times n)}$ be a symmetric matrix. Let B_1 and $B_2 \in \mathbb{R}^{(n\times k)}$ with $n \geq k$ be such that each has orthonormal columns and these two sets of columns each form a basis for the subspace $V \subseteq \mathbb{R}^n$. Then $B_1^T AB_1$ and $B_2^T AB_2$ have the same eigenvalues.

Proof. It is clear that $B_2^T AB_2$ and $B_1^T AB_1$ are symmetric matrices. Thus, the matrices $B_2^T AB_2$ and $B_1^T AB_1$ have k real eigenvalues and the corresponding eigenvectors form an orthonormal set. Suppose $\lambda_i \in \mathbb{R}$ is any eigenvalue of the matrix $B_2^T AB_2$ with corresponding eigenvector $x_i \in \mathbb{R}^k$. Then,

$$B_2^T AB_2 x_i = \lambda_i x_i.$$

It is straightforward to see that there exists an orthogonal matrix $Q \in \mathbb{R}^{(k\times k)}$ such that

$$B_2 = B_1 Q.$$

Therefore,

$$QQ^T B_1^T A B_1 Q x_i = \lambda_i Q x_i.$$

Since $QQ^T = I_{k\times k}$ we have that $B_1^T A B_1 (Qx_i) = \lambda_i Q x_i$ which implies, λ_i is an eigenvalue of the matrix $B_1^T A B_1$ with corresponding eigenvector Qx_i and establishes the theorem. □

Theorem 3.2: If S is a nonempty, closed and convex set then problem (12) has a unique solution.

Proof. Clearly f is continuous. Moreover, since we are assuming that no components of $h(x_k)$ are zero, f is also uniformly convex. Furthermore, the constraint set S is nonempty, closed and convex. This proves the theorem. □

The remainder of this work is concerned with finding the constraint set S and the numerical approach for solving problem (12). Any constraint set S that satisfies the hypothesis of Theorem 3.2 and that guarantees the positive definiteness of the Hessian matrix on the tangent space of the constraints, $H(c)$, is a good choice. However, if the constraint set S requires to know all the eigenvalues of the matrix $H(c)$, then solving the optimization subproblem (12) can be more complicated than solving the equality constrained optimization problem (1). Towards this end, we will define the constrait set S such that no eigenvalues of the matrix $H(c)$ are required.

4 The One Constraint Case $(m = 1)$

In this section, we consider problem (1) with only one constraint since it is of theoretical importance for the case of more constraints. For this particular case, we construct a constrained optimization problem, that will be solved at each iteration of the SQP-Newton method, to obtain an approximation to the solution of the ideal barrier problem (12).

Suppose we are at the k^{th} iteration of the SQP-Newton method. The reduced Hessian matrix $H(c)$ for the one constraint case can be written as,

$$H(c) = A + chD$$

where,

$$\begin{aligned} h &= h(x_k), \\ A &= B(x_k)^T(\nabla^2 f(x_k) + U(x_k)\nabla^2 h(x_k))B(x_k) \text{ and,} \\ D &= B(x_k)^T \nabla^2 h(x_k) B(x_k). \end{aligned}$$

In order to define the constraint set S we present a sufficient condition for a matrix $A \in \mathbb{R}^{n\times n}$ to be positive definite. This condition does not require the eigenvalues of the matrix A and it is an easy condition to evaluate. This condition can also be applied to nonsymmetric matrices and it was established by Tarazaga in [15] and [16]. In the following theorem the Tarazaga's condition is stated (see [16]).

Theorem 4.1: (Tarazaga)
Let $A \in \mathbb{R}^{n\times n}$. If

$$T(A) = \text{trace}(A) - (n-1)^{\frac{1}{2}} \|A\|_F > 0, \tag{13}$$

then A is positive definite.

In our current application $H(c)$ is an $(n-1)\times(n-1)$ matrix. Observe that if I is the $(n-1)\times(n-1)$ identity matrix, then for $\eta \geq 0$

$$T(\eta I) = \eta\left((n-1) - \sqrt{(n-2)}\sqrt{(n-1)}\right). \tag{14}$$

To avoid small eigenvalues in the Hessian matrix $H(c)$, let

$$\Omega = \left\{c \in \mathbb{R} : T(H(c)) \geq (n-1) - \sqrt{(n-2)}\sqrt{(n-1)}\right\}. \tag{15}$$

The constrained optimization problem that we propose to solve at each iteration of the SQP-Newton method to obtain the penalty parameter c_k is

$$\begin{array}{ll} \textit{minimize} & \phi(c) = \|ch_k\|_2^2 + \mu\|H(c) - I\|_F^2 \\ \textit{subject to} & c \in \Omega. \end{array} \tag{16}$$

Clearly, by the Tarazaga condition (13) any scalar in Ω makes the reduced Hessian matrix, $H(c)$, symmetric and positive definite.

Theorem 4.2: If $\Omega \neq \emptyset$, then the optimization problem (16) has a unique solution.

