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

Debora Cores
R.A. Tapia

October 1994

TR94-37
A Robust Choice of the Lagrange Multiplier in the SQP Newton Method *

Debora Cores† Richard A. Tapia‡

September 1994

Abstract
We study the choice of the Lagrange multipliers in the successive quadratic programming method (SQP) for equality constrained optimization.

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. Considerable numerical experimentation is included and shows that our approach merits further investigation.

---

*This research represents a portion of the first author doctoral thesis under the supervision of the second author in the Department of Computational and Applied Mathematics, Rice University, Houston, Texas

†Currently at INTEVEP (Research Center of the Venezuelan oil Company). This author would like to thank INTEVEP for support during part of her graduate studies

‡Department of Computational and Applied Mathematics and the Center for Research on Parallel Computations, Rice University, Houston, Texas, 77251-1892. This author was supported in part by AFOSR 89-0363, DOE DEFG05-86ER25017, and ARO 9DAAL03-90-G-0093.
Keywords: Augmented Lagrangian, successive quadratic programming, nonlinear programming, quadratic convergence, Lagrange multiplier estimate.

Abbreviated Title: Robust choice of multiplier.

AMS(MOS) subject classifications: 65K,49M,90C

1 Introduction

In this work we study the successive quadratic programming approach (SQP-Newton method) 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). 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. Little success, if any, was obtained in this direction and researchers eventually abandoned the use of the augmented Lagrangian and returned to the standard Lagrangian (choice of zero for the penalty parameter).

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
[9] 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 ma-
terial on the augmented Lagrangian SQP method for equality constrained
optimization and motivate our philosophical approach for picking the penalty
parameter. We begin Section 3 by posing an ideal constrained optimization
problem whose solution gives the desired penalty parameter. The constraint
consists of an inequality involving the Byrd and Nocedal measure function.
Next, with an eye towards computation, we observe that the Byrd-Nocedal
function serves effectively as a barrier function and replace our ideal problem
with an equivalent logarithmic barrier function problem. In Section 4 this
barrier function problem is studied closely for the special case of equality
constrained optimization with only one constraint. From this analysis we
conclude that solving the barrier function problem would not be numerically
efficient. Hence, we replace the barrier function problem with a problem
that preserves some of the flavor, but whose solution can be readily ob-
tained. 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 mul-
tiplier estimate [4], the Miele-Cragg-Levy multiplier [6], and the multiplier
associated with the solution of the quadratic program (see Tapia [10] )) did
not produce iteration sequences which converged. 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.

2 Preliminaries

In this work, we are concerned with the nonlinear equality constrained optimization problem:

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0
\end{align*}
\]  

(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

\[ \ell(x, \lambda) = f(x) + h(x)^T \lambda; \quad \text{and} \]

(2)

the augmented Lagrangian function associated with problem (1) is

\[ L(x, \lambda, c) = f(x) + h(x)^T \lambda + \frac{c}{2} h(x)^T h(x). \]

(3)

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

Throughout this work we will assume that problem (1) has a solution \( x_* \) with associated Lagrange multiplier \( \lambda_* \) 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) \( z^T \nabla^2 \ell(x_*, \lambda_*) z > 0 \) for all \( z \neq 0 \) satisfying \( \nabla h(x_*)^T z = 0 \).
Consider the augmented Lagrangian SQP-Newton method. For a given iterate \((x_k, \lambda_k)\) and penalty parameter \(c_k\) we let

\[
\begin{align*}
    x_{k+1} &= x_k + \Delta x_N \\
    \lambda_{k+1} &= \lambda_k + \Delta \lambda_N
\end{align*}
\]  
(4)

where \(\Delta x_N\) is the solution, and \(\Delta \lambda_N\) is the multiplier associated with the solution, of the quadratic program

\[
\begin{align*}
    \text{minimize} & \quad q(\Delta x) = \nabla_x L(x_k, \lambda_k, c_k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2_x L(x_k, \lambda_k, c_k) \Delta x \\
    \text{subject to} & \quad \nabla h(x_k)^T \Delta x + h(x_k) = 0.
\end{align*}
\]  
(5)

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

\[
s^T \nabla^2_x L(x_k, \lambda_k, c_k) s > 0 \quad \text{for all} \quad s \neq 0 \quad \text{such that} \quad \nabla h(x_k)^T s = 0.
\]  
(6)

Clearly, if \(\nabla^2_x L(x_k, \lambda_k, c_k)\) does not satisfy these sufficiency conditions, then problem (5) may not have a solution; however, even if \(\nabla^2_x L(x_k, \lambda_k, c_k)\) 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

\[
\begin{pmatrix}
    \nabla^2_x L(x_k, \lambda_k, c_k) & \nabla h(x_k) \\
    \nabla h(x_k)^T & 0
\end{pmatrix}
\begin{pmatrix}
    \Delta x_N \\
    \Delta \lambda_N
\end{pmatrix}
= \begin{pmatrix}
    -\nabla_x L(x_k, \lambda_k, c_k) \\
    -h(x_k)
\end{pmatrix}.
\]  
(8)

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

\[
\begin{pmatrix}
    \nabla^2_x \ell(x_k, \lambda_k + c_k h(x_k)) & \nabla h(x_k) \\
    \nabla h(x_k)^T & 0
\end{pmatrix}
\begin{pmatrix}
    \Delta x_N \\
    \Delta \lambda_N
\end{pmatrix}
= \begin{pmatrix}
    -\nabla_x \ell(x_k, \lambda_k) \\
    -h(x_k)
\end{pmatrix}.
\]  
(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 $c_k$ only through the multiplier estimate in the Hessian matrix $\nabla^2 \ell(x_k, \lambda_k + c_k h(x_k))$.

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

$$\hat{\lambda}_k = U(x_k, \lambda_k) + C_k h(x_k)$$

where $c_k \in \mathbb{R}^m$, $C_k = \text{diag}(c_k)$, and $U(x_k, \lambda_k)$ 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 $\nabla^2 \ell(x_k, \hat{\lambda}_k)$ is positive definite on the subspace $\{s \in \mathbb{R}^n : \nabla h(x_k)^T s = 0\}$.

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

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

We hope 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 $(x_k, \lambda_k)$ the $k^{th}$ iterate of the Lagrangian SQP-Newton method for problem (1). Let $U(x, \lambda)$ be a Lagrange multiplier formula, i.e., $\lambda_\ast = U(x_\ast, \lambda_\ast)$ when $(x_\ast, \lambda_\ast)$ 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 $c \in \mathbb{R}^m$ and $\rho \in \mathbb{R}$ we let

$$H(c) := B(x_k)^T \nabla^2 \ell(x_k, U(x_k, \lambda_k) + C h(x_k)) B(x_k)$$

(11)
and

\[ H(c, \rho) := H(c) + \rho I \]  \hspace{1cm} (12)

As before, \( C \) denotes the diagonal matrix \( \text{diag}(c) \).

The task at hand is to determine \( c_k \in \mathbb{R}^m \) so that the objectives (C1)-(C3) of Section 2 hold with \( \lambda_k \) given by (10). Hence, we look for \( c_k \) such that \( H(c_k) \) is positive definite and well conditioned. However, in general we will not be able to attain this goal, so we incorporate a standard regularization procedure and ask for \( c_k \) and \( \rho_k \) such that \( H(c_k, \rho_k) \) is positive definite and well conditioned. We then calculate our correction \((\Delta x, \Delta \lambda)\) to \((x_k, \lambda_k)\) as the solution of the linear system

\[
\begin{pmatrix}
D_k(c_k, \rho_k) & \nabla h(x_k) \\
\nabla h(x_k)^T & 0 \\
\end{pmatrix}
\begin{pmatrix}
\Delta x \\
\Delta \lambda \\
\end{pmatrix}
= 
\begin{pmatrix}
-\nabla_x \ell(x_k, U(x_k, \lambda_k)) \\
-h(x_k) \\
\end{pmatrix}
\]  \hspace{1cm} (13)

with \( D_k(c_k, \rho_k) = \nabla^2 \ell(x_k, U(x_k, \lambda_k) + C_k h(x_k)) + \rho_k I \).

