A Globally Convergent Primal-Dual Interior Point Method for Constrained Optimization

Hiroshi Yamashita*

Abstract

This paper proposes a primal-dual interior point method for solving general nonlinearly constrained optimization problems. The method is based on solving the barrier Karush-Kuhn-Tucker conditions for optimality by the Newton method. To globalize the iteration we introduce the barrier-penalty fucntion and the optimality condition for minimizing this function. Our basic iteration is the Newton iteration for solving the optimality conditions with respect to the barrier-penalty function which coincides with the Newton iteration for the barrier Karush-Kuhn-Tucker conditions if the penalty parameter is sufficiently large. It is proved that the method is globally convergent from an arbitrary initial point that strictly satisfies the bounds on the variables. Implementations of the given algorithm are done for small dense nonlinear programs . The method solves all the problems in Hock and Schittkowski's textbook efficiently. Thus it is shown that the method given in this paper possesses a good theoretical convergence property and is efficient in practice.

Key words: interior point method, primal-dual method, constrained optimization, nonlinear programming

*Mathematical Systems, Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo, Japan hy@msi.co.jp

1 Introduction

In this paper we propose a primal-dual interior point method that solves general nonlinearly constrained optimization problems. To obtain a fast algorithm for nonlinear optimization problems it is fairly clear from various experiences (for example the well known success of the SQP method) that we should eventually solve the Karush-Kuhn-Tucker conditions for optimality by a Newton-like method. However, solving the optimality conditions simply as a system of equations does not give an algorithm for solving optimization problems in general except for convex problems. Therefore it is not appropriate to treat the primal and dual variables equally to obtain globally convergent methods for general nonlinear optimization problems. Observing that there is a nice relation between the parameterized optimality conditions (barrier KKT condition) and the classical logarithmic barrier function of the primal variables, we can develop a globally convergent method for general nonlinear optimization problems. To prevent the possible divergence of the dual variables, we introduce the barrier penalty function and solve the optimality conditions by a Newton-like method.

For nonlinear programs it has been believed long time that interior point methods are not practical because of inevitable numerical difficulties which occur at the final stage of iterations. See Fiacco and McCormick [4], Fletcher [5] and other standard textbooks on nonlinear optimization. Thus the state of the art method for nonlinear programs today is the SQP method (see Powell [7], Fletcher [5]) which can also be interpreted as a Newton-like method for Karush-Kuhn-Tucker conditions near a solution. It is known that the SQP method is efficient and stable for wide range of problems. The SQP method requires the quadratic programming subproblems which handle the combinatorial aspect of the problem caused by inequality constraints. A solution of the quadratic programming problem itself requires rather expensive cost especially for large scale problems. Therefore it is desired to have an efficient and stable interior point method for nonlinear problems in the light of success of that in the field of large scale linear programs. This paper shows that it is in fact possible to construct such a method that is globally convergent theoretically, and efficient and stable in practice. The method is tested on 115 test problems in Hock and Schittkowski's collection [6]. For these problems our method solves all the problems with 20 to 30 function evaluations per problem and about 20 iterations per problem. Details are described in Section 5. A preliminary report of this paper has appeared in [11].

In this paper, problems to be solved are restricted to small to medium ones because we do not exploit the sparsity of the matrices here. However, recent report by Yamashita, Yabe and Tanabe [13] studies a trust region type method to utilize the sparsity of the Hessian of the Lagrangian. They report efficiency of the method that uses the barrier penalty function as in this paper. See also [1] for a trust region type interior point method. It is to be noted that the recent studies by Yamashita and Yabe [12] and Yabe and Yamashita [9] show the superlinear and/or quadratic convergence of a class of primal-dual interior point methods that use the Newton or quasi-Newton iteration for solving the barrier KKT conditions. Other reports on the local behavior of primal-dual interior point methods include [2] and [3].

In Section 2, we describe basic concepts in the primal-dual interior point method. The barrier penalty function which plays a key role in the method of this paper is introduced in Section 3, and analyzed there. A line search method that minimizes the barrier penalty function is described in Section 4, and is proved to be globally convergent. Section 5 reports the results of numerical experiment.

Notation. The subscript k denotes an iteration count. Subscripts i and j denote components of vectors and matrices. The superscript t denotes transposes of vectors and matrices. The vector e denotes the vector of all ones and the matrix I the identity matrix. For simplicity of description, we assume $\|\cdot\|$ denotes the l_2 norm for vectors. The symbol $\mathbf{R}^n$ denotes the n dimensional real vector space. The set $\mathbf{R}^n_+$ is defined by $\mathbf{R}^n_+ = \{x \in \mathbf{R}^n \,|x > 0\}$.

2 Primal-dual interior point method

In this paper, we consider the following constrained optimization problem:

$$(1) \qquad \begin{array}{lll} \text{minimize} & f(x), & x \in \mathbf{R}^n, \\ \text{subject to} & g(x) = 0, \ x \geq 0, & \end{array}$$

where we assume that the functions $f : \mathbf{R}^n \to \mathbf{R}^1$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice continuously differentiable.

Let the Lagrangian function of the above problem be defined by

$$(2) \qquad L(w) = f(x) - y^t g(x) - z^t x,$$

where $w = (x, y, z)^t$, and $y \in \mathbf{R}^m$ and $z \in \mathbf{R}^n$ are the Lagrange multiplier vectors which correspond to the equality and inequality constraints respectively. Then Karush-Kuhn-Tucker (KKT) conditions for optimality of the above problem are given by

$$(3) \qquad r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ XZe \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$(4) \qquad x \geq 0, \qquad z \geq 0,$$

where

$$\begin{aligned} \nabla_x L(w) &= \nabla f(x) - A(x)^t y - z, \\ A(x) &= \begin{pmatrix} \nabla g_1(x)^t \\ \vdots \\ \nabla g_m(x)^t \end{pmatrix}, \\ X &= \operatorname{diag}(x_1, \cdots, x_n), \\ Z &= \operatorname{diag}(z_1, \cdots, z_n). \end{aligned}$$

Now we approximate problem (1) by introducing the barrier function $F_B(\bullet; \mu) : \mathbf{R}^n_+ \to \mathbf{R}^1$,

$$(5) \qquad \begin{array}{ll} \text{minimize} & F_B(x; \mu) = f(x) - \mu \sum_{i=1}^{n} \log(x_i), \ x \in \mathbf{R}^n_+ \\ \text{subject to} & g(x) = 0, \end{array}$$

where the barrier parameter $\mu > 0$ is a given constant. It is well known that, if μ is sufficiently small, problem (5) is a good approximation to original problem (1) (see Fiacco and McCormick [4]). The optimality conditions for (5) are given by

$$\begin{aligned} \nabla f(x) - A(x)^t y - \mu X^{-1} e &= 0, \\ g(x) &= 0 \end{aligned} \tag{6}$$

and

$$x > 0,$$

where $y \in \mathbf{R}^m$ is the Lagrange multiplier for the equality constraints. If we introduce an auxiliary variable $z \in \mathbf{R}^n$ which is to be equal to $\mu X^{-1} e$, then the above conditions become the conditions

$$r(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ XZe - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{7}$$

and

$$x > 0, \qquad z > 0. \tag{8}$$

The introduction of the variable z is essential to the numerical success of the barrier based algorithm in this paper.