Proof. The proof follows from the fact that Ω is closed and convex and $\phi(c)$ is continuous and uniformly convex. □

Condition (13) is a sufficient condition for the matrix $H(c)$ to be positive definite. Moreover, for all $c \in \Omega$, $H(c)$ satisfies condition (13). Whether the condition $\Omega \neq \emptyset$ holds, in Theorem 4.2, depends heavily on the structure of the reduced Hessian matrix, $H(c)$, and also on the size of the problem. In fact, if $B(x_k)^T(\nabla^2 f(x_k) + U(x_k))B(x_k) =$

αI for $\alpha \geq 1$ then condition (15) is satisfied, therefore $0 \in \Omega$. Furthermore, we have observed that small variations of diagonal matrices also tend to produce $\Omega \neq \emptyset$. On the other hand, regardless of the structure, for relative small problems, we have noticed that Ω is frequently not empty. Nevertheless, at this point we would like to stress out that we are not claiming that condition (15) is the best condition for large scale problems.

4.1 Algorithm for the Penalty Parameter

The algorithm we present here determines $c_k \in \mathbb{R}$ such that objectives (C1)-(C3) hold. However, solving the optimization subproblem (16) will not always be possible, so we incorporate a standard regularization procedure and ask for c_k and ρ_k such that the reduced Hessian matrix, $H(c_k)+\rho_k I$ is positive definite and well conditioned. The algorithm we propose for obtaining (c_k, ρ_k) at the k^{th} iteration of the SQP-Newton method, is as follows

Algorithm 4.3
Given $\mu > 0$ do the following:
If $0 \in \Omega$
    Set $(c_k, \rho_k) = (0, 0)$
Else
    If the constrained problem (16) has a solution, c_k^*
        Set $c_k = c_k^*$ and $\rho_k = 0$
    Else
        Let c_k^* be the solution of the unconstrained problem:
        $minimize \quad \phi(c) = \|ch_k\|_2^2 + \mu\|H(c) - I\|_F^2$
        $c \in \mathbb{R}$.
        If $c_k^* \in$ SPD
            $\rho_k = 0$
        Else
            Find ρ_k such that $H(0) + \rho_k I \in$ SPD
        End
    End
End

Notice that in Algorithm 4.3 we compute ρ_k only if $0 \notin \Omega$ and $c_k^* \notin \Omega$. Moreover, in this case we take $c_k = 0$ and add $\rho_k I$ to the reduced Hessian matrix $H(0)$. This reasoning comes from the fact that c_k obtained in the first steps of the algorithm is not a satisfactory choice for the penalty parameter. Thus, in this case it makes better sense to correct the matrix $H(0)$ instead of the matrix $H(c_k)$. In order to compute ρ_k we use the modified Cholesky factorization as presented in Dennis and Schnabel [7]. On the other hand, an explicit expression of the penalty parameter, c_k^*, as the solution of the optimization problem (12), considering only one constraint, can be easily obtained using the Lagrange multiplier theory. From the first order necessary conditions associated with problem (16) when $m = 1$ we have

Case 1:

$$c_k = \frac{\mu\{\text{trace}(D) - \text{trace}(AD)\}}{h_k + \mu h_k \text{trace}(D^2)} \tag{17}$$

is a solution of the first order necessary conditions for problem (16) if

$$T(H(c_k)) \geq T(I).$$

Case 2:

c_k satisfying

$$T(H(c_k)) - T(I) = 0 \tag{18}$$

is a solution of the system of first order necessary conditions for problem (16) if the associated Lagrange multiplier, say γ, given by

$$\gamma = \frac{\{2c_k h_k^2 + \mu\{2h_k \text{trace}(AD) + 2c_k h_k^2 \text{trace}(D^2) - 2h_k \text{trace}(D)\}\} \;\; 2R(c_k)}{2R(c_k)\; h_k \text{trace}(D) - \sqrt{(n-2)} \;\; W(c_k)},$$

is positive, where

$$W(c) = 2h_k \text{trace}(AD) + 2ch_k^2 \text{trace}(D^2)$$

and,

$$R(c) = \sqrt{\text{trace}(A^2) + 2ch_k \text{trace}(AD) + (c)^2 h_k^2 \text{trace}(D^2)}.$$

Our next result characterizes the solutions of the equation (18).

Lemma 4.1: Let

$$\hat{T}(H(c)) = (\text{trace}(A) + ch_k \text{trace}(D) - T(I))^2 - (n-2)R(c)^2.$$

Then, any c satisfying $\hat{T}(H(c)) = 0$ such that

$$(\text{trace}(A) + ch_k \text{trace}(D) - T(I)) + \sqrt{(n-2)} \;\; R(c) \neq 0$$

is a solution of equation (18).