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

\[
\begin{align*}
\text{minimize} & \quad f(c, \rho) = \|((Ch(x_k))^T, \rho)\|^2_2 \\
\text{subject to} & \quad H(c, \rho) \in S
\end{align*}
\]  \hspace{1cm} (14)

where \( S \) is an appropriate set contained in the set of symmetric and positive definite matrices (SPD). 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 \( D_k \) in (13) or problem (14). For this reason we will always choose this component of \( c_k \) to be zero. Therefore, problem (14) will be only to select components of the penalty parameter corresponding to nonzero components of \( h(x_k) \). Hence, without any loss of generalization in the analysis of problems of the form (14) we will assume that the vector \( h(x_k) \) has no zero components.

In problem (14) we expect to define the set \( S \) in terms of the properties of the eigenvalues of \( H(c, \rho) \). Hence it is most satisfying that the eigenvalues of \( H(c, \rho) \) are invariant with respect to the choice of the matrix \( B(x_k) \) used in defining the regularized reduced Hessian \( H(c, \rho) \). 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 $B_1^T A B_1$ and $B_2^T A B_2$ have the same eigenvalues.

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

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

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 $B_1^T A B_1 (Q x_i) = \lambda_i Q x_i$ which implies, $\lambda_i$ is an eigenvalue of the matrix $B_1^T A B_1$ with corresponding eigenvector $Q x_i$ and establishes the theorem. \qed

Recently, Byrd and Nocedal [1] introduced a measure function which has the flavor of the condition of a matrix $A \in \text{SPD}$. This function was also studied by Fletcher [3]. It is defined as:

$$\Psi_{BN}(A) = \frac{1}{n}(\text{trace}(A) - \ln(\det(A))), \quad \text{for } A \in \text{SPD}. \quad (15)$$

The Byrd-Nocedal measure function (15) satisfies the following properties:

(P1) $\Psi_{BN}(A)$ is a continuous and strictly convex function on the set SPD (see [1] and [3]).

(P2) $\Psi_{BN}(A)$ is globally and uniquely minimized by $A = I$ over the set SPD (see [3]).

(P3) $\Psi_{BN}(A) \geq 1, \quad \text{for any } A \in \text{SPD} \text{ (see [1]).}$

(P4) $\Psi_{BN}(A) > \ln(\text{Cond}(A)), \quad \text{for any } A \in \text{SPD}$, where $\text{Cond}(A)$ denotes the 2-norm condition number of the matrix $A$. (see [1]).
In some sense, $\Psi_{BN}(A)$ can be considered to be a measure of the closeness between $A$ and the identity matrix.

Our ideal problem, whose solution gives $(c_k, \rho_k)$, to be used in $D_k(c_k, \rho_k)$ in the linear system (13) is

$$\begin{align*}
\text{minimize} & \quad f(c, \rho) = \|(Ch(x_k))^T, \rho\|_2^2 \\
\text{subject to} & \quad H(c, \rho) \in \text{SPD} \\
& \quad \Psi_{BN}(H(c, \rho)) \leq M
\end{align*}$$

(16)

for some choice of $M > 1$.

**Theorem 3.2** Problem (16) 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 is nonempty, closed and convex. This proves the theorem. □

The remainder of this work is concerned with finding a numerical approach for solving problem (16). Towards this end, we replace problem (16) with the barrier penalty function problem

$$\begin{align*}
\text{minimize} & \quad f_\mu(c, \rho) = \|(Ch(x_k))^T, \rho\|_2^2 + \mu \Psi_{BN}(H(c, \rho)) \\
\text{subject to} & \quad H(c, \rho) \in \text{SPD}
\end{align*}$$

(17)

for some choice of $\mu > 0$.

**Theorem 3.3**

(i) Problem (17) has a unique solution.

(ii) If $(c^*(\mu), \rho^*(\mu))$ solves problem (17), then $(c^*(\mu), \rho^*(\mu))$ solves problem (16) with

$$M = \Psi_{BN}(H((c^*(\mu), \rho^*(\mu))).$$

**Proof.** In order to prove (i), consider any $(\hat{c}, \hat{\rho})$ such that $H(\hat{c}, \hat{\rho}) \in \text{SPD}$. Let $\hat{S} = \{(c, \rho) \in \mathbb{R}^{n+1} : H(c, \rho) \in \text{SPD} \text{ and } f_\mu(c, \rho) \leq f_\mu(\hat{c}, \hat{\rho})\}$. Now, consider the optimization problem:

$$\begin{align*}
\text{minimize} & \quad f_\mu(c, \rho) = \|(Ch(x_k))^T, \rho\|_2^2 + \mu \Psi_{BN}(H(c, \rho)) \\
\text{subject to} & \quad (c, \rho) \in \hat{S}
\end{align*}$$

(18)
By the continuity, convexity of $f_\mu$ and the fact that the set $\mathcal{S}$ is closed we have that $\hat{\mathcal{S}}$ is closed and convex. On the other hand, $\hat{\mathcal{S}}$ is also bounded since $f_\mu$ has the infinity property on SPD, i.e., $f$ goes to infinity when the norm of of the vector $(c, \rho)$ goes to infinity, for all matrices $H(c, \rho)$ in SPD. Moreover, $f_\mu$ is a strictly convex function on $\hat{\mathcal{S}}$. Thus the optimization problem (18) has a unique solution. Finally, since $\hat{\mathcal{S}}$ is contained in SPD and $f_\mu$ is strictly convex on SPD we have that the optimization problem (17) has also a unique solution. The proof of part (ii) is straightforward.

\[\square\]

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 barrier penalty problem (17), with $\rho_k = 0$. The reason for considering $\rho_k = 0$ is that in this way the structure of the SQP-Newton method is preserved (see systems (8) and (13)). Moreover, we always obtain better results by first setting $\rho_k = 0$ and attempting to find $c_k$ which satisfies our criterion.

Suppose we are at the $k^{th}$ iteration of the SQP-Newton method. Consider the reduced Hessian matrix $H(c)$ for the one constraint case, i.e.,

\[
H(c) = B(x_k)^T \nabla^2 \ell(x_k, U(x_k)) + ch(x_k)B(x_k)
\]

\[
= B(x_k)^T \nabla^2 f(x_k) + U(x_k)\nabla^2 h(x_k)B(x_k)
+ ch(x_k)B(x_k)^T \nabla^2 h(x_k)B(x_k).
\]

We have that all the elements in (19) are known except the value of the penalty parameter $c$; so for the sake of simplicity, we can write (19) as

\[H(c) = A + ch_kD\]

where,

\[
h_k = h(x_k),
\]

\[
A = B(x_k)^T \nabla^2 f(x_k) + U(x_k)\nabla^2 h(x_k)B(x_k) \text{ and,}
\]

\[
D = B(x_k)^T \nabla^2 h(x_k)B(x_k).
\]
First, we characterize the set of minimizers of the Byrd and Nocedal measure function (15) over the set SPD.

