

## A PRIMAL-DUAL EXTERIOR POINT METHOD FOR NONLINEAR OPTIMIZATION\*

HIROSHI YAMASHITA<sup>†</sup> AND TAKAHITO TANABE<sup>†</sup>

**Abstract.** In this paper, primal-dual methods for general nonconvex nonlinear optimization problems are considered. The proposed methods are exterior point type methods that permit primal variables to violate inequality constraints during the iterations. The methods are based on the exact penalty type transformation of inequality constraints and use a smooth approximation of the problem to form primal-dual iteration based on Newton’s method as in usual primal-dual interior point methods. Global convergence and local superlinear/quadratic convergence of the proposed methods are proved. For global convergence, methods using line searches and trust region type searches are proposed. The trust region type method is tested with CUTEr problems and is shown to have similar efficiency to the primal-dual interior point method code IPOPT. It is also shown that the methods can be warm started easily, unlike interior point methods, and that the methods can be efficiently used in parametric programming problems.

**Key words.** primal-dual method, exterior point method, warm start, parametric programming

**AMS subject classifications.** 49M37, 90C30

**DOI.** 10.1137/060676970

**1. Introduction.** In this paper, we consider the following constrained optimization problem:

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

where we assume that the functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are smooth.

Let the Lagrangian function of the above problem be defined by

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

where  $w = (x, y, z)^t \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$  and  $y$  and  $z$  are the Lagrange multiplier vectors which correspond to the equality and inequality constraints, respectively. Then Karush–Kuhn–Tucker (KKT) conditions for the optimality of problem (1) are given by

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

and

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

where

$$\nabla_x L(w) = \nabla f(x) - A(x)^t y - z,$$

---

\*Received by the editors December 7, 2006; accepted for publication (in revised form) September 20, 2010; published electronically November 9, 2010.

<http://www.siam.org/journals/siopt/20-6/67697.html>

<sup>†</sup>Mathematical Systems Inc., 2-4-3, Shinjuku, Shinjuku-ku, Tokyo, Japan (hy@msi.co.jp, tanabe@msi.co.jp).

$$A(x) = \begin{pmatrix} \nabla g_1(x)^t \\ \vdots \\ \nabla g_m(x)^t \end{pmatrix},$$

$$X = \text{diag}(x_1, \dots, x_n), \quad Z = \text{diag}(z_1, \dots, z_n),$$

$$e = (1, \dots, 1)^t \in \mathbb{R}^n.$$

The interior point methods that use the log barrier function approximate problem (1) by the following:

$$(5) \quad \begin{aligned} &\text{minimize} && F_0(x) = f(x) - \mu \sum_{i=1}^n \log(x_i), \quad x \in \mathbb{R}^n, \\ &\text{subject to} && g(x) = 0, \quad x > 0, \end{aligned}$$

where  $\mu > 0$  is a barrier parameter. The KKT conditions of the above problem are

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

If we introduce the auxiliary variable  $z = \mu X^{-1}e$ , these conditions can be rewritten as

$$\begin{aligned} \nabla f(x) - A(x)^t y - z &= 0, \\ g(x) &= 0, \\ Xz &= \mu e, \quad x > 0, \quad z > 0. \end{aligned}$$

The primal-dual interior point methods try to solve the above conditions (barrier KKT conditions) by iterative methods. Usually the search direction is based on the Newton step for solving the equality part of the barrier KKT conditions. The iterates are kept in the interior region that satisfies  $x > 0$  and  $z > 0$  by definition.

Recent research on interior point methods for nonlinear optimization problems (see [1], [9], [11], [12], [13], [14], [15]) show good theoretical properties and practical performance for a wide range of problems. One possible drawback of the method is that the iterates should be kept strictly inside the interior region—the very basic nature of the algorithm. If the feasible region is “narrow,” iterates that start from a point far from a solution may take many iterations to arrive at the region near the solution. If an iterate happens to be near the boundary of the feasible region which is not close to a solution, it may not be easy to escape from the region and to arrive at the near center trajectory because of possible numerical difficulties when  $\mu$  is small. Also it is known that the warm start is not easy to utilize in the interior point method framework despite past research on this topic (see [5], [10]). Therefore it is of interest to consider an algorithm that does not need an interior point requirement and is able to utilize the warm start.

In this paper, we consider a primal-dual iteration that can lie outside the primal interior region. And we will show by various numerical experiments that the method is of similar performance with an interior point method for various test problems, that it can, in fact, utilize the warm start case, and that it is effective in parametric programming usage.

To this end we first define the following problem:

$$(6) \quad \begin{aligned} &\text{minimize} && F_0(x, \rho) = f(x) + \rho \sum_{i=1}^n |x_i|_-, \quad x \in \mathbb{R}^n, \\ &\text{subject to} && g(x) = 0, \end{aligned}$$

where  $\rho > 0$  is a penalty parameter and

$$|x|_- = \max \{-x, 0\} = \frac{|x| - x}{2}.$$

It is known that with sufficiently large  $\rho > 0$  and under certain conditions, the solution of (6) coincides with that of (1) as explained below. In this form, the nonnegativity restriction on the variable  $x$  in (1) is eliminated. Using a problem of the form (6) for solving (1) is not new, and there exist much research on nondifferentiable exact penalty function approaches to nonlinear optimization. See Chapters 12 and 14 of Fletcher [4] for a description of various aspects of this type of problem. We note that algorithmic discussions in [4] are mostly based on the sequential quadratic/linear programming type method that uses the active set method for solving subproblems that arise from approximating the original problem. Our intention in this paper is to show that it is possible to design practical primal-dual algorithms which use the smoothing of problem (6) as in the interior point method described above and to show that it is numerically efficient in practice. Thus we will consider solving problem (6) in the primal-dual space hereafter.