Proof. Let us write

$$T(H(c)) - T(I) = \frac{V_1(c)V_2(c)}{V_2(c)} = \frac{\hat{T}(H(c))}{V_2(c)} \tag{19}$$

where,

$$V_1(c) = (\text{trace}(A) + ch_k\text{trace}(D) - T(I)) - \sqrt{(n-2)}\ R(c)$$
$$V_2(c) = (\text{trace}(A) + ch_k\text{trace}(D) - T(I)) + \sqrt{(n-2)}\ R(c).$$

From equation (19), if c satisfies $\hat{T}(H(c)) = 0$ with $V_2(c) \neq 0$ then,

$$T(H(c)) - T(I) = 0.$$

□

5 The $m > 1$ Constraint Case

In this section we extend Algorithm 4.3 to the case where the number of constraints m is greater than one. We present two distinct algorithms for determining a penalty vector c_k in the Lagrange multiplier formula (10) keeping objectives (C1), (C2) and (C3) in mind.

In order to introduce the first algorithm we write the expression of the reduced Hessian matrix of the Lagrangian function $H(c)$ when the multiplier is given by (10) as:

$$H(c) = A + \sum_{i=1}^{m} U_k^i D^i + \sum_{i=1}^{m} c_i h_i D^i, \qquad (20)$$

where,

$$h_i = h_i(x_k)$$
$$U_k^i = U_i(x_k)$$
$$A = B(x_k)^T \nabla_x^2 f(x_k) B(x_k)$$
$$D^i = B(x_k)^T \nabla_x^2 h_i(x_k) B(x_k).$$

On the other hand, we can write (20) in the following way,

$$H(c) = \frac{1}{m}\{H_1(c_1) + H_2(c_2) + ... + H_m(c_m)\} \qquad (21)$$

where,

$$A_i = A + mU_k^i D^i \quad \text{and} \quad H_i(c) = A_i + mch_i D^i \quad \text{for } i = 1, ..., m.$$

Let us define the set Ω for the case $m \geq 1$ as

$$\Omega = \{c \in \mathbb{R}^m : T(H(c)) \geq T(I)\}, \qquad (22)$$

and we also define for $i = 1, ..., m$

$$\Omega_i = \{c_i \in \mathbb{R}^m : T(H_i(c_i)) \geq T(I)\}, \qquad (23)$$

where T was defined in Section 4 by (13). Now, we can state an extension of Algorithm 4.3, for computing the vector (c_k, ρ_k), when $m > 1$.

Algorithm 5.1 (Parallel)
Given $\mu > 0$ do the following:
If $0 \in \Omega$
  Set $(c_k, \rho_k) = (0, 0)$
Else
  For $i = 1, \ldots, m$ solve (in parallel) the constrained problems for c_i
    $(\text{Const})_{\text{i}} \equiv \quad \min_{c_i \in \Omega_i} \phi_i(c_i) = \|c_i h_i\|_2^2 + \mu \|H_i(c_i) - I\|_F^2$
    If problem $(Const)_i$ has a solution, c_i
      Set $c_i^* = c_i$
    Else
      Let c_i^* be the solution of unconstrained problem
      $(\text{Unconst})_{\text{i}} \equiv \quad \min \phi_i(c_i) = \|c_i h_i\|_2^2 + \mu \|H_i(c_i) - I\|_F^2$
    End
  End
  Set $c^* = (\frac{c_1^*}{m}, \frac{c_2^*}{m}, .., \frac{c_m^*}{m})$
  If $c^* \in \Omega$
    Set $(c_k, \rho_k) = (c^*, 0)$
  Else
    Set $(c_k, \rho_k) = (0, \rho_k)$ where ρ_k is computed such that
    $H(0) + \rho_k I \in$ SPD
  End
End

Notice that Algorithm 5.1 is actually a parallel version of Algorithm 4.3, given in Section 4, since it is possible to solve each problem $(\text{Const})_i$ for $i = 1, \ldots, m$ independently.

In order to state another version of Algorithm 4.3 we write the reduced Hessian matrix of the Lagrangian function, at iteration k, of the SQP-Newton method as:

$$\begin{aligned} H(c) &= A_0 + (\frac{1}{m}A + \frac{1}{m}A + ... + \frac{1}{m}A) + \sum_{i=2}^{m} U_k^i D^i + \sum_{i=1}^{m} c_i h_i D^i \\ &= A_{m-2} + \frac{1}{m}A + U_k^m D^m + \sum_{i=m-1}^{m} c_i h_i D^i \\ &= A_{m-1} + c_m h_m D^m \end{aligned} \tag{24}$$

where,

$$A_0 = \frac{1}{m}A + U_k^1 D^1$$

$$A_i = \frac{1}{m}A + A_{i-1} + U_k^{i+1}D^{i+1} + c_k^i h_i D^i \quad i = 1, ..., m-1$$

Let us denote, for $i = 1, ..., m$

$$H_i(c) = A_{i-1} + ch_iD^i \quad \text{and} \tag{25}$$

$$\Omega_i = \{c \in \mathbb{R}^m : T(H_i(c)) \geq \frac{i}{m}T(I)\}, \tag{26}$$

where T is given by (13).