**Theorem 4.1** Consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad \Psi_{BN}(A + ch_kD) \\
H(c) & \in \text{SPD}
\end{align*}
\]  

(20)

If \( h_k \neq 0 \), then the solution \( c_k \) of (20) satisfies

\[
\text{trace}(H(c_k)^{-1}D - D) = 0.
\]

**Proof.** Let us differentiate the function \( \Psi_{BN}(H(c)) \) with respect to the penalty parameter \( c \)

\[
\frac{d}{dc} (\Psi_{BN}(H(c))) = \frac{1}{n - 1} \left\{ h_k \text{trace}(D) - \frac{d}{dc} (\ln(\det(H(c)))) \right\}.
\]

(21)

On the other hand, by Lemma 1.4 of Fletcher [3] we have that

\[
\frac{d}{dc} (\Psi_{BN}(H(c))) = \frac{-h_k}{n - 1} \left\{ \text{trace}(H(c)^{-1}D - D) \right\}.
\]

(22)

If \( c_k \) is a solution of problem (20) and \( h_k \neq 0 \) we have that \( c_k \) satisfies the first-order necessary conditions, that is:

\[
\text{trace}(H(c_k)^{-1}D - D) = 0.
\]

(23)

From the above result we observe that if there exists \( c_k \) such that \( H(c_k) = I \), then equation (23) is satisfied and this \( c_k \) is the unique minimizer of problem (20). However, this is not always possible, nor desirable, and it is not trivial to obtain an expression for \( c_k \) from the first-order necessary condition (23). Therefore, we consider substituting \( H(c_k) \) for \( H(c_k)^{-1} \) in (23) to obtain

\[
\text{trace}(H(c_k)D - D) = 0.
\]

(24)

Now, consider the following equation

\[
c_k h_k + \mu \text{trace}(H(c_k)D - D) = 0.
\]

(25)
The first term on the left-hand side of (25) corresponds to the first-order necessary conditions for the minimization of $\|c_k h_k\|_2^2$ and we are assuming $h_k \neq 0$. Our next result shows that there exits a relationship between equation (25) and the following optimization problem:

$$\minimize \phi(c) = \|c h_k\|_2^2 + \mu\|H(c) - I\|_F^2,$$

c $\in \mathbb{R}$  

(26)

**Theorem 4.2** A necessary and sufficient condition for $c_k$ to be a minimizer of problem (26) is that $c_k$ satisfy equation (25).

**Proof.** Observe that $\phi'(c_k) = 0$ is equivalent to

$$2c_k h_k^2 + 2\mu h_k \text{trace}(H(c_k)D - D) = 0.$$

If $h_k \neq 0$, and trace($D^2$) $\neq 0$ then this necessary condition becomes

$$c_k h_k + \mu \text{trace}(H(c_k)D - D) = 0.$$  

(27)

On the other hand, we know that the function $\phi$ is uniformly convex on $\mathbb{R}$. Hence, equation (27) is a necessary and sufficient condition for $c_k$ to be a minimizer of problem (26).

Theorem 4.2, Theorem 4.1, and the fact that the minimizer of the Byrd and Nocedal measure function (15) over the set SPD is the identity matrix, motivate us to replace the barrier penalty function $f_\mu$ given in problem (17) with the following function:

$$\phi(c) = \|c h_k\|_2^2 + \mu\|H(c) - I\|_F^2, \quad (\mu > 0).$$  

(28)

It is clear, that for the minimizer $c_k$ of $\phi$ we can not guarantee that the matrix $H(c_k)$ is positive definite. For this reason, we need to restrict the minimization of $\phi$ to a set contained in SPD. Tarazaga in [11] and [12] presented a sufficiency condition for a matrix $A \in \mathbb{R}^{n \times n}$ to be positive definite. This condition does not require us to know the eigenvalues of the matrix $A$ and it is an easy condition to evaluate. We state this condition as the following theorem (see [12]).

**Theorem 4.3 (Tarazaga)**

*Let $A \in \mathbb{R}^{n \times n}$. If*

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

(29)

*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 $\eta \geq 0$

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

Let

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

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

$$\begin{align*}
\text{minimize} & \quad \phi(c) = \|c\|_2^2 + \mu\|H(c) - I\|_F^2 \\
\text{subject to} & \quad c \in \Omega.
\end{align*}$$

Clearly, by the Tarazaga condition (29) any matrix in $\Omega$, the constraint set of problem (32), is symmetric and positive definite; however $\Omega$ may be empty. An obvious way to enlarge $\Omega$ is to use $T(\eta I)$ for $\eta \in (0,1)$ instead of $T(I)$ in the definition of $\Omega$ given by (31). In the current study we did not investigate this option; it probably merits further investigation.

**Theorem 4.4** If $\Omega \neq \emptyset$, then the optimization problem (32) has a unique solution.

**Proof.** The proof follows from the fact that $\Omega$ is closed and convex and $\phi(c)$ is continuous and uniformly convex. \hfill \Box

### 4.1 Algorithm for the Penalty Parameter

Recall that $\Omega$ is given by (31). The algorithm we propose for obtaining $(c, \rho)$ at the $k^{th}$ iteration of the SQP-Newton method, is as follows

**Algorithm 4.5**

Given $\mu > 0$ do the following:

If $H(0) \in \Omega$

Set $(c_k, \rho_k) = (0, 0)$

Else

Let $c_k^*$ be the solution of the constrained problem (32), if it exists.

Otherwise, let $c_k^*$ be the solution of the unconstrained problem (26).
If $H(c_k^*) \in \Omega$
    Take $(c_k, \rho_k) = (c_k^*, 0)$
Else
    Take $(c_k, \rho_k) = (0, \rho_k)$ where $\rho_k$ is computed such that $H(0) + \rho_k I \in \text{SPD}$
End

Notice that in Algorithm 4.5 we compute $\rho_k$ only if $H(0) \notin \Omega$ and $H(c_k^*) \notin \Omega$. 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 $\rho_k$ we use the modified Cholesky factorization as presented in Dennis and Schnabel [2].

5 An Explicit Expression for the Penalty Parameter

In this section we use Lagrange multiplier theory to find an explicit solution of problem (32) for the case $m = 1$.

The system of first-order necessary conditions for problem (32) is

$$\nabla_{c_k} \ell(c_k, \gamma) = 2c_k h_k^2 + \mu(2h_k \text{trace} AD) + 2c_k h_k^2 \text{trace}(D^2) - 2h_k \text{trace}(D)$$

$$- \gamma \left\{ h_k \text{trace}(D) - \frac{(n - 2)W(c_k)}{2\sqrt{(n - 2)R(c_k)}} \right\}$$

$$= 0$$  \hspace{1cm} (33)

$$T(H(c_k)) \geq T(I)$$  \hspace{1cm} (34)

$$\gamma(T(H(c_k)) - T(I)) = 0$$  \hspace{1cm} (35)

$$\gamma \geq 0$$  \hspace{1cm} (36)

where,

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

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

The first-order necessary conditions will be satisfied if one of the following cases occurs.

Case 1:
If \( \gamma = 0 \), then (35) and (36) are satisfied and thus,
\[ c_k = \frac{\mu \{ \text{trace}(D) - \text{trace}(AD) \}}{h_k + \mu h_k \text{trace}(D^2)} \tag{37} \]
is a solution of the system of first-order necessary conditions ((33), (34), (35) and (36)) if \( T(H(c_k)) \geq T(I) \).

Case 2:
If \( \gamma > 0 \), then by the complementarity equation (35) we have that any \( c_k \) satisfying
\[ T(H(c_k)) - T(I) = 0 \quad \text{and such that,} \]
\[ \gamma \geq 0 \tag{39} \]
is a solution of the system of first order necessary conditions for problem (32) where,
\[ \gamma = \frac{2c_k \mu (2h_k \text{trace}(AD) + 2c_k h_k^2 \text{trace}(D^2) - 2h_k \text{trace}(D))\}}{2R(c_k) h_k \text{trace}(D) - \sqrt{(n-2)} W(c_k)}. \]

Our next result characterize the solutions of the equation (38).