In this paper we call conditions (7) the barrier KKT conditions, and a point $w(\mu) = (x(\mu), y(\mu), z(\mu))$ that satisfies these conditions is called the barrier KKT point. We will use an interior point method for searching a point that approximately satisfies the above conditions, and finally obtain a point that satisfies the Karush-Kuhn-Tucker conditions by letting $\mu \downarrow 0$. This means that we force x and z be strictly positive during iterations. Therefore we delete inequality conditions (8) hereafter, and always assume that x and z are strictly positive in what follows. Here we note that

$$r(w, \mu) = r_0(w) - \mu \hat{e}, \tag{9}$$

where

$$\hat{e} = \begin{pmatrix} 0 \\ 0 \\ e \end{pmatrix}.$$

An algorithm of this paper approximately solves the sequence of conditions (7) with a decreasing sequence of the barrier parameter μ that tends to 0, and thus obtains an approximate solution to KKT conditions. For definiteness, we describe a prototype of such algorithm as follows.

Algorithm IP

Step 0. (Initialize) Set $\varepsilon > 0$, $M_c > 0$ and $k = 0$. Let a positive sequence $\{\mu_k\}, \mu_k \downarrow 0$ be given.

Step 1. (Termination) If $\|r_0(w)\| \leq \varepsilon$, then stop.

Step 2. (Approximate barrier KKT point) Find a point w_{k+1} that satisfies

$$\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k. \tag{10}$$

Step 3. (Update) Set $k := k+1$ and go to Step 1. □

The following theorem shows the global convergence property of Algorithm IP.

Theorem 1 *Let $\{w_k\}$ be an infinite sequence generated by Algorithm IP. Suppose that the sequences $\{x_k\}$ and $\{y_k\}$ are bounded. Then $\{z_k\}$ is bounded, and any accumulation point of $\{w_k\}$ satisfies KKT conditions (3) and (4).*

Proof. Assume that there exists an i such that $(z_k)_i \to \infty$. Equation (10) yields

$$\left| \frac{(\nabla f(x_k) - A(x_k)^t y_k)_i}{(z_k)_i} - 1 \right| \le M_c \frac{\mu_{k-1}}{(z_k)_i},$$

which is a contradiction because of the boundedness of $\{x_k\}$ and $\{y_k\}$. Thus the sequence $\{z_k\}$ is bounded.

Let $\hat{w}$ be any accumulation point of $\{w_k\}$. Since the sequences $\{w_k\}$ and $\{\mu_k\}$ satisfy (10) for each k and μ_k approaches zero, $r_0(\hat{w}) = 0$ follows from the definition of $r(w, \mu)$. Therefore the proof is complete. □

We note that the barrier parameter sequence $\{\mu_k\}$ in Algorithm IP need not be determined beforehand. The value of each μ_k may be set adaptively as the iteration proceeds. An example of updating method of μ_k is described in Section 5. We call condition (10) the approximate barrier KKT condition, and call a point that satisfies this condition the approximate barrier KKT point.

To find an approximate barrier KKT point for a given $\mu > 0$, we use the Newton-like method in this paper. Let $\Delta w = (\Delta x, \Delta y, \Delta z)^t$ be defined by a solution of

$$J(w)\Delta w = -r(w, \mu), \tag{11}$$

where

$$J(w) = \begin{pmatrix} G & -A(x)^t & -I \\ A(x) & 0 & 0 \\ Z & 0 & X \end{pmatrix}. \tag{12}$$

Then the basic iteration of the Newton-like method may be described as

$$w_{k+1} = w_k + \Lambda_k \Delta w_k, \tag{13}$$

where $\Lambda_k = \mathrm{diag}(\alpha_{xk} I_n, \alpha_{yk} I_m, \alpha_{zk} I_n)$ is composed of step sizes in x, y and z variables. If $G = \nabla_x^2 L(w)$, then Δw becomes Newton's direction for solving (7). To solve (11), we split the equations into two groups. Thus we solve

$$\begin{pmatrix} G + X^{-1}Z & -A(x)^t \\ -A(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -\nabla_x L(w) + \mu X^{-1} e - z \\ g(x) \end{pmatrix}, \tag{14}$$

for $(\Delta x, \Delta y)^t$, then we obtain Δz by the third equation in (11). If $G + X^{-1}Z$ is positive definite and $A(x)$ is of full rank, then the coefficient matrix in (14) is nonsingular. It will be useful to note that (14) can be written as

$$\begin{pmatrix} G + X^{-1}Z & -A(x)^t \\ -A(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + \mu X^{-1}e \\ g(x) \end{pmatrix}, \tag{15}$$

where

$$\tilde{y} = y + \Delta y. \tag{16}$$

In this paper, we will deal with the case in which the matrix G can be assumed nonnegative definite. Therefore, to solve the general nonlinear problems, we will use a positive definite quasi-Newton approximation to the Hessian matrix of the Lagrangian function to obtain the desired property of the matrix G.

3 Barrier penalty function

To globalize the convergence property of an interior point algorithm based on the above iteration, we introduce two auxiliary problems. Firstly we define the following problem:

$$\begin{array}{ll} \text{minimize} & F_P(x; \bar{\rho}) = f(x) + \bar{\rho} \sum_{i=1}^{m} |g_i(x)|, \quad x \in \mathbf{R}^n, \\ \text{subject to} & x \geq 0, \end{array} \tag{17}$$

where the penalty parameter $\bar{\rho}$ is a given positive constant. The necessary conditions for optimality of this problem are (see 14.2 of Fletcher [5])

$$\begin{array}{rcl} \nabla_x L(w) & = & 0, \\ y & \in & -\partial \left\{ \bar{\rho} \sum_{i=1}^{m} |g_i(x)| \right\}, \\ XZe & = & 0, \ x \geq 0, z \geq 0, \end{array} \tag{18}$$

where the notation ∂ means the subdifferential of the function in the braces with respect to g. In our case the second condition in (18) is equivalent to

$$\begin{array}{ll} -\bar{\rho} \leq y_i \leq \bar{\rho}, & g_i(x) = 0, \\ y_i = -\bar{\rho}, & g_i(x) > 0, \\ y_i = \bar{\rho}, & g_i(x) < 0, \end{array} \tag{19}$$

for each $i = 1, \cdots, m$. This condition can be expressed as

$$\bar{\rho}\,|g_i(x)| = -y_i g_i(x), \ -\bar{\rho} \leq y_i \leq \bar{\rho}, \ i = 1, \cdots, m,$$

or

$$|g_i(x)| + \frac{y_i g_i(x)}{\bar{\rho}} = 0, \ -\bar{\rho} \leq y_i \leq \bar{\rho}, \ i = 1, \cdots, m.$$

Therefore conditions (18) can be written as

$$r_0(w) = \begin{pmatrix} \nabla_x L(w) \\ r_E(w) \\ XZe \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{20}$$

and

$$\begin{array}{l} x \geq 0, \; z \geq 0, \\ -\bar{\rho} \leq y_i \leq \bar{\rho}, \; i = 1, \cdots, m, \end{array} \tag{21}$$

where

$$r_E(w)_i = |g_i(x)| + \frac{y_i g_i(x)}{\bar{\rho}}, \; i = 1, \cdots, m.$$

Note that we are using the same symbol $r_0(w)$ to denote the residual vector of the optimality conditions as in Section 2 for simplicity. If $\|y\|_\infty < \bar{\rho}$, conditions (18) are equivalent to conditions (3) and (4). In this sense, problem (17) is equivalent to problem (1).