We will use the expression (24) for $H(c_k)$ and the notation (22), (25) and (26) to state another extension of Algorithm 4.3 when $m > 1$.

Algorithm 5.2 (Sequential)
For a given value of the constant $\mu > 0$ do the following:
If $0 \in \Omega$
 Set $(c_k, \rho_k) = (0, 0)$
Else
 For $i = 1, \dots, m$ solve the constrained problems for c_i
  $(\text{Const})_\text{i} \equiv \quad \min \phi_i(c_i) = \|c_ih_i\|_2^2 + \mu\|H_i(c_i) - \frac{i}{m}I\|_F^2$
   $c_i \in \Omega_i$
  If problem $(Const)_i$ have a solution, c_i
   Set $c_i^* = c_i$
  Else
   Let c_i^* be the solution of unconstrained problem
   $(\text{Unconst})_\text{i} \equiv \quad \min \phi_i(c_i) = \|c_ih_i\|_2^2 + \mu\|H_i(c_i) - \frac{i}{m}I\|_F^2$
  End
  Set $A_i = \frac{1}{m}A + A_{i-1} + c_i^*h_iD^i$
 End
 Set $c^* = (c_1^*, c_2^*, .., c_m^*)$
 If $c^* \in \Omega$
  Set $(c_k, \rho_k) = (c^*, 0)$
 Else
  Set $(c_k, \rho_k) = (0, \rho_k)$ where ρ_k is computed such that
  $H(0) + \rho_kI \in SPD$
 End
End

Algorithm 5.2 is a sequential version of Algorithm 4.3, since the solution of problem $(\text{Const})_{i+1}$ depends on the solution of problem $(\text{Const})_i$ for $i = 1, \dots, m-1$.

6 Local and q-Quadratic Convergence

In this section we study the local convergence properties of the SQP-Newton method with the Lagrange multiplier estimate given by formula (10). We present our convergence analysis in terms of a generic choice for the penalty parameter c in formula (10). Towards this end let us begin with the following definition.

Definition *Let x_* be a stationary point of problem (1) and consider the penalty choice function $c : \mathbb{R}^n \to \mathbb{R}^m$. We say that c is locally bounded at x_* if there exists $N(x_*)$, a neighborhood of x_*, such that $c(N(x_*))$ is a bounded subset of $\mathbb{R}^m$*

In addition to the standard Newton's method assumptions (A1)-(A3), our convergence theory will require the boundedness assumption:

(A4) The penalty choice function c is locally bounded at x_*.

In order to demonstrate the local and q-quadratic convergence of the SQP-Newton method for this new choice of the Lagrange multiplier, we write the first order necessary conditions (8) as a perturbation of Newton's method

$$x_+ = x - (\nabla_x^2 L(x, \lambda_*, \rho) + \hat{J}(x))^{-1}(\nabla_x L(x, \lambda_*, \rho) + \hat{F}(x))$$

where

$$\begin{aligned} \hat{J}(x) &= \nabla^2 h(x)(U(x) - \lambda_* + (c - \rho)h(x)) \\ \hat{F}(x) &= \nabla h(x)\Lambda(x)^{-1}(h(x) - \nabla h(x)^T B(x)^{-1} \nabla_x \ell(x, \lambda_* + \rho h(x))) \\ B(x) &= \nabla_x^2 \ell(x, U(x) + ch(x)) + \rho \nabla h(x) \nabla h(x)^T \\ \Lambda(x) &= \nabla h(x)^T B(x)^{-1} \nabla h(x) \quad \text{and,} \end{aligned}$$

λ_* is the multiplier associated with the solution x_* of problem (1) and ρ is a positive constant such that the matrix $\nabla_x^2 L(x_*, \lambda_*, \rho)$ is positive definite. Therefore, the local and q-quadratic convergence of the SQP-Newton method with the proposed Lagrange multiplier estimate follows inmediately from the following theorem and it is stated in the next corollary.

Theorem 6.1: Let x_* be a local solution of problem (1) with associated Lagrange multiplier λ_*. Assume the standard conditions (A1)-(A3) and the boundedness condition (A4). Also assume that in a neighborhood of the solution x_*

$$\|U(x) - \lambda_*\| = O(\|x - x_*\|). \tag{27}$$

Then the SQP-Newton method with the choice of Lagrange multiplier given by (10) is locally and q-quadratically convergent to x_*.

The proof of Theorem 6.1 can be found in [5], [6].

Corollary 6.1 *Let x_* be a local solution of problem (1) with associated Lagrange multiplier λ_*. Assume the standard assumptions (A1)-(A3). The SQP-Newton method* *with Lagrange multiplier estimate given by (10), where the approximation formula U is the least-squares formula*

$$U(x) = -(\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x), \tag{28}$$

or the Miele-Cragg-Levy formula

$$U(x) = (\nabla h(x)^T \nabla h(x))^{-1} (h(x) - \nabla h(x)^T \nabla f(x)), \tag{29}$$

with the penalty parameter c chosen according to Algorithm 5.1 or Algorithm 5.2, is locally and q-quadratically convergent to x_*.