The necessary conditions for optimality of problem (6) are (see section 14.2 of Fletcher [4])

$$(7) \quad \begin{aligned} \nabla_x L(w) &= 0, \\ g(x) &= 0, \\ z &\in -\partial \left\{ \rho \sum_{i=1}^n |x_i|_- \right\}, \end{aligned}$$

where the symbol  $\partial$  means the subdifferential of the function in the braces with respect to  $x$ . In our case the third condition in (7) is equivalent to

$$\begin{aligned} 0 \leq z_i \leq \rho, & \quad x_i = 0, \\ z_i = 0, & \quad x_i > 0, \\ z_i = \rho, & \quad x_i < 0 \end{aligned}$$

for each  $i = 1, \dots, n$ . The above conditions can be expressed as

$$(8) \quad |x_i| z_i - \rho |x_i|_- = 0, \quad 0 \leq z_i \leq \rho, \quad i = 1, \dots, n.$$

Therefore conditions (7) can be written as

$$(9) \quad r_0(w) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ r_C(w) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$(10) \quad 0 \leq z \leq \rho,$$

where

$$r_C(w)_i = |x_i| z_i - \rho |x_i|_-, \quad i = 1, \dots, n.$$

Note that we are using the same symbol  $r_0(w)$  to denote the residual vector of the optimality conditions as in (3) for simplicity. If  $\|z\|_\infty < \rho$ , conditions (9) and (10)

are equivalent to conditions (3) and (4). In this sense, problem (6) is equivalent to problem (1).

The next step is to construct a smooth approximation to problem (6). We approximate the nondifferentiable function  $|a|$ ,  $a \in \mathbb{R}$ , by a smooth differentiable function  $h(a, \mu)$ , where  $\mu > 0$  is a parameter that controls the accuracy of the approximation. In this paper, we use the following function:

$$(11) \quad h(a, \mu) = \frac{1}{2} \left( \sqrt{a^2 + \mu^2} - a \right).$$

For later reference, we write the first and second derivatives of  $h(a, \mu)$  as

$$(12) \quad h'(a, \mu) = \frac{1}{2} \left( \frac{a}{\sqrt{a^2 + \mu^2}} - 1 \right) = -\frac{h(a, \mu)}{\sqrt{a^2 + \mu^2}},$$

$$(13) \quad h''(a, \mu) = \frac{\mu^2}{2(a^2 + \mu^2)^{3/2}}$$

and note that

$$h(a, \mu) > 0, \quad -1 < h'(a, \mu) < 0, \quad h''(a, \mu) > 0, \quad a \in \mathbb{R}$$

for  $\mu > 0$ .

Before proceeding further, a brief note on the function  $h$  is given here. The function  $h$  is of similar form with the Fischer–Bermeister function (see [3])  $f_{FB}(a, b) : \mathbb{R}^2 \rightarrow \mathbb{R}$  which is used as  $f_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b = 0$  for transforming the complementarity equation  $ab = 0, a \geq 0, b \geq 0$ . However, the term  $\sqrt{a^2 + \mu^2}$  in the function  $h$  is simply a well-known smoothing form of  $|a|$ , and we believe there is no straightforward logical connection between the two.

By using the function  $h(a, \mu)$ , problem (6) is approximated by the following problem:

$$\begin{aligned} & \text{minimize} && f(x) + \rho \sum_{i=1}^n h(x_i, \mu), \quad x \in \mathbb{R}^n \\ & \text{subject to} && g(x) = 0. \end{aligned}$$

The KKT conditions for the above problem are

$$\begin{aligned} \nabla f(x) - A^t y + \rho H'(x, \mu) e &= 0, \\ g(x) &= 0, \end{aligned}$$

where

$$H(x, \mu) = \text{diag} \{h(x_1, \mu), \dots, h(x_n, \mu)\}, \quad H'(x, \mu) = \text{diag} \{h'(x_1, \mu), \dots, h'(x_n, \mu)\}.$$

By introducing the auxiliary variable  $z$  as

$$z = -\rho H'(x, \mu) e,$$

we rewrite the KKT conditions as

$$\begin{aligned} \nabla_x L(w) &= 0, \\ g(x) &= 0, \\ z + \rho H'(x, \mu) e &= 0. \end{aligned}$$

However, we modify the third equation to

$$(14) \quad \sqrt{x_i^2 + \mu^2} \cdot z_i - \rho h(x_i, \mu) = 0, \quad i = 1, \dots, n.$$

Then (14) can be viewed as a smooth approximation to (8). This procedure is similar to the primal-dual interior point case, where  $z = \mu X^{-1}e$  is converted to  $Xz = \mu e$  as explained above. Then the resulting Newton equations described later in this paper will be similar in form to the interior point case. Thus we express the KKT conditions as

$$(15) \quad r(w, \mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ U(x, \mu)z - \rho H(x, \mu)e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

where

$$(16) \quad u(x_i, \mu) = \sqrt{x_i^2 + \mu^2}, i = 1, \dots, n, \quad U(x, \mu) = \text{diag} \{u(x_1, \mu), \dots, u(x_n, \mu)\}.$$

The algorithm of this paper approximately solves the sequence of conditions (15) with a decreasing sequence of the parameter  $\mu$  that tends to 0 and thus obtains a solution of the KKT conditions. For definiteness, we describe a prototype of such algorithm as follows.