**Lemma 5.1** Let
\[ \hat{T}(H(c)) = (\text{trace}(A) + ch_k \text{trace}(D) - T(I))^2 - (n-2)R(c)^2. \]
Then,
\[ \{c \in \mathbb{R} : T(H(c)) - T(I) = 0\} \subseteq \{c \in \mathbb{R} : \hat{T}(H(c)) = 0\}. \]

**Proof.** Let us write
\[ T(H(c)) - T(I) = \frac{(T(H(c)) - T(I)) (T(H(c)) + T(I))}{T(H(c)) + T(I)} \tag{40} \]
If \( c \) satisfies \( T(H(c)) - T(I) = 0 \) then,

\[
\hat{T}(H(c)) = (T(H(c)) - T(I))(T(H(c)) + T(I)) = 0 \quad \text{and,} \\
T(H(c)) + T(I) \neq 0. \tag{41}
\]

From the above result we have that any zero of the polynomial of degree two, \( \hat{T}(H(c)) \), that satisfies equation (41), is a zero of equation (38).

## 6 The \( m > 1 \) Constraint Case

In this section we extend Algorithm 4.5 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:

\[
H(c) = A + \sum_{i=1}^{m} U_i^i D^i + \sum_{i=1}^{m} c_i h_i D^i, \tag{42}
\]

where,

\[
h_i = h_i(x_k) \\
U_i^i = U_i(x_k) \\
A = B(x_k)^T \nabla^2_x f(x_k) B(x_k) \\
D^i = B(x_k)^T \nabla^2_x h_i(x_k) B(x_k).
\]

On the other hand, we can write (42) in the following way,

\[
H(c) = \frac{1}{m} \{H_1(c_1) + H_2(c_2) + \ldots + H_m(c_m)\} \tag{43}
\]

where,

\[
A_i = A + mU_i^i D^i \quad \text{and} \quad H_i(c) = A_i + m c_i h_i D^i \quad \text{for} \quad i = 1, \ldots, m.
\]
Let us define the set $\Omega$ for the case $m \geq 1$ as
\[
\Omega = \{c \in \mathbb{R}^m : T(H(c)) \geq T(I)\},
\]
and we also define for $i = 1, \ldots, m$
\[
\Omega_i = \{c \in \mathbb{R}^m : T(H_i(c)) \geq T(I)\},
\]
where $T$ was defined in Section 4 by (29). Now, we can state an extension of Algorithm 4.5, for computing the vector $(c_k, \rho_k)$, when $m > 1$.

**Algorithm 6.1** (Parallel)
Given $\mu > 0$ do the following:
If $H(0) \in \Omega$
Set $(c_k, \rho_k) = (0, 0)$
Else
For $i = 1, \ldots, m$ solve the following constrained problems for $c_i$
\[
(\text{Const})_i \equiv \min_{c_i \in \Omega_i} \phi_i(c_i) = \|c_i h_i\|_2^2 + \mu \|H_i(c_i) - I\|_F^2
\]
If $H_i(c_i) \in \Omega_i$
Set $c_i^* = c_i$
Else
Let $c_i^*$ be the solution of unconstrained problem
\[
(\text{Unconst})_i \equiv \min c_i = \|c_i h_i\|_2^2 + \mu \|H_i(c_i) - I\|_F^2
\]
End
End
Set $c^* = (c_1^*, c_2^*, \ldots, c_m^*)$
If $H(c^*) \in \Omega$
Set $(c_k, \rho_k) = (c^*, 0)$
Else
Set $(c_k, \rho_k) = (0, \rho_k)$ where $\rho_k$ is computed such that $H(0) + \rho_k I \in \text{SPD}$
End
End

Notice that Algorithm 6.1 is actually a parallel version of Algorithm 4.5, 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 4.5 we write the reduced Hessian matrix of the Lagrangian function, at iteration \( k \), of the SQP-Newton method as:

\[
H(c) = A_0 + \left( \frac{1}{m} A + \frac{1}{m} A + \ldots + \frac{1}{m} A \right) + \sum_{i=2}^{m} U_k^i D^i + \sum_{i=1}^{m} c_i h_i D^i
\]

\[
= A_{m-2} + \frac{1}{m} A + U_k^m D^m + \sum_{i=m-1}^{m} c_i h_i D^i
\]

\[
= A_{m-1} + c_m h_m D^m
\]

(46)

where,

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

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

Let us denote, for \( i = 1, \ldots, m \)

\[
H_i(c) = A_{i-1} + c h_i D^i \quad \text{and} \quad \Omega_i = \{ c \in \mathbb{R}^m : T(H_i(c)) \geq \frac{i}{m} T(I) \},
\]

(47)

(48)

where \( T \) is given by (29).

We will use the expression (46) for \( H(c_k) \) and the notation (44), (47) and (48) to state another extension of Algorithm 4.5 when \( m > 1 \).

**Algorithm 6.2** (Sequential)

For a given value of the constant \( \mu > 0 \) do the following:

If \( H(0) \in \Omega \)

Set \( (c_k, \rho_k) = (0, 0) \)

Else

For \( i = 1, \ldots, m \) solve the constrained problems for \( c_i \)

\[
(\text{Const})_i \equiv \min \phi_i(c_i) = \| c_i h_i \|_2^2 + \mu \| H_i(c_i) - \frac{i}{m} I \|_F^2
\]

\( c_i \in \Omega_i \)
If $H_i(c_i) \in \Omega_i$
    Set $c_i^* = c_i$
Else
    Let $c_i^*$ be the solution of unconstrained problem
    \[
    (\text{Unconst})_i \equiv \min_{c_i} \phi_i(c_i) = \|c_i h_i\|_2^2 + \mu \|H_i(c_i) - \frac{1}{m} I\|_F^2
    \]
End
Set $A_i = \frac{1}{m} A + A_{i-1} + c_i^* h_i D_i$
End
Set \( c^* = (c_1^*, c_2^*, \ldots, c_m^*) \)
If $H(c^*) \in \Omega$
    Set $(c_k, \rho_k) = (c^*, 0)$
Else
    Set $(c_k, \rho_k) = (0, \rho_k)$ where $\rho_k$ is computed such that $H(0) + \rho_k I \in SPD$
End
End

Algorithm 6.2 is a sequential version of Algorithm 4.5, since the solution of problem $(\text{Const})_{i+1}$ depends on the solution of problem $(\text{Const})_i$ for $i = 1, \ldots, m - 1$.

7 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 \( x_* \).

We demonstrate local convergence and quadratic convergence for the SQP-Newton method using the Lagrange multiplier choice (10) under the standard assumptions (A1)-(A3) and the boundedness assumption (A4).

First, we present a useful general perturbation lemma that resembles Lemma 11.2.2. in Ortega and Rheinboldt [7]. However, our result is more general in the sense that we consider perturbations to both the Jacobian matrix and the right-hand side in the Newton linear system. To begin with consider the nonlinear system

\[
F(x) = 0,
\]

where \( F : \mathbb{R}^n \to \mathbb{R}^n \). We will use the notation \( J(x) = F'(x) \). Now, consider Newton's method applied to (49)

\[
x_+ = x - J(x)^{-1}F(x),
\]

and a perturbation of this method, say

\[
x_+ = x - (J(x) + \hat{J}(x))^{-1}(F(x) + \hat{F}(x)),
\]

where \( \hat{J} : \mathbb{R}^n \to \mathbb{R}^{n \times n} \) and \( \hat{F} : \mathbb{R}^n \to \mathbb{R}^n \). The standard assumptions for Newton's method applied to system (49) are:

(S1) There exists \( x_* \) such that \( F(x_*) = 0 \).