Proof. The penalty vectors c obtained by Algorithm 5.1 or Algorithm 5.2 satisfy the boundedness condition (A4). It is straightforward to prove that under the standard assumptions (A1)-(A3), there exists a neighborhood of the local solution x_* such that

$$\|U(x) - \lambda_*\| = O(\|x - x_*\|), \tag{30}$$

where the approximation formula U is given by formula (28) or by formula (29). □

A rather direct extension of these results shows that the SQP-Newton method with Lagrange multiplier formula given by (10), where formula U is given by the QP multiplier formula (the multiplier associated with the solution of subproblem (5)) is locally and q-quadratically convergent in the pair (x, λ); but not necessarily in x alone.

7 Numerical Results

In this section we discuss some issues concerning the implementation of the SQP-Newton method and present numerical results obtained from our implementation of the method.

The SQP-Newton method that we implemented to test our multiplier estimate is

$$x_{k+1} = x_k + \Delta x_k \tag{31}$$

where Δx_k is the solution of the linear system

$$\begin{pmatrix} \nabla_x^2 \ell(x_k, U(x_k) + C_k h(x_k)) & \nabla h(x_k) \\ \nabla h(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x \ell(x_k, U(x_k)) \\ -h(x_k) \end{pmatrix}. \tag{32}$$

The penalty vector $C_k = \text{diag}(c_k)$ is computed using Algorithm 5.1 or Algorithm 5.2. We also imposed the condition

$$\|c_k\|_2 \leq M, \tag{33}$$

for some positive constant M. We proceed in the following way: whenever c_k satisfies (33), it is acceptable. Otherwise, we set $c_k = 0$, and perform the modified Cholesky factorization, given in Dennis and Schnabel [7], on the matrix $H(0)$ to obtain ρ_k.

The problems tested were taken from Hock and Schittkowski [11] and will be referenced by the numbers given there. The SQP-Newton method with the choice of the multiplier (10) was implemented in Matlab 4.0 on a Sparc station 1. The choices for U in formula (10) are:

$$U_{LS}(x) = -(\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x) \tag{34}$$

$$U_{MCL}(x) = (\nabla h(x)^T \nabla h(x))^{-1} (h(x) - \nabla h(x)^T \nabla f(x)) \tag{35}$$

$$U_{QP}(x, \lambda) = (\nabla h(x)^T H \nabla h(x))^{-1} (h(x) - \nabla h(x)^T H \nabla f(x)) \tag{36}$$

where the matrix $H = \nabla_x^2 \ell(x, \lambda)^{-1}$.

In order to study the robustness of the Lagrange multiplier choice (10) in the SQP-Newton framework we tested each problem starting from various initial points x_0. Moreover, in order to make uniform comparisons, in all the experiments we use the same initial Lagrange multiplier

$$\lambda_k^0 = U_{LS}(x_0). \tag{37}$$

The numerical results are reported in Tables 1 through 4. The numbers in the column labeled PN give the number of the problem being tested. The numbers in the column labeled NIP give the number of starting points tested for each particular problem. We choose the initial iterate that appears in Hock and Schittkowski [11] and several other initial iterates that were presented in Williamson [17]. The numbers in the column under the label $\lambda(x_k) = U(x_k) + C_k h(x_k)$, give the number of different starting points for which the algorithm converged, for different multiplier estimates, i.e., U_{LS}, U_{MCL} and U_{QP} (labeled LS, MCL and QP respectively). Moreover, in the column labeled $c = c_k$, $\rho = 0$, $\mu = 500$ and $Seq.$ appears the number of starting points for which the SQP-Newton method, given by (31) with $\rho = 0$ and $\mu = 500$ in Algorithm 5.2, the sequential version of Algorithm 4.3, converged for different multipliers estimates. In this case $\rho = 0$ means that we did not add any diagonal matrix to the matrix $H(0)$ even when the matrix $H(c_k)$ was not positive definite and we obtain the iterate by solving the extended system (32). Moreover, $\rho = \rho_k$ means that we added to the diagonal of $H(0)$ the matrix $\rho_k I$. The numbers in the column labeled $\frac{it}{\rho}$ give the total number of iterations required to achieve convergence and the total number of times a nonzero for ρ was computed for all the starting points. In all the experiments we set $M = 10^5$.