Next we introduce the barrier penalty function $F(\bullet; \mu, \rho) : \mathbf{R}^n_+ \to \mathbf{R}^1$ by

$$F(x; \mu, \rho) = f(x) - \mu \sum_{i=1}^{n} \log x_i + \rho \sum_{i=1}^{m} |g_i(x)|, \tag{22}$$

where μ and ρ are given positive constants. This function plays an essential role in the method given in this paper. Let us approximate problem (17) by the following problem

$$\text{minimize} \quad F(x; \mu, \bar{\rho}), \quad x \in \mathbf{R}^n_+. \tag{23}$$

The necessary conditions for optimality of a solution to the above problem are

$$\begin{array}{rcl} \nabla_x L(w) & = & 0, \\ y & \in & -\partial \left\{ \bar{\rho} \sum_{i=1}^{m} |g_i(x)| \right\} \\ XZe & = & \mu e, \; x > 0, \; z > 0, \end{array} \tag{24}$$

where we introduce an auxiliary variable $z \in \mathbf{R}^n$ as in (7). As above, conditions (24) can be written as

$$r(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ r_E(w) \\ XZe - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{25}$$

and

$$\begin{array}{l} x > 0, \; z > 0, \\ -\bar{\rho} \leq y_i \leq \bar{\rho}, \; i = 1, \cdots, m, \end{array} \tag{26}$$

where we use the same symbol $r(w, \mu)$ as in Section 2 for simplicity. If $\bar{\rho} > \|y\|_\infty$ then conditions (25) and (26) coincide with (7) and (8). We call a point $w(\mu) = (x(\mu), y(\mu), z(\mu)) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ that satisfies (24) for a given $\mu > 0$ the barrier KKT point for this μ as before. We can use Algorithm IP to solve (17). The following theorem shows the global convergence property of Algorithm IP for solving (17).

Theorem 2 *Let $\{w_k\}$ be an infinite sequence generated by Algorithm IP for solving (17). Suppose that the sequences $\{x_k\}$ is bounded. Then $\{z_k\}$ is bounded, and any accumulation point of $\{w_k\}$ satisfies the optimality conditions (20) and (21) for problem (17).* □

Now we formulate a Newton-like iteration for solving the above conditions (25). Thus we calculate the first order change of (24) with respect to a change in w. This gives

$$\begin{aligned} (G + X^{-1}Z)\Delta x - A(x)^t \Delta y &= -\nabla_x L(w) + \mu X^{-1}e - z, \\ \tilde{y} \equiv y + \Delta y &\in -\partial \left\{ \bar{\rho} \sum_{i=1}^{m} \left| g_i(x) + \nabla g_i(x)^t \Delta x \right| \right\}, \\ \Delta z &= -X^{-1}Z\Delta x + \mu X^{-1}e - z. \end{aligned} \tag{27}$$

Following lemma gives a basic property of the iteration vector $\Delta w = (\Delta x, \Delta y, \Delta z)^t$.

Lemma 1 *Suppose that Δw satisfies (27) at an interior point w.*
(i) If $\bar{\rho} > \|\tilde{y}\|_\infty$, then Δw is identical to the one given by (11).
(ii) If $\Delta w = 0$, then the point w is a barrier KKT point that satisfies (24).
(iii) If $\Delta x = 0$, then the point $(x, y + \Delta y, z + \Delta z)$ is a barrier KKT point that satisfies (24). □

If we consider the subproblem:

(28)

$$\text{minimize } \frac{1}{2}\Delta x^t (G + X^{-1}Z)\Delta x + (\nabla f(x) - \mu X^{-1}e)^t \Delta x + \bar{\rho} \sum_{i=1}^{m} \left| g_i(x) + \nabla g_i(x)^t \Delta x \right|, \ \Delta x \in \mathbf{R}^n,$$

the solution vector Δx and the corresponding multiplier vector $\tilde{y}$ that satisfy the necessary conditions for optimality also satisfy conditions (27). If $G + X^{-1}Z$ is positive definite we can solve the problem (28) by a straightforward active set method which starts with an active set that contains all the constraints $g_i(x) + \nabla g_i(x)^t \Delta x = 0, i = 1, \cdots m$. An example of the procedure is described in Fletcher [5]. It is important to note that, if the vectors Δx and $\tilde{y}$ obtained by solving (15) satisfy $\bar{\rho} > \|\tilde{y}\|_\infty$, then the desired iteration vector that satisfies (27) are also obtained.

The procedure described above is devised instead of the simple Newton iteration given in Section 2 in order to show a way of preventing possible divergence of the dual variable y. However for the practical purpose it seems sufficient to solve only equation (11) once per iteration as shown in sections below in which practical experiences obtained by the author on general nonlinear programming problems are described. Therefore this procedure is of only theoretical importance at present.

As shown above, optimality conditions (20) and (21) are identical to optimality conditions (3) and (4) if we set $\overline{\rho} = \infty$. Also, Newton-like iteration (27) coincides with (11) when $\overline{\rho} = \infty$. Therefore we will regard that the method explained below includes a method for solving (1) when $\overline{\rho} = \infty$ for simplicity of exposition. We note that the penalty parameter ρ which will appear below should be finite even if $\overline{\rho} = \infty$.

Another interesting point to be noted in the above iteration is that it can give a solution to an infeasible problem. Even if the constraints in problem (1) are incompatible, the method described above will give a solution to problem (17) as shown below. A solution to problem (17) with infeasible constraints may give useful information about problem (1).

Now we proceed to an analysis of the properties of the barrier penalty function. The directional derivative $F'(x;\mu,\rho;s)$ of the function $F(x;\mu,\rho)$ along an arbitrary given direction s is defined by

$$\begin{aligned} F'(x;\mu,\rho;s) &= \lim_{\alpha\downarrow 0}\frac{F(x+\alpha s;\mu,\rho)-F(x;\mu,\rho)}{\alpha} \\ &= \nabla f(x)^t s - \mu e^t X^{-1}s + \rho\sum_{+}\nabla g_i(x)^t s \\ &\quad +\rho\sum_{0}\left|\nabla g_i(x)^t s\right| - \rho\sum_{-}\nabla g_i(x)^t s, \end{aligned}$$

where the summations in the above equation are to be understood as

$$\sum_{+} a_i = \sum_{g_i>0} a_i, \quad \sum_{0} a_i = \sum_{g_i=0} a_i, \quad \sum_{-} a_i = \sum_{g_i<0} a_i.$$

We introduce a first order approximation F_l of $F(x+s;\mu,\rho)$ by

$$\begin{aligned} F_l(x;\mu,\rho;s) &\equiv f(x)+\nabla f(x)^t s - \mu\sum_{i=1}^{n}\left(\log(x_i)+\frac{s_i}{x_i}\right) \\ &\quad +\rho\sum_{i=1}^{m}\left|g_i(x)+\nabla g_i(x)^t s\right|, \end{aligned} \tag{29}$$

and an estimate of the first order change ΔF_l of F by

$$\begin{aligned} \Delta F_l(x;\mu,\rho;s) &\equiv F_l(x;\mu,\rho;s) - F(x;\mu,\rho), \\ &= \nabla f(x)^t s - \mu e^t X^{-1} s \\ &\quad +\rho\sum_{i=1}^{m}|g_i(x)+\nabla g_i(x)s| - \rho\sum_{i=1}^{m}|g_i(x)| \end{aligned} \tag{30}$$

for an arbitrary given direction s.

We have following properties for the quantities defined above which give an extension to the similar properties in the case of differentiable functions (see for example Yamashita [10]).

Lemma 2 *Let $\mu>0$, $\rho>0$ and $s\in\mathbf{R}^n$ be given. Then the following assertions hold.*
(i) The function $F_l(x;\mu,\rho;\alpha s)$ is convex with respect to the variable α.
(ii) There holds the relation

$$F(x;\mu,\rho)+F'(x;\mu,\rho;s)\le F_l(x;\mu,\rho;s). \tag{31}$$

(iii) Further, there exists a $\theta\in(0,1)$ such that

$$F(x+s;\mu,\rho)\le F(x;\mu,\rho)+F'(x+\theta s;\mu,\rho;s), \tag{32}$$

whenever $x+s>0$.