(S2) The function \( F \) is continuously differentiable in an open convex set \( D \) containing \( x_* \), and the Jacobian operator \( J \) is Lipschitz continuous in \( D \).

(S3) The Jacobian matrix \( J(x_*) \) is nonsingular.

Lemma 7.1 (Perturbation Lemma)

Assume the standard assumptions (S1)-(S3). Assume further that in a neighborhood \( \hat{D} \subset D \) of the local solution \( x_* \)

(i) \( \|\hat{J}(x)\| = O(\|x - x_*\|) \),

(ii) \( \|\hat{F}(x)\| = O(\|x - x_*\|^2) \).
Then, the iterative procedure (51) is locally and $q$-quadratically convergent to $x_*$.

**Proof.** From assumption (i) and Theorem 3.6 of Stewart [8] we have

$$x_+ - x_* = x - x_* - (J^{-1}(x) + W(x)J^{-1}(x))(F(x) + \hat{F}(x))$$

where

$$\|W(x)\| \leq \frac{\|J^{-1}(x)\hat{J}(x)\|}{1 - \|J^{-1}(x)\hat{J}(x)\|},$$

for all $x$ in a neighborhood of $x_*$. Hence,

$$\|x_+ - x_*\| \leq \|J^{-1}(x)(F(x) - F(x) - J(x))(x_* - x)\|$$

$$+ \|J^{-1}(x)\|\|\hat{F}(x)\| + \|W(x)\|\|J^{-1}(x)\|\|F(x) - F(x_*)\|$$

$$+ \|W(x)\|\|J^{-1}(x)\|\|\hat{F}(x)\|. \quad (52)$$

By the standard assumptions (S1)-(S3) and assumption (i) we have that there exists a neighborhood of $x_*$, such that for all $x$ in this neighborhood

$$\|W(x)\| = O(\|x - x_*\|). \quad (53)$$

Moreover, from (52), (53) and assumptions (i), (ii) we have that for all $x$ in a neighborhood of $x_*$

$$\|x_+ - x_*\| = O(\|x - x_*\|^2). \quad (54)$$

It follows from (54) that the iterative procedure (51) is locally and quadratically convergent to $x_*$. \hfill \Box

In order to utilize the perturbation lemma we consider Newton’s method on the nonlinear system of equations

$$\nabla_x L(x, \lambda_*, \hat{\rho}) = 0, \quad (55)$$

where $\lambda_*$ is the multiplier associated with the solution $x_*$ of problem (1) and $\hat{\rho}$ is a positive constant such that the matrix $\nabla_x^2 L(x_*, \lambda_*, \hat{\rho})$ is positive definite. It is well known that such constants exist under the assumptions (A1)-(A3). Hence, our ideal algorithm will be

$$x_+ = x - \nabla_x^2 L(x, \lambda_*, \hat{\rho})^{-1}\nabla_x L(x, \lambda_*, \hat{\rho}). \quad (56)$$
Recall that the iterative procedure (56) is locally and \( q \)-quadratically convergent to \( x_\ast \).

The SQP-Newton method with our choice of the Lagrange multiplier formula (10) can be written

\[
x_+ = x - R(x)^{-1} \{ \nabla f(x) + \nabla h(x) \Lambda(x)(h(x) - \nabla h(x)^T R(x)^{-1} \nabla f(x)) \} \\
= x - (\nabla_x^2 L(x, \lambda_\ast, \rho) + Q(x))^{-1}(\nabla_x L(x, \lambda_\ast, \rho) + E(x)) \quad (57)
\]

where

\[
R(x) = \nabla_x^2 L(x, \lambda_\ast, \rho) + Q(x) \\
Q(x) = \nabla_x^2 \ell(x, U(x)) - \nabla_x^2 \ell(x, \lambda_\ast) + Ch(x) \nabla^2 h(x) - \rho h(x) \nabla^2 h(x) \\
\Lambda(x) = (\nabla h(x)^T R(x)^{-1} \nabla h(x))^{-1} \\
E(x) = \nabla h(x) \{ \Lambda(x)(h(x) - \nabla h(x)^T R(x)^{-1} \nabla f(x)) - \lambda_\ast - \rho h(x) \}.
\]

Hence, the iterative procedure (57) has been written as a perturbation of the ideal Newton method given by (56), where the perturbations are given by \( Q(x) \) and \( E(x) \).

Our next result demonstrates that the SQP-Newton method with Lagrange multiplier estimate given by formula (10) satisfies conditions (i) and (ii) of the previous lemma, whenever the approximation formula \( U \) in the Lagrange multiplier (10) satisfies a mild condition.

**Lemma 7.2** Let \( x_\ast \) be a local solution of problem (1) with associated multiplier \( \lambda_\ast \). Assume that in a neighborhood of the solution \( x_\ast \)

\[
||U(x) - \lambda_\ast|| = O(||x - x_\ast||). \quad (58)
\]

Also assume the standard conditions (A1)-(A3) and the boundedness condition (A4). Then there exists a neighborhood \( N(x_\ast) \), of the local solution \( x_\ast \), such that for all \( x \in N(x_\ast) \)

(a) \( ||Q(x)|| = O(||x - x_\ast||) \),

(b) \( ||E(x)|| = O(||x - x_\ast||^2) \).

22
Proof. From the boundedness condition (A4), the standard assumption (A1) and the condition (58) it follows that there exists a neighborhood of the solution \( x_* \), such that for all \( x \) in that neighborhood
\[
\| Q(x) \| \leq \| \nabla_x^2 \ell(x, \lambda_\star) \| + \| e \| \| h(x) - h(x_\star) \| \| \nabla^2 h(x) \| \\
+ \| \lambda \| \| h(x) - h(x_\star) \| \| \nabla^2 h(x) \| \\
= O(\| x - x_\star \|). (59)
\]

Therefore condition (a) holds. For the sake of simplicity we will denote the \( i^{th} \) component of the gradient vector of \( h \), evaluated at \( x_\star, (\nabla h_i(x_\star)) \) by \( \nabla h_i \).

Similar notation will be used for similar quantities. Using this notation we have
\[
\Lambda\{ h - \nabla^T R^{-1} \nabla f \} = \lambda_\star - \lambda h = \\
= \Lambda\{ P(h - h(x_\star)) - \nabla^T R^{-1} (\nabla f + \nabla h \lambda_\star) \\
+ \nabla^T R^{-1} (\nabla f(x_\star) + \nabla h(x_\star) \lambda_\star) \}
\]
\[
= \Lambda \left\{ P \int_0^1 \nabla^T (x_\star + t(x - x_\star))(x - x_\star) dt \\
- \nabla^T R^{-1} \int_0^1 \nabla^2 f(x_\star + t(x - x_\star))(x - x_\star) dt \\
- \nabla^T R^{-1} \sum_{i=1}^m \int_0^1 \nabla^2 h_i(x_\star + t(x - x_\star))(x - x_\star) \lambda_\star dt \right\}
\]
where
\[
P = I_{m \times m} - \hat{\lambda} A^{-1}.
\]