In Table 1 the stopping criteria employed was either

$$\|(\nabla_x \ell(x_k, U(x_k)), h(x_k))\|_2 \leq 10^{-7},$$

or the number of iterations reached 250. In this table the results presented were obtained allowing the reduced Hessian matrix $H(c_k)$ to be indefinite. This means that when the reduced Hessian matrix $H(c_k)$ is not positive definite and problem (5) may therefore not have a solution, we obtain the iterate instead by solving the extended system (32). In Table 2 we did not allow the reduced Hessian matrix $H(c_k)$ to be indefinite. Therefore, we not only used the previous stopping criteria, we also consider that the algorithm failed if the reduced Hessian matrix $H(c_k)$ had an eigenvalue less than 10^{-6}. In Tables 3 and 4 we did not allow the reduced Hessian matrix $H(c_k)$ to be indefinite. However, in this case, the stopping criteria was the same as in Table 1. For these tables, we followed Algorithm 5.1 or Algorithm 5.2, i.e., we added $\rho_k I$ to the reduced Hessian matrix $H(0)$ when $H(c_k)$ was not positive definite.

We observe from Table 1 that we can achieve convergence to a minimizer x_*, in many problems from different starting points, just by computing the penalty vector c_k proposed in this work. Also Table 2, indicates that this new choice of the multiplier generates a positive definite reduced Hessian matrix more frequently than the traditional multiplier formulas (least-squares multiplier (34), Miele-Cragg-Levy multiplier (35) and the QP multiplier (36)). Moreover, from Tables 3 and 4 we observe that, when the matrix $H(c_k)$ is not positive definite, the number of times we compute ρ_k for our choice of the Lagrange multiplier, compared with the number of times we compute ρ_k for the traditional multipliers is smaller. Finally, we can achieve convergence to a minimizer in almost all the problems from most starting points.

It is important to mention that the algorithms do not always converge to the same points. Our numerical experiments seem to indicate that the region of local convergence is larger with our choice of the multipliers, indeed example 56 shows clearly this conclusion. Therefore, we can conclude that our choice of the Lagrange multipliers is robust in the sense that far from the solution we can get convergence in most of the cases. This also implies that the global behavior of the augmented Lagrangian SQP-Newton method can be improved just by choosing the penalty parameter as in Algorithm 5.1 or 5.2. On the other hand, our numerical results seem to indicate that Algorithm 5.1 and Algorithm 5.2 are not very sensitive to the choice of the parameter μ.

The most important contribution of this work is the natural way in which we have defined the Lagrange multipliers as a function of the penalty parameter to satisfy conditions (C1)-(C3). In this work we propose to solve the simple subproblem (12) to obtain the penalty parameter. Subproblem (12) is not necessarily the only simple subproblem we can solve to get the penalty parameter satisfying conditions (C1)-(C3). Moreover, the q-quadratic convergence of the method is preserved with the additional cost of solving subproblem (12). Nevertheless, our numerical results

PN	NIP	$\lambda_k = U(x_k) + C_k h(x_k)$														
		$c = 0$, $\rho = 0$			$c = c_k$, $\rho = 0$, $\mu = 500$, *Parallel*			$c = c_k$, $\rho = 0$, $\mu = 1000$, *Parallel*			$c = c_k$, $\rho = 0$, $\mu = 500$, *Seq.*			$c = c_k$, $\rho = 0$, $\mu = 1000$, *Seq.*		
		LS	MCL	QP	LS	MCL	QP	LS	MCL	QP	LS	MCL	QP	LS	MCL	Q P
6	9	2	3	9	8	8	9	8	7	9	8	8	9	8	7	9
7	4	0	3	0	4	4	4	4	4	4	4	4	4	4	4	4
26	5	2	2	2	5	5	5	5	5	5	5	5	5	5	5	5
27	7	5	7	7	7	7	7	7	7	7	7	7	7	7	7	7
60	4	1	1	2	4	4	4	4	4	4	4	4	4	4	4	4
39	11	0	5	10	9	9	9	9	9	10	7	9	10	7	9	10
40	12	4	1	4	11	9	4	7	9	6	7	7	6	8	6	5
42	12	4	7	3	9	9	9	9	9	9	10	10	10	10	10	10
77	10	7	9	7	10	9	8	10	9	8	10	9	9	10	9	10
78	12	6	7	10	12	8	10	10	8	10	10	11	10	10	10	1 0
79	10	6	7	7	7	8	8	7	8	8	8	9	8	8	9	7
46	8	7	7	7	8	6	6	7	7	6	8	6	7	6	7	7
47	11	5	6	6	9	11	11	10	11	10	10	10	11	11	7	10
56	9	0	0	0	0	3	3	1	2	3	7	4	8	7	7	8

Tab. 1: Number of starting points for which the SQP-Newton method converges (indefinite reduced Hessian allowed).