**Algorithm EP**

**Step 0.** (Initialize) Set  $\varepsilon > 0$ ,  $M_c > 0$ ,  $\rho > 0$ , 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 KKT point) Find a point  $w_{k+1}$  that satisfies

$$(17) \quad \|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k,$$

$$0 \leq z_{k+1} \leq \rho.$$

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

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

**THEOREM 1.** *Let  $\{w_k\}$  be an infinite sequence generated by Algorithm EP. Then any accumulation point of  $\{w_k\}$  is a KKT point of problem (6).*

*Proof.* Let  $\hat{w} = (\hat{x}, \hat{y}, \hat{z})$  be any accumulation point of  $\{w_k\}$ . Since the sequences  $\{w_k\}$  and  $\{\mu_k\}$  satisfy (17) for each  $k$  and  $\mu_k$  approaches zero,  $\nabla_x L(\hat{w}) = 0$  and  $g(\hat{x}) = 0$  follow from the definition of  $r(w, \mu)$ . By the relation

$$\left| \sqrt{(x_k)_i^2 + \mu_{k-1}^2} (z_k)_i - \rho h((x_k)_i, \mu_{k-1}) \right| \leq M_c \mu_{k-1}, \quad i = 1, \dots, n,$$

we have

$$|\hat{x}_i| \hat{z}_i - \rho |\hat{x}_i|_- = 0$$

for  $\hat{x}_i \neq 0$ . We also have  $0 \leq \hat{z}_i \leq \rho$  for  $\hat{x}_i = 0$  because we pose the condition  $0 \leq z_k \leq \rho$  in Step 2. Therefore, the proof is complete.  $\square$

We note that the parameter sequence  $\{\mu_k\}$  in Algorithm EP need not be determined beforehand. The value of each  $\mu_k$  may be set adaptively as the iteration proceeds.

**2. Newton iteration and merit function.** To find an approximate KKT point for a given  $\mu > 0$ , we use the Newton-like method. Let  $\Delta w = (\Delta x, \Delta y, \Delta z)^t$  be defined by a solution of

$$(18) \quad J(w, \mu)\Delta w = -r(w, \mu),$$

where

$$(19) \quad J(w, \mu) = \begin{pmatrix} G & -A(x)^t & -I \\ A(x) & 0 & 0 \\ V(w, \mu) & 0 & U(x, \mu) \end{pmatrix},$$

$$(20) \quad \begin{aligned} v(w_i, \mu) &= z_i u'(x_i, \mu) - \rho h'(x_i, \mu) \\ &= \frac{x_i(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} + \frac{\rho}{2}, \quad i = 1, \dots, n, \\ V(w, \mu) &= \text{diag} \{v(w_1, \mu), \dots, v(w_n, \mu)\}, \end{aligned}$$

and  $G = \nabla_x^2 L(w)$  or  $G$  is an approximation to the Hessian  $\nabla_x^2 L(w)$ .

As in the primal-dual interior point method, we can solve the above set of equations by directly solving (18) or by solving

$$(21) \quad (G + U(x, \mu)^{-1}V(w, \mu))\Delta x - A^t \Delta y = -\nabla f(x) + A^t y - \rho H'(x, \mu)e,$$

$$(22) \quad A\Delta x = -g(x)$$

for  $\Delta x$  and  $\Delta y$ , and then

$$(23) \quad \Delta z = -z - \rho H'(x, \mu)e - U(x, \mu)^{-1}V(w, \mu)\Delta x.$$

The following lemma gives basic properties of the iteration vector  $\Delta w$  and is apparent from (21) to (23).

LEMMA 1. *Suppose that  $\Delta w$  satisfies (18) at a point  $w$ .*

- (i) *If  $\Delta w = 0$ , then the point  $w$  is a KKT point that satisfies (15).*
- (ii) *If  $\Delta x = 0$ , then the point  $(x, y + \Delta y, z + \Delta z)$  is a KKT point that satisfies (15).*

In order to generate a descent direction in the following (see Lemma 3), we need to have a positive definite  $V(w, \mu)$ . The next lemma shows a condition for this property to hold.

LEMMA 2. *If  $\mu \neq 0$  and  $0 \leq z_i \leq \rho$ , then  $v(w_i, \mu) \in (0, \rho)$  for  $i = 1, \dots, n$ .*

*Proof.* Because  $z_i - \rho/2 \in [-\rho/2, \rho/2]$  and  $x_i/\sqrt{x_i^2 + \mu^2} \in (-1, 1)$ , we have

$$-\frac{\rho}{2} < \frac{x_i(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} < \frac{\rho}{2}.$$

From (20), we obtain the desired result. □

In the following, we will describe methods that use a line search algorithm and a trust region algorithm, respectively, to obtain an approximate KKT point. To ensure global convergence of the proposed algorithms that use the Newton iteration, we need a merit function. To this purpose, the penalty function

$$(24) \quad F(x) = f(x) + \rho \sum_{i=1}^n h(x_i, \mu) + \rho' \sum_{i=1}^m |g_i(x)|$$

will be used throughout the paper. In the above,  $\rho' > 0$  also serves as a penalty parameter that controls the equality constraints violation. We delete the dependence of the function  $F$  to the parameters  $\rho, \rho'$ , and  $\mu$  for notational simplicity in the following.