Adding and subtracting some terms conveniently we obtain
\[
\Lambda\{ h - \nabla^T R^{-1} \nabla f \} = \lambda_\star - \lambda h = \\
\Lambda \left\{ P \int_0^1 A^1(t)(x - x_\star) dt - \nabla^T R^{-1} \int_0^1 A^2(t)(x - x_\star) dt \\
- \nabla^T R^{-1} \sum_{i=1}^m \int_0^1 A^3_i(t)(\lambda_\star)(x - x_\star) dt \right\} \\
+ \Lambda Z(x - x_\star) (60)
\]
where
\[
A^1(t) = (\nabla h(x_\star + t(x - x_\star)) - \nabla h(x_\star))^T, \\
A^2(t) = \nabla^2 f(x_\star + t(x - x_\star)) - \nabla^2 f(x_\star), \\
A^3_i(t) = \nabla^2 h_i(x_\star + t(x - x_\star)) - \nabla^2 h_i(x_\star),
\]
and

23
\[ Z = P \nabla h(x_*)^T - \nabla h^T R^{-1} \nabla^2 f(x_*) - \nabla h^T R^{-1} \sum_{i=1}^{m} \nabla^2 h_i(x_*)(\lambda_*)_i. \]  

By assumption (A1) and the fact that 0 ≤ t ≤ 1 we have that in a neighborhood of the solution \( x_* \)

\[
\| A^1(t) \| = O(\| x - x_* \|),
\| A^2(t) \| = O(\| x - x_* \|),
\| A^3(t) \| = O(\| x - x_* \|). \tag{62}
\]

On the other hand, adding and subtracting terms to (61) we have

\[
Z = \nabla h^T R^{-1} \left\{ (\nabla^2 f - \nabla^2 f(x_*)) + \sum_{i=1}^{m} (\nabla^2 h_i - \nabla^2 h_i(x_*))(\lambda_*)_i \right\}
\]

\[
+ \hat{\rho} \nabla h(\nabla h - \nabla h(x_*))^T + \sum_{i=1}^{m} \nabla^2 h_i(U_i - (\lambda_*)_i) \right\}
\]

\[
+ \binv{R} \nabla h(x_*) - \nabla h \right)^T \{ R^{-1} \nabla^2 f + \hat{\rho} R^{-1} \nabla h \nabla h^T
\]

\[
+ R^{-1} \sum_{i=1}^{m} \nabla^2 h_i U_i \} + \nabla h(x_*)^T R^{-1} \sum_{i=1}^{m} c_i \nabla^2 h_i(h_i - h_i(x_*)). \tag{63}
\]

By assumptions (A1)-(A4) and condition (58) we have that in a neighborhood of the solution \( x_* \)

\[
\| Z \| = O(\| x - x_* \|) \tag{64}
\]

From (60), (62) and (64) it follows that there exists a neighborhood of the solution \( x_* \) such that

\[
\| E(x) \| \leq \| \nabla h \| \| h - \nabla h^T R^{-1} \nabla f \| - \lambda_* - \hat{\rho} h \| = O(\| x - x_* \|^2). \tag{65}
\]

**Theorem 7.1** Let \( x_* \) be a local solution of problem (1) with associated Lagrange multiplier \( \lambda_* \). Assume the standard conditions (A1)-(A3) and the boundedness condition (A4). Also assume that in a neighborhood of the solution \( x_* \)

\[
\| U(x) - \lambda_* \| = O(\| x - x_* \|).
\]

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

24
Proof. It is not difficult to prove and well known that the standard assumptions (A1)-(A3) imply the standard assumptions (S1)-(S3) for the nonlinear system of equations (55). On the other hand, by Lemma 7.2 conditions (ii) and (iii) of Lemma 7.1 are satisfied. Therefore, we have the assumptions of Lemma 7.1 and thus the iterative procedure (57) is locally and q-quadratically convergent to \( x_* \). □

**Corollary 7.1** Let \( x_* \) be a local solution of problem (1) with associated Lagrange multiplier \( \lambda_* \). 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

\[
U(x) = - (\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x),
\]

or the Miele-Cragg-Levy formula

\[
U(x) = (\nabla h(x)^T \nabla h(x))^{-1} (h(x) - \nabla h(x)^T \nabla f(x)),
\]

with the penalty parameter \( c \) chosen accordingly to Algorithm 6.1 or Algorithm 6.2, is locally and q-quadratically convergent to \( x_* \).

Proof. The penalty vectors \( c \) obtained by Algorithm 6.1 or Algorithm 6.2 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

\[
\|U(x) - \lambda_*\| = O(\|x - x_*\|),
\]

where the approximation formula \( U \) is given by formula (66) or by formula (67). □

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 \( (x, \lambda) \); but not necessarily in \( x \) alone.

### 8 Numerical Results

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

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

\[ x_{k+1} = x_k + \Delta x_k \]  

(69)

where \( \Delta x_k \) is the solution of the linear system

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

(70)

The penalty vector \( C_k = \text{diag}(c_k) \) is computed using Algorithm 6.1 or Algorithm 6.2. We also imposed the condition

\[ \| c_k \|_2 \leq M, \]  

(71)

for some positive constant \( M \). We proceed in the following way: whenever \( c_k \) satisfies (71), it is acceptable. Otherwise, we set \( c_k = 0 \), and perform the modified Cholesky factorization, given in Dennis and Schnabel [2], on the matrix \( H(0) \) to obtain \( \rho_k \).

The problems tested were taken from Hock and Schittkowski [5] and will be referenced by the number 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:

\[ U_{LS}(x) = - (\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x) \]  

(72)

\[ U_{MCL}(x) = (\nabla h(x)^T \nabla h(x))^{-1} (h(x) - \nabla h(x)^T \nabla f(x)) \]  

(73)

\[ U_{QP}(x, \lambda) = (\nabla h(x)^T H \nabla h(x))^{-1} (h(x) - \nabla h(x)^T H \nabla f(x)) \]  

(74)

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

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

\[ \lambda^0_k = U_{LS}(x_0). \]  

(75)

The numerical results are reported in Tables 1 through 9. 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 [5] and several other initial iterates that were presented in Williamson [13]. 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 example, 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 (69) with \(\rho = 0\) and \(\mu = 500\) in Algorithm 6.2, the sequential version of Algorithm 4.5, converged. 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. 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 \(\frac{\mu}{\rho}\) give the number of iterations required to achieve convergence over the number of times a nonzero for \(\rho\) was computed.

In Tables 1, 2 and 3 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 these tables 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 (70). In Tables 4, 5 and 6 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 \(H(c_k)\) given in (42) had an eigenvalue less than \(10^{-6}\). In Tables 7, 8 and 9 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 Tables 1, 2 and 3. For these tables, we followed Algorithm 6.1 or Algorithm 6.2, 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 Tables 1, 2, 3, 4, 5 and 6 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 Tables 4, 5 and 6 indicate that this new choice of the multiplier generates a positive definite reduced Hessian matrix more frequently than the traditional multiplier formulas (least-squares multiplier (72), Miele-Cragg-Levy multiplier (73) and
\[ \lambda_k = U_L(x_k) + C_L h(x_k) \]

<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>( c = 0 )</th>
<th>( \rho = 0 )</th>
<th>( c = c_k )</th>
<th>( \rho = 0 )</th>
<th>( c = c_k )</th>
<th>( \rho = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>0</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>4</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

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

the QP multiplier \((74)\). Moreover, from Tables 7, 8 and 9 we observe that the number of times we compute \(\rho_k\) compared with the number of times we compute \(\rho_k\) 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. On the other hand, our numerical results seem to indicate that Algorithm 6.1 and Algorithm 6.2 are not very sensitive to the choice of the parameter \(\mu\).