Proof. The first statement of the lemma is obvious. If $\alpha > 0$ is sufficiently small, we have

$$\begin{aligned} F_l(x;\mu,\rho;\alpha s) &= F(x;\mu,\rho) + F'(x;\mu,\rho;\alpha s) \\ &= F(x;\mu,\rho) + \alpha F'(x;\mu,\rho;s). \end{aligned}$$

Since $F_l(x;\mu,\rho;\alpha s)$ is convex with respect to α and coincides with a linear function of $\alpha > 0$ when α is sufficiently small, we obtain (31). Now we show (32). If we consider $F'(x+\tau s;\mu,\rho;s)$ as a function of the variable $\tau \in [0,1]$, the number of discontinuous points of F' is finite. Therefore there exists a $\theta \in (0,1)$ such that

$$\begin{aligned} F'(x+\theta s;\mu,\rho;s) &\geq \int_0^1 F'(x+\tau s;\mu,\rho;s)d\tau \\ &= F(x+s;\mu,\rho) - F(x;\mu,\rho). \end{aligned}$$

This completes the proof. □

Lemma 3 *Let $\varepsilon_0 \in (0,1)$ be a given constant and $s \in \mathbf{R}^n$ be given. If $\Delta F_l(x;\mu,\rho;s) < 0$, then*

(33) $$F(x+\alpha s;\mu,\rho) - F(x;\mu,\rho) \leq \varepsilon_0 \alpha \Delta F_l(x;\mu,\rho;s),$$

for sufficiently small $\alpha > 0$.

Proof. From (32), there exists a $\theta \in (0,1)$ such that

(34) $$\begin{aligned} F(x+\alpha s;\mu,\rho) - F(x;\mu,\rho) &\leq F'(x+\theta\alpha s;\mu,\rho;\alpha s) \\ &= \alpha F'(x+\theta\alpha s;\mu,\rho;s). \end{aligned}$$

From (31) we have

(35) $$F'(x+\theta\alpha s;\mu,\rho;s) \leq \Delta F_l(x+\theta\alpha s;\mu,\rho;s).$$

Because $\Delta F_l(\bullet;\mu,\rho;s)$ is continuous, and $\Delta F_l(x;\mu,\rho;s) < 0$ by the assumption, we obtain

(36) $$\Delta F_l(x+\theta\alpha s;\mu,\rho;s) \leq \varepsilon_0 \Delta F_l(x;\mu,\rho;s),$$

for sufficiently small $\|\theta\alpha s\|$. From (34) - (36) we obtain (33) for sufficiently small $\alpha > 0$.
□

Lemma 4 *Suppose that Δw satisfies (27). If $0 < \rho \leq \bar{\rho}$, then*

(37) $$\Delta F_l(x;\mu,\rho;\Delta x) \leq -\Delta x^t(G + X^{-1}Z)\Delta x - \sum_{i=1}^{m}(\rho - |\tilde{y}_i|)|g_i(x)|.$$

Further if, G is positive semidefinite and $\|\tilde{y}\|_\infty \leq \rho$, then $\Delta F_l(x;\mu,\rho;\Delta x) \leq 0$, and $\Delta F_l(x;\mu,\rho;\Delta x) = 0$ yields $\Delta x = 0$.

Proof. From (27) and (30) we have

$$\Delta F_l(x;\mu,\rho;\Delta x) = -\Delta x^t(G+X^{-1}Z)\Delta x+\Delta x^t A(x)^t\tilde{y}+\rho\sum_{i=1}^{m}\left|g_i(x)+\nabla g_i(x)^t\Delta x\right|-\rho\sum_{i=1}^{m}|g_i(x)|.$$

The i-th components in the summations in the last three terms give

$$\begin{aligned}(38)\qquad & \tilde{y}_i\nabla g_i(x)^t\Delta x+\rho\left|g_i(x)+\nabla g_i(x)^t\Delta x\right|-\rho\left|g_i(x)\right| \\ & \le \tilde{y}_i\nabla g_i(x)^t\Delta x+\bar{\rho}\left|g_i(x)+\nabla g_i(x)^t\Delta x\right|-\rho\left|g_i(x)\right| \\ & = -\tilde{y}_i g_i(x)-\rho\left|g_i(x)\right|,\end{aligned}$$

where the inequality in the second line follows from $\rho \le \bar{\rho}$, and the equality in the third line follows from the property

$$\begin{array}{ll} -\bar{\rho}\le\tilde{y}_i\le\bar{\rho}, & g_i(x)+\nabla g_i(x)^t\Delta x=0, \\ \tilde{y}_i=-\bar{\rho}, & g_i(x)+\nabla g_i(x)^t\Delta x>0, \\ \tilde{y}_i=\bar{\rho}, & g_i(x)+\nabla g_i(x)^t\Delta x<0. \end{array}$$

The relation (38) gives (37). □

4 Line search algorithm

To obtain a globally convergent algorithm to a barrier KKT point for a fixed $\mu > 0$, it is necessary to modify the basic Newton iteration with the unit step length somehow. Our iterations consist of

$$(39)\qquad \begin{aligned} x_{k+1} &= x_k+\alpha_{xk}\Delta x_k, \\ y_{k+1} &= y_k+\alpha_{yk}\Delta y_k, \\ z_{k+1} &= z_k+\alpha_{zk}\Delta z_k, \end{aligned}$$

where α_{xk}, α_{yk} and α_{zk} are step sizes determined by the line search procedures described below.

The main iteration is to decrease the value of the barrier penalty function for fixed μ. Thus the step size of the primal variable x is determined by the sufficient decrease rule of the merit function. The step size of the dual variable z is determined so as to stabilize the iteration. The explicit rules follow in order.

We adopt Armijo's rule as the line search rule for the variable x. At the point x_k, we calculate the maximum allowed step to the boundary of the feasible region by

$$(40)\qquad \alpha_{k\max}=\min_i\left\{-\frac{(x_k)_i}{(\Delta x_k)_i}\,\middle|\,(\Delta x_k)_i<0\right\},$$

i.e., the step size $\alpha_{k\max}$ gives an infinitely large value of the barrier penalty function F if it exists, because of the barrier terms, and a step size $\alpha\in[0,\alpha_{k\max})$ gives a strictly feasible primal variable. A step to the next iterate is given by

$$(41)\qquad \alpha_{xk}=\bar{\alpha}_k\beta^{l_k},\qquad \bar{\alpha}_k=\min\{\gamma\alpha_{k\max},1\},$$

where $\gamma \in (0,1)$ and $\beta \in (0,1)$ are fixed constants and l_k is the smallest nonnegative integer such that

$$F(x_k + \bar{\alpha}_k \beta^{l_k} \Delta x_k; \mu, \rho) - F(x_k; \mu, \rho) \leq \varepsilon_0 \bar{\alpha}_k \beta^{l_k} \Delta F_l(x_k; \mu, \rho; \Delta x_k), \tag{42}$$

where $\varepsilon_0 \in (0,1)$. Typical values of these parameters are $\beta = 0.5$, $\gamma = 0.9995$ and $\varepsilon_0 = 10^{-6}$. Therefore we will try the sequence

$$x_k + 0.9995\alpha_{k\max}\Delta x_k, \quad x_k + 0.5 \times 0.9995\alpha_{k\max}\Delta x_k, \quad x_k + 0.25 \times 0.9995\alpha_{k\max}\Delta x_k, \quad \cdots$$

for example, and will find a step size that satisfies (42). If G is positive semidefinite, then $\Delta F_l(x_k; \mu, \rho; \Delta x_k) \leq 0$ by Lemma 4, and therefore the existence of such steps is assured by Lemma 3.