We denote the first order and second order approximation to  $F(x + s)$  by  $F_l(x, s)$  and  $F_q(x, s)$ , respectively, i.e.,

$$\begin{aligned} F_l(x, s) &= F(x) + \nabla f(x)^t s + \rho \sum_{i=1}^n h'(x_i, \mu) s_i \\ &\quad + \rho' \sum_{i=1}^m (|g_i(x) + \nabla g_i(x)^t s| - |g_i(x)|), \\ F_q(x, s) &= F_l(x, s) + \frac{1}{2} s^t Q s, \end{aligned}$$

where

$$Q = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 g_i(x) + \rho H''(x, \mu),$$

$$H''(x, \mu) = \text{diag} \{h''(x_1, \mu), \dots, h''(x_n, \mu)\}.$$

We also need the differences of these quantities with respect to the value  $F(x)$ :

$$\begin{aligned} \Delta F_l(x, s) &= F_l(x, s) - F(x), \\ \Delta F_q(x, s) &= F_q(x, s) - F(x). \end{aligned}$$

The following lemma plays a key role later in the proof of the global convergence property (Theorem 2) because it gives a condition for the Newton direction being a descent direction of the merit function  $F$ .

LEMMA 3. *Suppose that  $\Delta w$  satisfies (18) at a point  $w$ . Then there holds*

$$\begin{aligned} (25) \quad \Delta F_l(x, \Delta x) &\leq -\Delta x^t (G + U(x, \mu)^{-1} V(w, \mu)) \Delta x \\ &\quad - (\rho' - \|y + \Delta y\|_\infty) \sum_{i=1}^m |g_i(x)|. \end{aligned}$$

If  $\mu \neq 0$ ,  $\rho' \geq \|y + \Delta y\|_\infty$ , and  $G$  is positive semidefinite, then  $\Delta F_l(x, \Delta x) \leq 0$ , and  $\Delta F_l(x, \Delta x) = 0$  yields  $\Delta x = 0$ .

*Proof.* From (21) and (22) we have

$$\begin{aligned} \Delta F_l(x; \Delta x) &= \nabla f(x)^t \Delta x + \rho \sum_{i=1}^n h'(x_i, \mu) \Delta x_i - \rho' \sum_{i=1}^m |g_i(x)| \\ &= -\Delta x^t (G + U(x, \mu)^{-1} V(w, \mu)) \Delta x + \Delta x^t A(x)^t (y + \Delta y) - \rho' \sum_{i=1}^m |g_i(x)| \\ &= -\Delta x^t (G + U(x, \mu)^{-1} V(w, \mu)) \Delta x - (y + \Delta y)^t g(x) - \rho' \sum_{i=1}^m |g_i(x)|. \end{aligned}$$

This equality gives the desired result (25).

The proof of the second statement is easy because two terms in (25) are nonpositive by the assumption.  $\square$

Let  $F'(x; s)$  be a directional derivative of the function  $F(x)$  along an arbitrary given direction  $s \in \mathbb{R}^n$ ,

$$F'(x; s) = \lim_{\alpha \downarrow 0} \frac{F(x + \alpha s) - F(x)}{\alpha}.$$

Then the following lemma holds.

LEMMA 4. *Let  $s \in \mathbb{R}^n$  be given. Then the following assertions hold.*

- (i) *The function  $F_l(x, \alpha s)$  is convex with respect to the variable  $\alpha$ .*
- (ii) *There holds the relation*

$$(26) \quad F(x) + F'(x; s) \leq F_l(x; s).$$

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

$$(27) \quad F(x + s) \leq F(x) + F'(x + \theta s, s).$$

*Proof.* We can prove the lemma by the same way as the proof of Lemma 2 in [12].  $\square$

**3. Line search algorithm.** In this section, we describe an algorithm that uses line searches, and we prove its global convergence. The algorithm is similar to the interior point method proposed by Yamashita [12]. The basic iteration of the line search algorithm may be described as

$$(28) \quad w_{k+1} = w_k + \Lambda_k \Delta w_k,$$

where  $\Lambda_k = \text{diag}(\alpha_{xk} I_n, \alpha_{yk} I_m, \alpha_{zk} I_n)$  is composed of the step sizes in the  $x$ ,  $y$ , and  $z$  variables.

The main iteration is to decrease the value of the merit function  $F(x)$ . 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 to satisfy the condition  $0 \leq z \leq \rho$ . The explicit rules follow in order.