So far, we have considered the local properties of the new choice of multipliers \((10)\). In the near future we would like to embed this technique in a globalization strategy; and consider the global properties of this new choice of the multipliers.
\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{PN} & \textbf{NIP} & \multicolumn{6}{|c|}{\lambda_k = U_{MCL}(x_k) + C_k h(x_k)} \\
\hline
 & & \begin{tabular}{c|c|c|c|c|}
\hline
$c = 0$ & $c = c_k$ & $c = c_k$ & $c = c_k$ & $c = c_k$
\hline
$\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 1000$
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
Parallel & Seq. & Parallel & Seq. & Parallel & Seq.
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
$c = c_k$ & $c = c_k$ & $c = c_k$ & $c = c_k$
\hline
$\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 1000$
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
Parallel & Seq. & Parallel & Seq. & Parallel & Seq.
\hline
\end{tabular} \\
\hline
6 & 9 & 3 & 8 & 8 & 7 & 7 \\
7 & 4 & 3 & 4 & 4 & 4 & 4 \\
26 & 5 & 2 & 5 & 5 & 5 & 5 \\
27 & 7 & 7 & 7 & 7 & 7 & 7 \\
60 & 4 & 1 & 4 & 4 & 4 & 4 \\
39 & 11 & 5 & 9 & 9 & 9 & 9 \\
40 & 12 & 1 & 9 & 7 & 9 & 6 \\
42 & 12 & 7 & 9 & 10 & 9 & 10 \\
77 & 10 & 9 & 9 & 9 & 9 & 9 \\
78 & 12 & 7 & 8 & 11 & 8 & 10 \\
79 & 10 & 7 & 8 & 9 & 8 & 9 \\
46 & 8 & 7 & 6 & 6 & 7 & 7 \\
47 & 11 & 6 & 11 & 10 & 11 & 7 \\
56 & 9 & 0 & 3 & 4 & 2 & 7 \\
\hline
\end{tabular}
\caption{Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian allowed).}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{PN} & \textbf{NIP} & \multicolumn{6}{|c|}{\lambda_k = U_{QP}(x_k) + C_k h(x_k)} \\
\hline
 & & \begin{tabular}{c|c|c|c|c|}
\hline
$c = 0$ & $c = c_k$ & $c = c_k$ & $c = c_k$ & $c = c_k$
\hline
$\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 1000$
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
Parallel & Seq. & Parallel & Seq. & Parallel & Seq.
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
$c = c_k$ & $c = c_k$ & $c = c_k$ & $c = c_k$
\hline
$\rho = 0$ & $\mu = 500$ & $\rho = 0$ & $\mu = 1000$
\hline
\end{tabular} & \begin{tabular}{c|c|c|c|c|}
\hline
Parallel & Seq. & Parallel & Seq. & Parallel & Seq.
\hline
\end{tabular} \\
\hline
6 & 9 & 9 & 9 & 9 & 9 & 9 \\
7 & 4 & 0 & 4 & 4 & 4 & 4 \\
26 & 5 & 2 & 5 & 5 & 5 & 5 \\
27 & 7 & 7 & 7 & 7 & 7 & 7 \\
60 & 4 & 2 & 4 & 4 & 4 & 4 \\
39 & 11 & 10 & 9 & 10 & 10 & 10 \\
40 & 12 & 4 & 4 & 6 & 6 & 5 \\
42 & 12 & 3 & 9 & 10 & 9 & 10 \\
77 & 10 & 7 & 8 & 9 & 8 & 10 \\
78 & 12 & 10 & 10 & 10 & 10 & 10 \\
79 & 10 & 7 & 8 & 8 & 8 & 7 \\
46 & 8 & 7 & 6 & 7 & 6 & 7 \\
47 & 11 & 6 & 11 & 11 & 10 & 10 \\
56 & 9 & 0 & 3 & 4 & 3 & 8 \\
\hline
\end{tabular}
\caption{Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian allowed).}
\end{table}
Table 4: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).

<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>$\lambda_k = U_L x_k + C_k h(x_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = 500$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = 1000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parallel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seq.</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).

<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>$\lambda_k = U_{MCL} x_k + C_k h(x_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = 500$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = 1000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parallel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seq.</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

30
<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>$\lambda_k = U_{QP}(\bar{x}_k, \lambda_k) + C_k A(\bar{x}_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$c = 0$  $c = c_s$  $c = c_s$  $c = c_s$  $c = c_s$  $c = c_s$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\rho = 0$ $\rho = 0$ $\rho = 0$ $\rho = 0$ $\rho = 0$ $\rho = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = 500$ $\mu = 1000$ $\mu = 500$ $\mu = 1000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parallel  Parallel  Seq.  Seq.</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>6  8  8  8  8  8</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0  4  4  4  4  4</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>2  5  5  5  5  5</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>1  5  5  5  5  5</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td>1  4  4  4  4  4</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>0  6  6  6  6  6</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>0  1  1  1  1  1</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>4  9  9  10 10 10</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>3  3  3  3  3  3</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>4  5  5  4  4  4</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>4  4  4  5  5  5</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>0  0  0  0  0  0</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>2  5  5  5  5  5</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>0  0  0  0  0  0</td>
</tr>
</tbody>
</table>

Table 6: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).
\[ \lambda_k = U_{LS}(z_k) + C_k h(x_k) \]