For the variable z, we adopt the box constraints rule, i.e., we force x and z to satisfy the condition

$$c_{Lki} \leq ((x_k)_i + \alpha_{xk}(\Delta x_k)_i)((z_k)_i + \alpha_{zk}(\Delta z_k)_i) \leq c_{Uki}, \quad i = 1, \cdots, n \tag{43}$$

at the end of each iteration, where the bounds c_{Lk} and c_{Uk} satisfy

$$0 < c_{Lki} < \mu < c_{Uki}, \quad i = 1, \cdots, n. \tag{44}$$

To this end, we let

$$\begin{aligned} c_{Lki} &= \min\left\{\frac{\mu}{M_L}, ((x_k)_i + \alpha_{xk}(\Delta x_k)_i)(z_k)_i\right\}, \\ c_{Uki} &= \max\left\{M_U\mu, ((x_k)_i + \alpha_{xk}(\Delta x_k)_i)(z_k)_i\right\}, \end{aligned} \tag{45}$$

where $M_L > 1$ and $M_U > 1$ are given constants. The construction of the above bounds shows that current z satisfies

$$\frac{c_{Lki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)} \leq (z_k)_i \leq \frac{c_{Uki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)}, \quad i = 1, \cdots, n. \tag{46}$$

The step size α_z is determined by

$$\begin{aligned} \alpha_{zk} \quad = \quad \min\Bigg\{\min_i \Bigg\{\max_{\alpha_i}\Bigg\{\alpha_i \Bigg| & \frac{c_{Lki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)} \\ & \leq ((z_k)_i + \alpha_i(\Delta z_k)_i) \leq \frac{c_{Uki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)}\Bigg\}\Bigg\}, 1\Bigg\}. \end{aligned} \tag{47}$$

The rule (47) means that the step size α_z is the maximal allowed step that satisfies the box constraints with the restriction of being not greater than the unit step length.

Lemma 5 *Suppose that an infinite sequence $\{w_k\}$ is generated for fixed $\mu > 0$. Then if $\liminf_{k\to\infty}(x_k)_i > 0$ and $\limsup_{k\to\infty}(x_k)_i < \infty$, then $\liminf_{k\to\infty}(c_{Lk})_i > 0$ and $\limsup_{k\to\infty}(c_{Uk})_i < \infty$ for $i = 1, \cdots, n$.*

Proof. Suppose that $(c_{Lk})_i \to 0$ for an i and some subsequence $K \subset \{0, 1, 2, \cdots\}$. Then by the definition of $(c_{Lk})_i$ in (45), $(z_k)_i \to 0, k \in K$. However, in order for a subsequence of $\{(z_k)_i\}$ to tend to 0, there must be an iteration k at which the lower bound $(c_{Lk})_i/(x_{k+1})_i$ of $(z_k)_i$ is arbitrary small and the value of $(z_k)_i$ at the iteration is strictly larger than that bound, i.e. at the iteration the value of $(z_k)_i$ decreases to a strictly smaller value. This means that at iteration k, $(c_{Lk})_i = \mu/M_L$ from the definition (45), and therefore the value of $(x_{k+1})_i$ must be arbitrary large because $\mu/M_L < (x_{k+1})_i(z_k)_i$ and $(z_k)_i \to 0, k \in K$. This is impossible because of the assumption of the lemma. The proof of the boundedness of $(c_{Uk})_i$ is similar. □

In actual calculation we modify the direction Δz_k by

$$(\Delta z_k')_i = \begin{cases} 0, & \text{if } (z_k)_i = c_{Lki}/(x_{k+1})_i \text{ and } (\Delta z_k)_i < 0, \\ 0, & \text{if } (z_k)_i = c_{Uki}/(x_{k+1})_i \text{ and } (\Delta z_k)_i > 0, \\ (\Delta z_k)_i, & \text{otherwise.} \end{cases} \tag{48}$$

This modification means that we project the direction along the boundary of the box constraints if the point z_k is on that boundary and the direction Δz_k points outward of the box. This procedure is adopted because it gives better numerical results. The global convergence results shown in the following are equally valid for both unmodified and modified directions.

For the variable y, there exist three obvious choices for the step length:

$$\alpha_{yk} = 1 \quad \text{or} \quad \alpha_{xk} \quad \text{or} \quad \alpha_{zk}. \tag{49}$$

The global convergence property given below holds for these choices. We choose $\alpha_{yk} = \alpha_{zk}$ from numerical experiments.

The following algorithm describes the iteration for fixed $\mu > 0$ and $\rho > 0$. We note that this algorithm corresponds to Step 2 of Algorithm IP in Section 2.

Algorithm LS

Step 0. (Initialize) Let $w_0 \in \mathbf{R}^n_+ \times \mathbf{R}^m \times \mathbf{R}^n_+$, and $\mu > 0$, $\rho > 0$. Set $\varepsilon' > 0$, $\gamma \in (0, 1)$, $\beta \in (0, 1)$, $\varepsilon_0 \in (0, 1)$, $M_L > 1$ and $M_U > 1$. Let $k = 0$.

Step 1. (Termination) If $\|r(w_k, \mu)\| \le \varepsilon'$, then stop.

Step 2. (Compute direction) Calculate the direction Δw_k by (27).

Step 3. (Stepsize) Set

$$\alpha_{k\max} = \min_i \left\{ -\frac{(x_k)_i}{(\Delta x_k)_i} \middle| (\Delta x_k)_i < 0 \right\}, \quad \bar{\alpha}_k = \min\{\gamma \alpha_{k\max}, 1\}.$$

Find the smallest nonnegative integer l_k that satisfies

$$F(x_k + \bar{\alpha}_k \beta^{l_k} \Delta x_k, \mu, \rho) - F(x_k, \mu, \rho) \le \varepsilon_0 \bar{\alpha}_k \beta^{l_k} \Delta F_l(x_k; \mu, \rho; \Delta x_k).$$

Calculate

$$\alpha_{xk} = \bar{\alpha}_k \beta^{l_k},$$

$$c_{Lki} = \min\left\{\frac{\mu}{M_L}, (x_k + \alpha_{xk}\Delta x_k)_i (z_k)_i\right\},$$
$$c_{Uki} = \max\left\{M_U\mu, (x_k + \alpha_{xk}\Delta x_k)_i (z_k)_i\right\},$$

$$\alpha_{zk} = \min\left\{\min_i\left\{\max_{\alpha_i}\left\{\alpha_i \,\middle|\, \frac{c_{Lki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)} \right.\right.\right.$$
$$\left.\left.\left. \leq ((z_k)_i + \alpha_i(\Delta z_k)_i) \leq \frac{c_{Uki}}{((x_k)_i + \alpha_{xk}(\Delta x_k)_i)}\right\}\right\}, 1\right\},$$
$$\alpha_{yk} = \alpha_{zk},$$
$$\Lambda_k = \mathrm{diag}\{\alpha_{xk}I_n, \alpha_{yk}I_m, \alpha_{zk}I_n\}.$$

Step 4. (Update variables) Set
$$w_{k+1} = w_k + \Lambda_k \Delta w_k.$$

Step 5. Set $k := k + 1$ and go to Step 1. □

To prove global convergence of Algorithm LS, we need the following assumptions.

Assumption G

(1) The functions f and $g_i, i = 1, ..., m$, are twice continuously differentiable.

(2) The level set of the barrier penalty function at an initial point $x_0 \in \mathbf{R}^n_+$, which is defined by $\left\{x \in \mathbf{R}^n_+ \,|F(x;\mu,\rho) \leq F(x_0;\mu,\rho)\right\}$, is compact for given $\mu > 0$.