We adopt Armijo's rule as the line search rule for the variable  $x$ . In contrast to the interior point methods, where the primal variable  $x$  should always satisfy the positivity condition, there is no such restriction here. Therefore, Armijo's step size rule is the same as in the unconstrained optimization. The step to the next iterate is given by  $\alpha_{xk} = \beta^{l_k}$ , where  $\beta \in (0, 1)$  is a fixed constant and  $l_k$  is the smallest nonnegative integer such that

$$(29) \quad F(x_k + \beta^{l_k} \Delta x_k) - F(x_k) \leq \varepsilon_0 \beta^{l_k} \Delta F_l(x_k),$$

where  $\varepsilon_0 \in (0, 1)$ . Typical values of the parameters are  $\beta = 0.5$  and  $\varepsilon_0 = 10^{-6}$ . If  $G$  is positive semidefinite and  $\|y + \Delta y\|_\infty \leq \rho'$ , then  $\Delta F_l(x_k, \Delta x_k) \leq 0$  by Lemma 3.

For the variable  $z$ , we always force  $z$  to satisfy the condition  $0 \leq z \leq \rho$ . If the value  $\alpha_{zk} = 1$  violates the condition  $0 \leq z_k + \Delta z_k \leq \rho$ , then the step size is reduced to satisfy the condition, i.e.,

$$\alpha_{zk} = \min_i \left\{ \max_{\alpha_i} \{ \alpha_i | 0 \leq (z_k)_i + \alpha_{zk} (\Delta z_k)_i \leq \rho, 0 \leq \alpha_{zk} \leq 1 \} \right\}.$$

In our implementation of the algorithm explained below, if  $(z_k)_i$  is at the boundary, i.e.,  $(z_k)_i = 0$  or  $(z_k)_i = \rho$ , and if  $\alpha_{zk} = 0$  from the above step size calculation, we project  $\Delta z_k$  along the boundary by setting the corresponding  $(\Delta z_k)_i = 0$ . This procedure is not necessary for the global convergence proof given below but is adopted for better actual performance.

For the variable  $y$ , there exist two choices for the step length:

$$(30) \quad \alpha_{yk} = 1 \quad \text{or} \quad \alpha_{zk}.$$

The global convergence property given below holds for both choices.

The following algorithm describes the iterations for fixed  $\mu > 0$ ,  $\rho > 0$ , and  $\rho' > 0$ . We note that this algorithm corresponds to Step 2 of Algorithm EP in section 1, and the parameter  $\varepsilon'$  in Step 0 of the following algorithm corresponds to the quantity  $M_c \mu_k$  in (17).

**Algorithm LS**

**Step 0.** (Initialize) Let  $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n_\rho$ , where  $\mathbb{R}^n_\rho = \{z \in \mathbb{R}^n \mid 0 \leq z_i \leq \rho, i = 1, \dots, n\}$ , and  $\mu > 0, \rho > 0, \rho' > 0$ . Set  $\varepsilon' > 0, \beta \in (0, 1), \varepsilon_0 \in (0, 1)$ . Let  $k = 0$ .

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

**Step 2.** (Compute direction) Calculate the direction  $\Delta w_k$  by (18).

**Step 3.** (Step size) Find the smallest nonnegative integer  $l_k$  that satisfies

$$F(x_k + \beta^{l_k} \Delta x_k) - F(x_k) \leq \varepsilon_0 \beta^{l_k} \Delta F_l(x_k, \Delta x_k).$$

Calculate

$$\alpha_{xk} = \beta^{l_k},$$

$$\alpha_{zk} = \min_i \left\{ \max_{\alpha_i} \{ \alpha_i \mid 0 \leq (z_k)_i + \alpha_i (\Delta z_k)_i \leq \rho, 0 \leq \alpha_i \leq 1 \} \right\},$$

$$\alpha_{yk} = 1 \quad \text{or} \quad \alpha_{zk},$$

$$\Lambda_k = \text{diag}\{\alpha_{xk} I_n, \alpha_{yk} I_m, \alpha_{zk} I_n\}.$$

**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 GLS**

- (1) The functions  $f$  and  $g_i, i = 1, \dots, m$ , are twice continuously differentiable.
- (2) The level set of the function  $F(x)$  at an initial point  $x_0 \in \mathbb{R}^n$ , which is defined by  $\{x \in \mathbb{R}^n \mid F(x) \leq F(x_0)\}$ , is compact.
- (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 over all  $k$ .
- (5) The penalty parameter  $\rho'$  satisfies  $\rho' \geq \|y_k + \Delta y_k\|_\infty$  for each  $k = 0, 1, \dots$  □

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 derivatives of functions in Assumption GLS(1). We also note that if  $\Delta F_l(x_k, \Delta x_k) = 0$  at 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 KKT point from Lemmas 1 and 3. Therefore, in the following, we may assume  $\Delta F_l(x_k, \Delta x_k) < 0$  for all  $k$  if an infinite sequence is generated by Algorithm LS. The following theorem gives a convergence of an infinite sequence generated by Algorithm LS.