PN	NIP	$\lambda_k = U(x_k) + C_k h(x_k)$														
		$c = 0$, $\rho = 0$			$c = c_k$, $\rho = 0$, $\mu = 500$, *Parallel*			$c = c_k$, $\rho = 0$, $\mu = 1000$, *Parallel*			$c = c_k$, $\rho = 0$, $\mu = 500$, *Seq.*			$c = c_k$, $\rho = 0$, $\mu = 1000$, *Seq.*		
		LS	MCL	QP	LS	MCL	QP	LS	MCL	QP	LS	MCL	QP	LS	MCL	Q P
6	9	2	3	6	7	7	8	7	6	8	7	7	8	7	6	8
7	4	0	0	0	4	4	4	4	4	4	4	4	4	4	4	4
26	5	2	2	2	4	4	5	4	0	4	4	4	5	4	0	4
27	7	1	5	1	7	7	5	7	7	5	7	7	5	7	7	5
60	4	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4
39	11	0	3	0	5	6	6	6	6	6	8	7	6	7	6	6
40	12	2	3	0	3	2	1	2	2	1	3	2	1	2	3	1
42	12	3	3	4	10	10	9	10	9	9	10	10	10	10	10	10
77	10	4	4	3	4	3	3	4	3	3	3	3	3	3	2	3
78	12	6	6	4	7	7	5	7	7	5	7	7	4	7	7	4
79	10	5	5	4	5	4	4	5	4	4	5	5	5	6	5	5
46	8	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
47	11	2	2	2	10	4	5	8	4	5	2	5	5	4	4	5
56	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tab. 2: Number of starting points for which the SQP-Newton method converges (indefinite reduced Hessian is not allowed).

PN	NIP	$\lambda_k = U_{LS}(x_k) + C_k h(x_k)$									
		$c = 0$ $\rho = \rho_k$	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 500$ Parallel	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 1000$ Parallel	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 500$ Seq.	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 1000$ Seq.	$\frac{it}{\rho}$
6	9	3	$\frac{24}{11}$	9	$\frac{392}{6}$	9	$\frac{544}{44}$	9	$\frac{392}{6}$	9	$\frac{544}{44}$
7	4	0	$\frac{0}{0}$	4	$\frac{53}{0}$	4	$\frac{53}{0}$	4	$\frac{53}{0}$	4	$\frac{53}{0}$
26	5	2	$\frac{110}{12}$	5	$\frac{145}{3}$	5	$\frac{145}{3}$	5	$\frac{145}{3}$	5	$\frac{145}{3}$
27	7	7	$\frac{173}{46}$	7	$\frac{149}{0}$	7	$\frac{147}{0}$	7	$\frac{149}{0}$	7	$\frac{147}{0}$
60	4	4	$\frac{59}{5}$	4	$\frac{62}{0}$	4	$\frac{62}{0}$	4	$\frac{62}{0}$	4	$\frac{62}{0}$
39	11	6	$\frac{430}{290}$	9	$\frac{241}{37}$	9	$\frac{243}{54}$	10	$\frac{301}{34}$	9	$\frac{142}{15}$
40	12	9	$\frac{145}{88}$	9	$\frac{144}{31}$	9	$\frac{121}{42}$	9	$\frac{146}{23}$	6	$\frac{80}{13}$
42	12	12	$\frac{184}{8}$	12	$\frac{114}{2}$	12	$\frac{116}{2}$	12	$\frac{117}{2}$	12	$\frac{117}{2}$
77	10	9	$\frac{219}{34}$	9	$\frac{203}{30}$	9	$\frac{216}{31}$	9	$\frac{220}{47}$	9	$\frac{220}{47}$
78	12	8	$\frac{50}{5}$	10	$\frac{119}{30}$	9	$\frac{88}{25}$	8	$\frac{54}{4}$	9	$\frac{81}{12}$
79	10	10	$\frac{115}{12}$	10	$\frac{121}{12}$	10	$\frac{119}{12}$	10	$\frac{119}{13}$	10	$\frac{109}{10}$
46	8	8	$\frac{268}{39}$	8	$\frac{238}{30}$	8	$\frac{242}{32}$	8	$\frac{200}{31}$	8	$\frac{186}{32}$
47	11	11	$\frac{218}{38}$	11	$\frac{221}{6}$	11	$\frac{216}{11}$	11	$\frac{240}{22}$	11	$\frac{249}{17}$
56	9	1	$\frac{8}{2}$	2	$\frac{35}{20}$	2	$\frac{27}{15}$	1	$\frac{6}{1}$	0	$\frac{0}{0}$

Tab. 3: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).