<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>( c = 0 )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \rho = \rho_k )</td>
<td>( \rho = \rho_k )</td>
<td>( \mu = 500 )</td>
<td>( \mu = 1000 )</td>
<td>( \rho = \rho_k )</td>
<td>( \mu = 1000 )</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>3 ( \frac{52}{11} )</td>
<td>9 ( \frac{32}{6} )</td>
<td>9 ( \frac{544}{44} )</td>
<td>9 ( \frac{32}{6} )</td>
<td>9 ( \frac{544}{44} )</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0 ( \frac{0}{0} )</td>
<td>4 ( \frac{53}{0} )</td>
<td>4 ( \frac{53}{0} )</td>
<td>4 ( \frac{53}{0} )</td>
<td>4 ( \frac{53}{0} )</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>2 ( \frac{110}{12} )</td>
<td>5 ( \frac{145}{3} )</td>
<td>5 ( \frac{145}{3} )</td>
<td>5 ( \frac{145}{3} )</td>
<td>5 ( \frac{145}{3} )</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>7 ( \frac{173}{46} )</td>
<td>7 ( \frac{149}{0} )</td>
<td>7 ( \frac{149}{0} )</td>
<td>7 ( \frac{149}{0} )</td>
<td>7 ( \frac{149}{0} )</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>4 ( \frac{52}{8} )</td>
<td>4 ( \frac{52}{0} )</td>
<td>4 ( \frac{62}{0} )</td>
<td>4 ( \frac{62}{0} )</td>
<td>4 ( \frac{62}{0} )</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>6 ( \frac{430}{250} )</td>
<td>9 ( \frac{241}{37} )</td>
<td>9 ( \frac{243}{34} )</td>
<td>10 ( \frac{301}{34} )</td>
<td>9 ( \frac{142}{16} )</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>9 ( \frac{145}{88} )</td>
<td>9 ( \frac{144}{31} )</td>
<td>9 ( \frac{121}{42} )</td>
<td>9 ( \frac{146}{23} )</td>
<td>6 ( \frac{89}{13} )</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>12 ( \frac{184}{8} )</td>
<td>12 ( \frac{114}{2} )</td>
<td>12 ( \frac{116}{2} )</td>
<td>12 ( \frac{117}{2} )</td>
<td>12 ( \frac{117}{2} )</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>9 ( \frac{210}{34} )</td>
<td>9 ( \frac{203}{30} )</td>
<td>9 ( \frac{216}{31} )</td>
<td>9 ( \frac{220}{47} )</td>
<td>9 ( \frac{220}{47} )</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>8 ( \frac{50}{5} )</td>
<td>10 ( \frac{110}{30} )</td>
<td>9 ( \frac{88}{25} )</td>
<td>8 ( \frac{54}{4} )</td>
<td>9 ( \frac{81}{12} )</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>10 ( \frac{115}{12} )</td>
<td>10 ( \frac{121}{12} )</td>
<td>10 ( \frac{110}{13} )</td>
<td>10 ( \frac{110}{13} )</td>
<td>10 ( \frac{110}{13} )</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>8 ( \frac{268}{30} )</td>
<td>8 ( \frac{288}{30} )</td>
<td>8 ( \frac{242}{32} )</td>
<td>8 ( \frac{200}{31} )</td>
<td>8 ( \frac{186}{32} )</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>11 ( \frac{218}{38} )</td>
<td>11 ( \frac{221}{6} )</td>
<td>11 ( \frac{216}{11} )</td>
<td>11 ( \frac{240}{22} )</td>
<td>11 ( \frac{240}{17} )</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>1 ( \frac{8}{2} )</td>
<td>2 ( \frac{38}{20} )</td>
<td>2 ( \frac{27}{6} )</td>
<td>1 ( \frac{5}{1} )</td>
<td>0 ( \frac{5}{0} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).
<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>$c = 0$</th>
<th>$c = c_k$</th>
<th>$c = c_k$</th>
<th>$c = c_k$</th>
<th>$c = c_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\rho = \rho_k$</td>
<td>$\mu = 500$</td>
<td>$\mu = 1000$</td>
<td>$\mu = 500$</td>
<td>$\mu = 1000$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\frac{\mu}{\rho}$</td>
<td>Parallel</td>
<td>Seq.</td>
<td>Parallel</td>
<td>Seq.</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>9</td>
<td>544 148</td>
<td>9</td>
<td>504 58</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4</td>
<td>52 4</td>
<td>4</td>
<td>52 0</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>5</td>
<td>114 3</td>
<td>5</td>
<td>147 3</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>7</td>
<td>88 6</td>
<td>7</td>
<td>122 0</td>
<td>7</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>4</td>
<td>52 6</td>
<td>4</td>
<td>63 0</td>
<td>4</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>8</td>
<td>142 34</td>
<td>10</td>
<td>225 20</td>
<td>9</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>9</td>
<td>114 61</td>
<td>5</td>
<td>46 7</td>
<td>6</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>12</td>
<td>153 10</td>
<td>12</td>
<td>128 3</td>
<td>12</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>9</td>
<td>418 246</td>
<td>9</td>
<td>308 580</td>
<td>8</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>8</td>
<td>51 6</td>
<td>10</td>
<td>86 5</td>
<td>10</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>10</td>
<td>122 20</td>
<td>10</td>
<td>110 10</td>
<td>10</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>8</td>
<td>226 43</td>
<td>8</td>
<td>221 13</td>
<td>8</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>11</td>
<td>260 68</td>
<td>11</td>
<td>221 13</td>
<td>11</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>6</td>
<td>898 726</td>
<td>9</td>
<td>277 266</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 8: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).
\[
\lambda_k = UQP(z_k, \lambda_k) + C_k h(x_k)
\]

<table>
<thead>
<tr>
<th>PN</th>
<th>NIP</th>
<th>( c = 0 )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
<th>( c = c_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \rho = \rho_k )</td>
<td>( \mu = 500 ) Parallel</td>
<td>( \rho = \rho_k )</td>
<td>( \mu = 1000 ) Parallel</td>
<td>( \rho = \rho_k )</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>9 ( \frac{50}{22} )</td>
<td>9 ( \frac{282}{2} )</td>
<td>9 ( \frac{180}{2} )</td>
<td>9 ( \frac{282}{2} )</td>
<td>9 ( \frac{180}{2} )</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4 ( \frac{496}{64} )</td>
<td>4 ( \frac{52}{0} )</td>
<td>4 ( \frac{52}{0} )</td>
<td>4 ( \frac{52}{0} )</td>
<td>4 ( \frac{52}{0} )</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>5 ( \frac{116}{16} )</td>
<td>5 ( \frac{187}{0} )</td>
<td>5 ( \frac{187}{0} )</td>
<td>5 ( \frac{187}{0} )</td>
<td>5 ( \frac{187}{0} )</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>6 ( \frac{61}{6} )</td>
<td>6 ( \frac{130}{0} )</td>
<td>6 ( \frac{130}{0} )</td>
<td>6 ( \frac{130}{0} )</td>
<td>6 ( \frac{130}{0} )</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>4 ( \frac{50}{6} )</td>
<td>4 ( \frac{50}{6} )</td>
<td>4 ( \frac{50}{6} )</td>
<td>4 ( \frac{50}{6} )</td>
<td>4 ( \frac{50}{6} )</td>
</tr>
<tr>
<td>39</td>
<td>11</td>
<td>10 ( \frac{210}{50} )</td>
<td>10 ( \frac{244}{57} )</td>
<td>9 ( \frac{170}{42} )</td>
<td>9 ( \frac{170}{42} )</td>
<td>9 ( \frac{170}{42} )</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>9 ( \frac{145}{68} )</td>
<td>9 ( \frac{144}{31} )</td>
<td>9 ( \frac{121}{42} )</td>
<td>9 ( \frac{146}{23} )</td>
<td>6 ( \frac{80}{13} )</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>9 ( \frac{160}{68} )</td>
<td>9 ( \frac{190}{36} )</td>
<td>7 ( \frac{115}{28} )</td>
<td>6 ( \frac{164}{32} )</td>
<td>8 ( \frac{184}{55} )</td>
</tr>
<tr>
<td>77</td>
<td>10</td>
<td>8 ( \frac{320}{140} )</td>
<td>8 ( \frac{304}{23} )</td>
<td>8 ( \frac{210}{66} )</td>
<td>9 ( \frac{331}{23} )</td>
<td>9 ( \frac{253}{65} )</td>
</tr>
<tr>
<td>78</td>
<td>12</td>
<td>9 ( \frac{204}{216} )</td>
<td>10 ( \frac{140}{63} )</td>
<td>10 ( \frac{184}{61} )</td>
<td>11 ( \frac{187}{58} )</td>
<td>11 ( \frac{253}{62} )</td>
</tr>
<tr>
<td>79</td>
<td>10</td>
<td>10 ( \frac{145}{65} )</td>
<td>10 ( \frac{230}{54} )</td>
<td>10 ( \frac{153}{22} )</td>
<td>10 ( \frac{123}{14} )</td>
<td>10 ( \frac{118}{9} )</td>
</tr>
<tr>
<td>46</td>
<td>8</td>
<td>8 ( \frac{224}{66} )</td>
<td>8 ( \frac{224}{66} )</td>
<td>8 ( \frac{316}{75} )</td>
<td>8 ( \frac{224}{30} )</td>
<td>8 ( \frac{227}{40} )</td>
</tr>
<tr>
<td>47</td>
<td>11</td>
<td>11 ( \frac{211}{36} )</td>
<td>11 ( \frac{220}{19} )</td>
<td>11 ( \frac{204}{12} )</td>
<td>11 ( \frac{218}{18} )</td>
<td>11 ( \frac{236}{21} )</td>
</tr>
<tr>
<td>56</td>
<td>9</td>
<td>0 ( \frac{0}{0} )</td>
<td>2 ( \frac{109}{85} )</td>
<td>4 ( \frac{818}{596} )</td>
<td>3 ( \frac{26}{7} )</td>
<td>0 ( \frac{0}{0} )</td>
</tr>
</tbody>
</table>

Table 9: Number of starting points for which the SQP-Newton method converges, (indefinite reduced Hessian not allowed).
References


35