(3) The matrix $A(x)$ is of full rank on the level set defined in (2).

(4) The matrix G_k is positive semidefinite and uniformly bounded.

(5) The penalty parameter ρ satisfies $\overline{\rho} \geq \rho \geq \|y_k + \Delta y_k\|_\infty$ for each $k = 0, 1, \ldots$. □

We note that if a quasi-Newton approximation is used for computing the matrix G_k, then we need the continuity of only the first order derivatives of functions in Assumption G-(1). We also note that if $\Delta F_l(x_k;\mu,\rho;\Delta x_k) = 0$, at an iteration k, then the step sizes $\alpha_{xk} = \alpha_{yk} = \alpha_{zk} = 1$ are adopted and $(x_{k+1}, y_{k+1}, z_{k+1})$ gives a barrier KKT point from Lemma 1 and Lemma 4. The following theorem gives a convergence of an infinite sequence generated by Algorithm LS.

Theorem 3 *Let an infinite sequence $\{w_k\}$ be generated by Algorithm LS. Then there exists at least one accumulation point of $\{w_k\}$, and any accumulation point of the sequence $\{w_k\}$ is a barrier KKT point.*

Proof. First we note that each component of the sequence $\{x_k\}$ is bounded away from zero and bounded above by the assumption and the existence of the log barrier term. Therefore the sequence $\{x_k\}$ has at least one accumulation point. The sequence $\{z_k\}$ also has these properties by Lemma 5. Thus there exists a positive number M such that

$$\frac{\|p\|^2}{M} \le p^t(G_k + X_k^{-1}Z_k)p \le M\|p\|^2, \quad \forall p \in \mathbf{R}^n, \tag{50}$$

by the assumption. From (37) and (50), we have

$$\Delta F_l(x_k;\mu,\rho;\Delta x_k) \le -\frac{\|\Delta x_k\|^2}{M} < 0, \tag{51}$$

and from (42),

$$\begin{aligned} F(x_{k+1};\mu,\rho) - F(x_k;\mu,\rho) &\le \varepsilon_0\bar{\alpha}_k\beta^{l_k}\Delta F_l(x_k;\mu,\rho;\Delta x_k) \\ &\le -\varepsilon_0\bar{\alpha}_k\beta^{l_k}\frac{\|\Delta x_k\|^2}{M} < 0. \end{aligned} \tag{52}$$

Because the sequence $\{F(x_k;\mu,\rho)\}$ is decreasing and bounded below, the left hand side of (52) converges to 0. Since $\liminf_{k\to\infty}(x_k)_i > 0, i = 1,\cdots,n$, we have $\liminf_{k\to\infty}\bar{\alpha}_k > 0$. Suppose that there exists a subsequence $K \subset \{0,1,\cdots\}$ and a δ such that

$$\liminf_{k\to\infty}\|\Delta x_k\| \ge \delta > 0, \quad k \in K. \tag{53}$$

Then we have $l_k \to \infty, k \in K$ from (52) because the left most expression tends to zero, and therefore we can assume $l_k > 0$ for sufficiently large $k \in K$ without loss of generality. If $l_k > 0$ then the point $x_k + \alpha_{xk}\Delta x_k/\beta$ does not satisfy the condition (42). Thus, we have

$$F(x_k + \alpha_{xk}\Delta x_k/\beta;\mu,\rho) - F(x_k;\mu,\rho) > \varepsilon_0\alpha_{xk}\Delta F_l(x_k;\mu,\rho;\Delta x_k)/\beta. \tag{54}$$

By (32) and (31), there exists a $\theta_k \in (0,1)$ such that

(55)

$$\begin{aligned} F(x_k + \alpha_{xk}\Delta x_k/\beta;\mu,\rho) - F(x_k;\mu,\rho) &\le \alpha_{xk}F'(x_k + \theta_k\alpha_{xk}\Delta x_k/\beta;\mu,\rho;\Delta x_k)/\beta \\ &\le \alpha_{xk}\Delta F_l(x_k + \theta_k\alpha_{xk}\Delta x_k/\beta;\mu,\rho;\Delta x_k)/\beta, k \in K. \end{aligned}$$

Then, from (54) and (55), we have

$$\varepsilon_0\Delta F_l(x_k;\mu,\rho;\Delta x_k) < \Delta F_l(x_k + \theta_k\alpha_{xk}\Delta x_k/\beta;\mu,\rho;\Delta x_k).$$

This inequality yields

$$\begin{aligned} &\Delta F_l(x_k + \theta_k\alpha_{xk}\Delta x_k/\beta;\mu,\rho;\Delta x_k) - \Delta F_l(x_k;\mu,\rho;\Delta x_k) \\ > \; &(\varepsilon_0 - 1)\Delta F_l(x_k;\mu,\rho;\Delta x_k) > 0. \end{aligned} \tag{56}$$

Because Δx_k is a solution of problem (28) and there holds (50), $\|\Delta x_k\|$ is uniformly bounded above. Then by the property $l_k \to \infty$, we have $\|\theta_k\alpha_{xk}\Delta x_k/\beta\| \to 0, k \in K$.

Thus the left hand side of (56) and therefore $\Delta F_l(x_k;\mu,\rho;\Delta x_k)$ converges to zero when $k\to\infty, k\in K$. This contradicts the assumption (53) because we have $\Delta x_k\to 0, k\in K$ from (51). Therefore we proved

$$\lim_{k\to\infty,k\in K}\|\Delta x_k\| = 0. \tag{57}$$

Let an arbitrary accumulation point of the sequence $\{x_k\}$ be $\hat{x}\in \mathbf{R}^n_+$ and let $x_k\to\hat{x}, k\in K$ for $K\subset\{0,1,\cdots\}$. Thus

$$x_k\to\hat{x},\quad \Delta x_k\to 0,\quad x_{k+1}\to\hat{x},\quad k\in K. \tag{58}$$

Because $\left\{X_k^{-1}Z_k\right\}$ is bounded, we have

$$\lim_{k\to\infty,k\in K}\left\|z_k+\Delta z_k-\mu X_k^{-1}e\right\| = 0$$

from (27). If we define $\hat{z}=\mu\hat{X}^{-1}e$ where $\hat{X}=\mathrm{diag}(\hat{x}_1,\cdots,\hat{x}_n)$, then we have

$$z_k+\Delta z_k\to\hat{z},\quad k\in K.$$

Hence from (45) we have

$$(c_{Lk})_i\le\frac{\mu}{M_L}\le(x_{k+1})_i(z_k+\Delta z_k)_i\le M_U\mu\le(C_{Uk})_i,\quad i=1,\cdots,n$$

for $k\in K$ sufficiently large, which shows that the point $z_k+\Delta z_k$ is always accepted as z_{k+1} for sufficiently large $k\in K$.