**THEOREM 2.** *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 KKT point.*

*Proof.* Because  $\Delta F_l(x_k, \Delta x_k) < 0$  by the Armijo rule adopted in the line searches, the sequence  $\{F(x_k)\}$  is strictly decreasing. Therefore, by Assumption GLS(2), the sequence  $\{x_k\}$  is bounded and has at least one accumulation point. The sequence  $\{z_k\}$  is also bounded. Thus there exists a positive number  $M$  such that

$$(31) \quad \frac{\|p\|^2}{M} \leq p^t(G_k + U(x_k, \mu)^{-1}V(w_k, \mu))p \leq M\|p\|^2 \quad \text{for all } p \in \mathbb{R}^n$$

because  $G_k$  is positive semidefinite by Assumption GLS(4), and the sequence  $\{U(x_k, \mu)^{-1}V(w_k, \mu)\}$  is strictly positive definite for bounded  $\{x_k\}$  by (16) and Lemma 2. From (29), (25), and (31), we have

$$(32) \quad F(x_{k+1}) - F(x_k) \leq \varepsilon_0 \beta^{l_k} \Delta F_l(x_k, \Delta x_k) \leq -\varepsilon_0 \beta^{l_k} \frac{\|\Delta x_k\|^2}{M} < 0.$$

The left-hand side of the above inequalities tends to zero since the sequence  $\{F(x_k)\}$  is decreasing and bounded below. Therefore, if there exists a number  $N > 0$  such that  $l_k < N$  for all  $k$  in a subsequence of  $\{0, 1, \dots\}$ , then  $\Delta x_k \rightarrow 0$  in this subsequence from (32). Now suppose that there exists a subsequence  $K \subset \{0, 1, \dots\}$  such that  $l_k \rightarrow \infty, k \in K$ . Then 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 (29), and we have

$$(33) \quad F(x_k + \alpha_{xk} \Delta x_k / \beta) - F(x_k) > \varepsilon_0 \alpha_{xk} \Delta F_l(x_k, \Delta x_k) / \beta.$$

By Lemma 4, there exists a  $\theta_k \in (0, 1)$  such that

$$(34) \quad \begin{aligned} F(x_k + \alpha_{xk} \Delta x_k / \beta) - F(x_k) &\leq \alpha_{xk} F'(x_k + \theta_k \alpha_{xk} \Delta x_k / \beta, \Delta x_k) / \beta \\ &\leq \alpha_{xk} \Delta F_l(x_k + \theta_k \alpha_{xk} \Delta x_k / \beta, \Delta x_k) / \beta, k \in K. \end{aligned}$$

Now from (33) and (34), we have

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

This inequality yields

$$(35) \quad \begin{aligned} \Delta F_l(x_k + \theta_k \alpha_{xk} \Delta x_k / \beta, \Delta x_k) - \Delta F_l(x_k, \Delta x_k) \\ > (\varepsilon_0 - 1) \Delta F_l(x_k, \Delta x_k) > 0. \end{aligned}$$

Because  $\Delta x_k$  satisfies (21) and (22) and there holds (31), by Assumption GLS(3),  $\|\Delta x_k\|$  is uniformly bounded above. Then by the assumption  $l_k \rightarrow \infty, k \in K$ , we have  $\|\theta_k \alpha_{xk} \Delta x_k / \beta\| \rightarrow 0, k \in K$ . Thus the left-hand side of (35) and therefore

$\Delta F_l(x_k, \Delta x_k)$  converge to zero when  $k \rightarrow \infty, k \in K$ . This yields  $\Delta x_k \rightarrow 0, k \in K$  because we have

$$\Delta F_l(x_k, \Delta x_k) \leq -\frac{\|\Delta x_k\|^2}{M} < 0$$

also from (25) and (31).

Now we proved  $\Delta x_k \rightarrow 0$ . Let an arbitrary accumulation point of the sequence  $\{x_k\}$  be  $\hat{x} \in \mathbb{R}^n$ , and let  $x_k \rightarrow \hat{x}, k \in K$  for a subsequence  $K \subset \{0, 1, \dots\}$ . Thus

$$(36) \quad x_k \rightarrow \hat{x}, \quad \Delta x_k \rightarrow 0, \quad x_{k+1} \rightarrow \hat{x}, \quad k \in K.$$

Because  $\{U(x_k, \mu)^{-1}V(w_k, \mu)\}$  is bounded, we have

$$\lim_{k \rightarrow \infty} \|z_k + \Delta z_k + \rho H'(x_k, \mu)e\| = 0$$

from (23). If we define  $\hat{z} = -\rho H'(\hat{x}, \mu)e$ , then  $0 < \hat{z} < \rho$ , and

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

This shows that the point  $z_k + \Delta z_k$  is always accepted as  $z_{k+1}$  (i.e.,  $\alpha_{z_k} = 1$ ) for sufficiently large  $k \in K$ . Since  $\alpha_{z_k} = 1$  is accepted for  $k \in K$  sufficiently large, so is  $\alpha_{y_k} = 1$ . Therefore we obtain

$$\lim_{k \rightarrow \infty, k \in K} \nabla_x L(\hat{x}, y_k + \Delta y_k, \hat{z}) = 0.$$

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

$$\begin{aligned} \nabla_x L(\hat{x}, \hat{y}, \hat{z}) &= 0, \\ g(\hat{x}) &= 0, \\ U(\hat{x}, \mu)\hat{z} &= \rho H(\hat{x}, \mu)e, \quad 0 < \hat{z} < \rho. \end{aligned}$$

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.  $\square$

**4. Trust region algorithm.** In this section, we describe an algorithm that uses trust region type iterations. The basic algorithm is the same as the primal-dual interior point trust region method proposed by Yamashita, Yabe, and Tanabe [14].