PN	NIP	$\lambda_k = U_{MCL}(x_k) + C_k h(x_k)$									
		$c = 0$ $\rho = \rho_k$	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 500$ Parallel	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 1000$ Parallel	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 500$ Seq.	$\frac{it}{\rho}$	$c = c_k$ $\rho = \rho_k$ $\mu = 1000$ Seq.	$\frac{it}{\rho}$
6	9	9	$\frac{544}{148}$	9	$\frac{504}{58}$	9	$\frac{504}{42}$	9	$\frac{504}{58}$	9	$\frac{504}{42}$
7	4	4	$\frac{85}{4}$	4	$\frac{52}{0}$	4	$\frac{52}{0}$	4	$\frac{52}{0}$	4	$\frac{52}{0}$
26	5	5	$\frac{114}{12}$	5	$\frac{147}{3}$	5	$\frac{163}{12}$	5	$\frac{147}{3}$	5	$\frac{163}{12}$
27	4	7	$\frac{88}{3}$	7	$\frac{122}{0}$	7	$\frac{132}{0}$	7	$\frac{122}{0}$	7	$\frac{132}{0}$
60	4	4	$\frac{59}{5}$	4	$\frac{63}{0}$	4	$\frac{64}{0}$	4	$\frac{63}{0}$	4	$\frac{64}{0}$
39	11	8	$\frac{142}{34}$	10	$\frac{225}{20}$	9	$\frac{164}{3}$	10	$\frac{216}{22}$	9	$\frac{146}{24}$
40	12	9	$\frac{114}{61}$	5	$\frac{46}{7}$	6	$\frac{62}{8}$	7	$\frac{108}{40}$	7	$\frac{85}{4}$
42	12	12	$\frac{163}{10}$	12	$\frac{128}{3}$	12	$\frac{128}{3}$	12	$\frac{115}{2}$	12	$\frac{115}{2}$
77	10	9	$\frac{418}{245}$	9	$\frac{398}{230}$	8	$\frac{161}{19}$	8	$\frac{152}{23}$	9	$\frac{178}{29}$
78	12	8	$\frac{51}{6}$	10	$\frac{86}{5}$	10	$\frac{86}{5}$	9	$\frac{98}{36}$	9	$\frac{98}{36}$
79	10	10	$\frac{122}{20}$	10	$\frac{119}{16}$	10	$\frac{124}{14}$	10	$\frac{111}{13}$	10	$\frac{120}{12}$
46	8	8	$\frac{326}{53}$	8	$\frac{271}{43}$	8	$\frac{265}{42}$	7	$\frac{217}{50}$	8	$\frac{269}{57}$
47	11	11	$\frac{260}{58}$	11	$\frac{221}{13}$	11	$\frac{224}{16}$	11	$\frac{262}{31}$	11	$\frac{242}{22}$
56	9	6	$\frac{856}{720}$	9	$\frac{277}{200}$	6	$\frac{76}{35}$	1	$\frac{6}{1}$	0	$\frac{0}{0}$

Tab. 4: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).

show the improved global behavior of augmented Lagrangian SQP-Newton method with this choice of the penalty parameter. In the near future we would like to consider other subproblems to get the penalty parameter, and also to embed this technique in a globalization strategy.

Finally, we would like to extend our approach for large scale problems.

Acknowledgment: We would like to thank three anonymous referees for their constructive comments and helpful suggestions.

References

[1] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1976.

[2] Bertsekas, D. P., "On the penalty and multiplier methods for constrained minimization". *SIAM J. on Control and Optimization*, 14 (1976), 216-235.

[3] Boggs P. and Toole J., "Sequential quadratic programming". *Acta Numerica*, 4 (1995), 1-51.

[4] Byrd R. H. and Nocedal J., "A tool for the analysis of Quasi-Newton methods with application to unconstrained minimization". *SIAM J. Numer. Anal.*, 26 (1989), 727-739.

[5] Cores D., *A Robust Choice of the Lagrange Multipliers Estimates in the Successive Quadratic Programming Framework.* Ph.D. Thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1993.

[6] Cores, D. and Tapia R., "A perturbation lemma for the Newton's method with application to the SQP Newton method". To appear in *Journal of Optimization Theory and Applications*, 97(2), (1998).

[7] Dennis J. E. and Schnabel R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, New Jersy, 1983.

[8] Fletcher R., *Practical Methods of Optimization.* John Willey and Sons, New York, 1987.

[9] Fletcher R., "A new variational result for Quasi-Newton formulae". *SIAM J. on Optimization*, 1 (1991), 18-21.

[10] Haarhoff P. C. and Buys J. D., "A new method for the optimization of a nonlinear function subject to nonlinear constraints". *Computer Journal*, 13 (1970), 178-184.

[11] Hock W. and Schittkowski K., *Test Examples for Nonlinear Programming Codes.* Springer Verlag, New York, 1981.

[12] Miele A., Cragg E. E. and Levy A. V., "On the method of multipliers for mathematical programming problems, part 2". *Journal of Optimization Theory and Applications*, 10 (1972), 1-33.

[13] Tapia R. A., "On secant updates for use in general constrained optimization". *Mathematics of Computation*, 51 (1988), 181-202.

[14] Tapia R. A., "Diagonalized multiplier methods and Quasi-Newton methods for constrained optimization". *Journal of Optimization Theory and Applications*, 22 (1977), 135-194.

[15] Tarazaga P., "Eigenvalue estimates for symmetric matrices". *Linear Algebra and its Applications*, 135 (1990), 171-179.

[16] Tarazaga P., "More estimates for eigenvalues and singular values". *Linear Algebra and its Applications*, 149 (1991), 97-110.

[17] Williamson K. A., *A Robust Trust Region Algorithm for Nonlinear Programming*. Ph.D. Thesis, Department of Mathematical Sciences, Rice University, Houston, Texas, 1990.