Since $\alpha_{zk}=1$ is accepted for $k\in K$ sufficiently large, so is $\alpha_{yk}=1$. Therefore we obtain

$$\begin{aligned}\lim_{k\to\infty,k\in K}\nabla_x L(\hat{x},y_k+\Delta y_k,\hat{z}) &= 0,\\ \lim_{k\to\infty,k\in K} y_k+\Delta y_k &\in -\partial\left\{\bar{\rho}\sum_{i=1}^{n}|g_i(\hat{x})|\right\}.\end{aligned}$$

Because the matrix $A(\hat{x})$ is of full rank, the sequence $\{y_k+\Delta y_k\}, k\in K$ converges to a point $\hat{y}\in\mathbf{R}^m$ which satisfies

$$\begin{aligned}\nabla_x L(\hat{x},\hat{y},\hat{z}) &= 0,\\ \hat{y} &\in -\partial\left(\bar{\rho}\sum_{i=1}^{m}|g_i(\hat{x})|\right),\\ \hat{X}\hat{z} &= \mu e,\ \hat{x}>0, \hat{z}>0.\end{aligned}$$

This completes the proof because we proved that there exists at least one accumulation point of $\{x_k\}$, and for an arbitrary accumulation point $\hat{x}$ of $\{x_k\}$, there exist unique $\hat{y}$ and $\hat{z}$ that satisfy the above. □

5 Numerical Result

In this section, we report numerical results of an implementation of the algorithm given in this paper for nonlinear programming problems. We set $\bar{\rho} = \infty$ in this experiment. The software is called NUOPT and the code is written by Takahito Tanabe. In order to have an appropriate positive semidefinite matrix G by a reasonable cost for nonlinear problems, we resort to a quasi-Newton approximation to the Hessian matrix of the Lagrangian function. We use updating formula suggested by Powell[7] for the SQP method:

$$G_{k+1} = G_k - \frac{G_k s_k s_k^t G_k}{s_k^t G_k s_k} + \frac{u_k u_k^t}{s_k^t u_k},$$

where u_k is calculated by

$$\begin{aligned} s_k &= x_{k+1} - x_k, \\ v_k &= \nabla_x L(x_{k+1}, y_{k+1}, z_{k+1}) - \nabla_x L(x_k, y_{k+1}, z_{k+1}), \\ u_k &= \theta_k v_k + (1 - \theta_k) G_k s_k, \\ \theta_k &= \begin{cases} 1, & s_k^t v_k \geq 0.2 s_k^t G_k s_k, \\ \frac{0.8 s_k^t G_k s_k}{s_k^t G_k s_k - s_k^t v_k}, & s_k^t v_k \leq 0.2 s_k^t G_k s_k, \end{cases} \end{aligned}$$

to satisfy $s_k^t u_k > 0$ for the hereditary positive definiteness of the update.

Method for updating the barrier parameter μ_k is as follows. Suppose we have an approximate barrier KKT point w_{k+1} that satisfies

$$\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k,$$

in Step 2 of Algorithm IP. Then μ_{k+1} is defined by

$$\mu_{k+1} = \max \left\{ \frac{\|r(w_{k+1}, \mu_k)\|}{M_\mu}, \frac{\mu_k}{M_0} \right\},$$

where $0 < M_c < M_\mu$ and $M_0 > 1$ should be satisfied. In our experiment we set $M_c = 30, M_\mu = 40, M_0 = 50$.

As in the SQP method, we expect fast local convergence of the method if μ is sufficiently small near a solution because it is based on the Newton iteration for the optimality conditions and a quasi-Newton approximation to the second derivative matrix. As noted in the above, this expectation is proved by Yamashita and Yabe [12]. Linear equation (14) is solved by using the Bunch-Parlett factorization.

The test problems for nonlinear problems are adopted from the book by Hock and Schittkowski [6]. The results are summarized in Table 1 at the end of this paper. Following list explains the notations used in Table 1:

n= number of variables.
m=number of constraints.
obj=final objective function value.
res=norm of final KKT condition residual.
itr=iteration count.

neval= number of function evaluations.
nfact=number of factorizations.

From the textbook [6] we adopt 115 problems. All the problems tried are solved by our code from the starting point mentioned in the text book. Of these, one is solved by a separate run because of the reason explained below. Accuracies listed in Table 1 for 110 test problems are obtained by an identical set of parameters:

$$\beta = 0.5, \gamma = 0.9995, \varepsilon_0 = 1 \times 10^{-6}, M_{Lz} = 2.5, M_{Uz} = 10, M_c = 175.$$

Of these problems we obtained local optimal points for 7 problems. Problem HS13 does not satisfy the constraint qualifications, but our code can solve it successfully. However our code requires large number of iterations for this problem and therefore we list this result separately. We obtained a correct approximation to the primal variables, but the norm of Karush-Kuhn-Tucker conditions does not tend to 0.

From these experiments it can be said that the method given in this paper is efficient and stable. In the first consecutive tests for 114 problems the method requires 2576 function evaluations in 2094 iterations. It can be claimed that the globally convergent algorithm given in this paper is efficient and stable for small dense nonlinear programming problems.

Table 1. Numerical Results on Problems by Hock and Schittkowski

problem	n	m	obj	res	itr	neval	nfact	
HS1	2	1	5.00995e-11	1.0e-07	37	50	37	
HS2	2	1	4.94124	3.4e-07	16	18	16	*l
HS3	2	1	6.02158e-08	6.0e-08	11	13	11	
HS4	2	1	2.66667	1.6e-08	6	8	6	
HS5	2	1	-1.91322	2.2e-08	6	8	6	
HS6	2	2	6.0196e-17	2.7e-08	9	11	9	
HS7	2	2	-1.73205	5.8e-08	9	15	9	
HS8	2	3	-1	1.8e-11	5	8	5	
HS9	2	2	-0.5	7.7e-10	6	8	6	
HS10	2	2	-1	1.1e-07	13	15	13	
HS11	2	2	-8.49846	1.1e-06	7	9	7	
HS12	2	2	-30	3.1e-08	9	11	9	
HS14	2	3	1.39346	2.2e-09	6	8	6	
HS15	2	3	306.5	5.3e-07	9	11	9	
HS16	2	3	0.250033	3.4e-07	19	26	19	
HS17	2	3	1.00003	2.8e-07	15	19	15	
HS18	2	3	5	5.8e-09	14	16	14	
HS19	2	3	-6961.81	1.2e-06	10	17	10	
HS20	2	4	40.1989	6.3e-07	7	9	7	*l
HS21	2	2	-99.96	6.2e-07	7	9	7	
HS22	2	3	1	1.0e-08	7	9	7	
HS23	2	6	2	4.9e-07	11	13	11	
HS24	2	4	-1	2.5e-07	13	16	13	
HS25	3	1	1.81845e-16	1.7e-13	53	63	53	*t
HS26	3	2	2.24353e-11	1.1e-06	27	29	27	
HS27	3	2	0.04	3.7e-07	20	22	20	
HS28	3	2	3.76506e-17	2.4e-08	10	12	10	
HS29	3	2	-22.6274	3.5e-07	14	17	14	
HS30	3	2	1	3.7e-07	13	16	13	
HS31	3	2	6	1.1e-06	7	9	7	
HS32	3	3	1.00001	4.9e-07	9	11	9	
HS33	3	3	-4.58579	2.1e-07	17	23	17	
HS34	3	3	-0.834032	1.2e-06	9	11	9	
HS35	3	2	0.111111	2.5e-07	8	10	8	
HS36	3	2	-3300	1.2e-07	8	11	8	
HS37	3	3	-3456	1.2e-07	8	10	8	
HS38	4	1	6.42044e-07	8.7e-07	34	41	34	
HS39	4	3	-1	1.8e-07	12	14	12	
HS40	4	4	-0.25	2.4e-07	6	8	6	
HS41	4	2	1.92593	1.5e-07	9	11	9	
HS42	4	3	13.8579	1.8e-08	6	8	6	
HS43	4	4	-44	4.1e-07	8	10	8	
HS44	4	7	-15	5.5e-07	12	15	12	
HS45	5	1	1	2.4e-07	13	15	13	
HS46	5	3	5.66038e-10	1.3e-06	24	26	24	