As in [14], we define a reference direction that will be used to form the actual step with Newton’s direction and to obtain the global convergence property of the algorithm by

$$(37) \quad \begin{pmatrix} D & -A(x)^t & -I \\ A(x) & 0 & 0 \\ V(w, \mu) & 0 & U(x, \mu) \end{pmatrix} \begin{pmatrix} \Delta x_{SD} \\ \Delta y_{SD} \\ \Delta z_{SD} \end{pmatrix} = -r(w, \mu),$$

where  $D$  is a positive definite possibly diagonal matrix. We call the direction  $\Delta w_{SD} = (\Delta x_{SD}, \Delta y_{SD}, \Delta z_{SD})^t$  the steepest descent direction by an analogy with the case in unconstrained optimization. We note that the direction is not the steepest descent for  $D \neq I$ . However, we use the term “steepest descent” in the following mainly because we use positive diagonal  $D$  in our implementation.

Replacing  $G$  by  $D$  in Lemma 3, we have

$$(38) \quad \begin{aligned} \Delta F_l(x; \Delta x_{SD}) &\leq -\Delta x_{SD}^t (D + U(x, \mu)^{-1} V(w, \mu)) \Delta x_{SD} \\ &\quad - (\rho' - \|y + \Delta y_{SD}\|_\infty) \sum_{i=1}^m |g_i(x)|. \end{aligned}$$

In the following, we assume  $\rho' > \|y + \Delta y_{SD}\|_\infty$  so that  $\Delta F_l(x; \Delta x_{SD}) \leq 0$  is satisfied. Then the vector  $\Delta x_{SD}$  is a descent direction of the merit function  $F(x)$ .

A trust region algorithm that finds a KKT point for a fixed  $\mu$  may proceed as follows. At iteration  $k$ , let us assume that the trust region radius  $\delta_k > 0$  and that the vectors  $\Delta w_k$  and  $\Delta w_{SDk}$  are given. From these two vectors, the step  $s_k$  that satisfies the trust region constraint  $\|s_k\| \leq \delta_k$  will be calculated. The step  $s_k$  must satisfy

$$(39) \quad \Delta F_q(x_k; s_k) \leq \frac{1}{2} \Delta F_q(x_k; \alpha^*(x_k, \Delta x_{SDk}) \Delta x_{SDk}),$$

where  $\alpha^*(x, d)$  is defined by

$$(40) \quad \alpha^*(x, d) = \arg \min \{F_q(x; \alpha d) \mid \|\alpha d\| \leq \delta\}$$

for  $x \in \mathbb{R}^n, d \in \mathbb{R}^n$ . The step size  $\alpha^*(x, d)$  gives a minimum point of the function  $F_q$  along the direction  $d$  in the interval defined by the trust region radius  $\delta$ . Therefore, condition (39) is a sufficient decrease condition based on the steepest descent step.

Now we present an algorithm of the trust region type method as follows.

**Algorithm TR**

**Step 0.** An initial point  $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_\rho^n$  and positive parameters  $\mu, \rho$ , and  $\rho'$  are given. Set parameters  $\varepsilon' > 0, \delta_0 > 0$ , and set  $k = 0$ .

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

**Step 2.** Calculate the vectors  $\Delta w_k$  and  $\Delta w_{SDk}$  that satisfy (18) and (37), respectively. If  $G_k = \nabla_x^2 L(w_k)$  gives a too large vector that does not satisfy the first inequality of (42) given below,  $G_k$  is modified to satisfy (42) by adding an appropriate positive diagonal matrix.

**Step 3.** Calculate a direction  $s_k \in \mathbb{R}^n$  that satisfies the conditions

$$(41) \quad \|s_k\| \leq \delta_k,$$

$$\Delta F_q(x_k, s_k) \leq \frac{1}{2} \Delta F_q(x_k, \alpha^*(x_k, \Delta x_{SDk}) \Delta x_{SDk}).$$

**Step 4.** Update the trust region radius  $\delta_{k+1}$  by the following:

$$\text{If } \Delta F(x_k, s_k) > \frac{1}{4} \Delta F_q(x_k, s_k), \text{ then } \delta_{k+1} = \frac{1}{2} \delta_k;$$

$$\text{If } \Delta F(x_k, s_k) \leq \frac{3}{4} \Delta F_q(x_k, s_k), \text{ then } \delta_{k+1} = 2\delta_k;$$

$$\text{otherwise, } \delta_{k+1} = \delta_k,$$

$$\text{where } \Delta F(x_k, s_k) = F(x_k + s_k) - F(x_k).$$

**Step 5.** If  $\Delta F(x_k, s_k) \leq 0$ , then set  $x_{k+1} = x_k + s_k$ , compute  $\alpha_{yk}$  and  $\alpha_{zk}$ , and set  $y_{k+1} = y_k + \alpha_{yk} \Delta y_k$  and  $z_{k+1} = z_k + \alpha_{zk} \Delta z_k$ . Otherwise, set  $w_{k+1} = w_k$ .

**Step 6.** Set  $k = k + 1$ , and return to Step 1. □

In the above algorithm, step sizes for the variables  $y$  and  $z$  are determined according to the rule of the previous section.

Before proving global convergence of Algorithm TR, we list the necessary assumptions.