problem	n	m	obj	res	itr	neval	nfact	
HS47	5	4	1.72265e-09	1.0e-06	22	27	22	
HS48	5	3	9.18249e-12	5.7e-07	10	12	10	
HS49	5	3	2.74891e-07	2.1e-07	28	30	28	
HS50	5	4	1.94337e-10	3.9e-08	19	21	19	
HS51	5	4	1.0279e-19	3.5e-07	3	5	3	
HS52	5	4	5.32665	8.6e-09	8	10	8	
HS53	5	4	4.09302	1.7e-07	7	9	7	
HS54	6	2	-0.867409	1.1e-07	39	46	39	
HS54	6	2	-0.903547	3.7e-11	80	87	80	*lt
HS55	6	7	6.66667	1.2e-08	7	9	7	*l
HS56	7	5	-3.456	4.1e-07	9	15	9	
HS57	2	2	0.0284597	2.6e-11	34	36	34	*t
HS59	2	4	-6.7495	4.3e-07	18	27	18	*l
HS60	3	2	0.0325682	5.7e-08	8	10	8	
HS61	3	3	-143.646	2.1e-09	7	9	7	
HS62	3	2	-26272.5	4.3e-08	8	12	8	
HS63	3	3	961.715	5.5e-08	8	10	8	
HS64	3	2	6299.85	1.1e-06	29	31	29	
HS65	3	2	0.953529	1.8e-07	15	17	15	
HS66	3	3	0.518163	2.7e-07	8	10	8	
HS67	3	15	-1162.12	1.1e-09	31	33	31	*b
HS68	4	3	-0.920425	9.5e-07	21	25	21	
HS69	4	3	-956.713	4.8e-07	14	20	14	
HS70	4	2	0.269086	1.2e-07	10	12	10	*l
HS71	4	3	17.014	1.1e-06	8	10	8	
HS72	4	3	727.679	1.3e-06	43	49	43	
HS73	4	4	29.8944	1.0e-07	12	16	12	
HS74	4	6	5126.5	1.2e-06	11	13	11	
HS75	4	6	5174.41	1.4e-10	12	20	12	
HS76	4	4	-4.68182	1.0e-06	8	10	8	
HS77	5	3	0.241505	9.2e-07	10	12	10	
HS78	5	4	-2.9197	1.4e-06	7	9	7	
HS79	5	4	0.0787768	1.3e-06	8	10	8	
HS80	5	4	0.0539498	1.2e-07	6	8	6	
HS81	5	4	0.0539498	9.0e-09	10	12	10	
HS83	5	4	-30665.5	6.8e-09	16	21	16	
HS84	5	4	-5280330	3.0e-10	21	25	21	
HS85	5	22	-2.2156	9.8e-07	35	41	35	*b
HS86	5	11	-32.3487	2.2e-07	12	14	12	
HS87	6	5	8927.6	6.4e-07	28	45	28	
HS88	2	2	1.36266	6.9e-07	22	30	22	
HS89	3	2	1.36266	7.0e-07	23	32	23	
HS90	4	2	1.36266	1.6e-09	29	45	29	
HS91	5	2	1.36266	5.9e-07	25	43	25	
HS92	6	2	1.36266	9.9e-07	26	45	26	
HS93	6	3	135.076	7.3e-07	29	34	29	

problem	n	m	obj	res	itr	neval	nfact	
HS95	6	5	0.0156327	7.0e-08	13	16	13	
HS96	6	5	0.0156272	2.5e-07	12	17	12	
HS97	6	5	3.13581	2.8e-07	22	26	22	
HS98	6	5	3.13583	1.3e-07	20	24	20	
HS99	7	3	-8.3108e+08	2.2e-07	10	12	10	
HS100	7	5	680.63	1.3e-06	16	18	16	
HS101	7	6	1809.76	6.1e-07	25	28	25	
HS102	7	6	911.881	2.8e-07	27	34	27	
HS103	7	6	543.668	5.0e-07	26	32	26	
HS104	8	6	3.95116	1.7e-07	19	21	19	
HS105	8	2	1044.61	3.6e-07	56	64	56	*b
HS106	8	7	7049.25	1.1e-07	39	42	39	
HS107	9	7	5055.01	1.1e-06	10	13	10	
HS108	9	14	-0.674981	1.4e-06	62	67	62	*l
HS109	9	11	5362.07	8.4e-07	21	23	21	
HS110	10	1	-45.7785	4.1e-08	6	10	6	
HS111	10	4	-47.7611	1.2e-06	57	64	57	
HS112	10	4	-47.7611	8.4e-07	17	23	17	*b
HS113	10	9	24.3062	6.5e-07	25	27	25	
HS114	10	12	-1768.81	7.5e-08	47	54	47	
HS116	13	15	97.5875	2.6e-07	82	94	82	
HS117	15	6	32.3487	4.4e-07	36	43	36	
HS118	15	18	664.82	1.4e-08	34	54	34	
HS119	16	9	244.9	5.6e-07	28	30	28	
TOTAL(114 prob.)					2094	2576	2094	
AVERAGE	4	4		4.2e-07	18.4	22.6	18.4	
HS13*c	2	2	1.01967	2.1e+02	101	108	100	*i

*l: local optimum obtained
*t: tighter convergence criterion (eps=1.e-10) needed
*b: better solution obtained
*i: iteration limit reached
*c: constraint qualification not satisfied

References

[1] R.H.Byrd, J.C.Gilbert and J.Nocedal, (1996) *A trust region method based on interior point techniques for nonlinear programming*, Technical Report OTC 96/02, Optimization Technology Center, Argonne National Laboratory

[2] R.H.Byrd, G.Liu and J.Nocedal, (1998) On the local behaviour of an interior point method for nonlinear programming, in *Numerical Analysis 1997*, D.F.Griffiths, D.J.Higham and G.A.Watson eds. (Longman), 37-56.

[3] A.S.El-Bakry, R.A.Tapia, T.Tsuchiya and Y.Zhang, (1996) On the formulation and theory of the Newton interior-point method for nonlinear programming, *Journal of Optimization Theory and Applications*, **89**, 507-541.

[4] A.V.Fiacco and G.P.McCormick, (1968) *Nonlinear Programming: Sequential Unconstrained Minimization Technique* (John Wiley and Sons)

[5] R.Fletcher (1987) *Practical Methods of Optimization*, Second Edition (John Wiley and Sons)

[6] W.Hock and K.Schittkowski, (1981) *Test Examples for Nonlinear Programming Codes*, (Springer-Verlag)

[7] M.J.D.Powell (1978) A fast algorithm for nonlinearly constrained optimization calculations, in *Numerical analysis, Dundee 1977* Ed. G.A.Watson, Lecture notes in Mathematics **630** (Springer-Verlag)

[8] J.-P.Vial (1994) Computational experiences with a primal-dual interior-point method for smooth convex programming, *Optimization methods and Software* **3** 285-310.

[9] H.Yabe and Y.Yamashita, (1997) Q-superlinear convergence of primal-dual interior point quasi-Newton methods for constrained optimization, *Journal of the Operations Research Society of Japan*, **40** 415-436.

[10] H.Yamashita (1981) Convergence conditions for optimization methods, in *The Newton Method and Related Topics*, eds. T.Yamamoto and K.Tanabe (Kinokuniya Book-Store Co.) 77-104.

[11] H.Yamashita (1992) *A globally convergent primal-dual interior point method for constrained optimization*, Technical Report, Mathematical Systems Institute Inc.

[12] H.Yamashita and H.Yabe (1996) Superlinear and quadratic convergence of primal-dual interior point methods for constrained optimization, *Mathematical Programming* **75** 377-397.

[13] H.Yamashita, H.Yabe and T.Tanabe, (1997) *A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization*, Technical Report, Mathematical Systems, Inc.