**Assumption GTR**

- (1) The functions  $f$  and  $g_i, i = 1, \dots, m$ , are twice continuously differentiable.
- (2) The level set of the merit function at an initial point  $x_0 \in \mathbb{R}^n$  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  $D$  is uniformly positive definite and uniformly bounded. The matrix  $G$  is uniformly bounded.
- (5) There exists a number  $M > 0$  such that

$$(42) \quad \|\Delta x_k\| \leq M \|\Delta x_{SDk}\|, \|s_k\| \leq M \|\Delta x_{SDk}\|$$

for each  $k = 0, 1, \dots$ .

- (6) The penalty parameter  $\rho'$  satisfies  $\rho' \geq \|y_k + \Delta y_{SDk}\|_\infty$  for each  $k = 0, 1, \dots$ .  
 $\square$

It follows from Assumption GTR that the linear system of equations (37) has a unique solution and that the direction  $\Delta x_{SDk}$  is uniformly bounded on the compact level set defined in GTR(2). The following lemma shows the basic property of the search directions.

LEMMA 5. (1) *If  $\Delta w_k = 0$  or  $\Delta w_{SDk} = 0$  at a point  $w_k$ , then the point  $w_k$  satisfies the KKT conditions.*

- (2) *If  $\Delta x_k = 0$ , then  $\Delta x_{SDk} = 0$ .*
- (3) *If  $\Delta x_{SDk} = 0$ , then  $\Delta x_k = 0$  and  $s_k = 0$ .*
- (4) *If  $\Delta x_k = 0$ , then  $\alpha_{zk} = 1$  and  $\alpha_{yk} = 1$  are adopted in Algorithm TR, and the point  $w_{k+1}$  satisfies the barrier KKT conditions.*

*Proof.* (1) It is clear from (18) and (37).

- (2) Since  $(0, \Delta y_k, \Delta z_k)^t$  satisfies (37) and the coefficient matrix of (37) is nonsingular, the uniqueness of the solution to (37) implies  $\Delta x_{SDk} = 0$ .
- (3) This follows from GTR(5).
- (4) If  $\Delta x_k = 0$ , then by (23) we have

$$z_k + \Delta z_k = -\rho H'(x_k, \mu)e \in (0, \rho).$$

This implies that the stepsize  $\alpha_{zk} = 1$  is accepted, and so is  $\alpha_{yk} = 1$ . Then it follows from (21)–(23) that  $w_{k+1} = (x_k, y_k + \Delta y_k, z_k + \Delta z_k)$  satisfies the KKT conditions. Therefore, the lemma is proved.  $\square$

Now we prove the global convergence property of the above algorithm. From the above lemma, we observe that if  $\Delta x_{SDk} = 0$  at some iteration  $k$ , then the next point  $w_{k+1}$  is a KKT point. Therefore, we will assume that  $\Delta x_{SDk} \neq 0$  for each  $k = 0, 1, \dots$  in the following.

We state the following simple lemma first.

LEMMA 6. *If a vector  $d \in \mathbb{R}^n$  satisfies*

$$g(x) + A(x)d = 0,$$

*then there holds the relation*

$$\Delta F_l(x; \alpha d) = \alpha \Delta F_l(x; d), \quad \alpha \in [0, 1].$$

*Proof.* Since  $g_i(x) + \nabla g_i(x)^t d = 0$  for all  $i$ , we have

$$\begin{aligned} \Delta F_l(x; \alpha d) &= \alpha(\nabla f(x) + \rho H'(x, \mu)e)^t d + \rho' \sum_{i=1}^m ((1 - \alpha)|g_i(x)| - |g_i(x)|) \\ &= \alpha \left[ (\nabla f(x) + \rho H'(x, \mu)e)^t d + \rho \sum_{i=1}^m (|g_i(x) + \nabla g_i(x)^t d| - |g_i(x)|) \right]. \end{aligned}$$

Thus the proof is complete.  $\square$

LEMMA 7. Let  $x \in \mathbb{R}^n$ ,  $0 \neq d \in \mathbb{R}^n$ , and  $\delta > 0$  be given. Assume that  $\Delta F_l(x, d) < 0$  and that

$$g(x) + A(x)d = 0.$$

Then the step size defined by (40) can be expressed as

$$(43) \quad \alpha^*(x, d) = \min \left\{ 1, \frac{\delta}{\|d\|}, -\frac{\Delta F_l(x; d)}{\max\{d^t G d, 0\}} \right\},$$

where the last term in the braces in the right-hand side is assumed to give the value  $\infty$  if the value of the denominator is 0. Further we have

$$(44) \quad \Delta F_q(x; \alpha^*(x, d)d) \leq \frac{1}{2} \alpha^*(x, d) \Delta F_l(x; d).$$

*Proof.* By the definition of the function  $F_q$  and Lemma 6, we have

$$(45) \quad F_q(x, \alpha d) = F(x, \mu) + \alpha \Delta F_l(x; d) + \frac{1}{2} \alpha^2 d^t Q d, \quad \alpha \in [0, 1].$$

Suppose that  $d^t Q d > 0$  for the moment. Then the unconstrained minimum  $\hat{\alpha}$  of the function in the right-hand side of the above equality is calculated by

$$\hat{\alpha} = -\frac{\Delta F_l(x, d)}{d^t Q d}.$$

Therefore we obtain

$$(46) \quad \alpha^*(x, d) = \min \left\{ \frac{\delta}{\|d\|}, -\frac{\Delta F_l(x, d)}{d^t Q d} \right\}$$

in this case. From this relation, we have

$$(47) \quad d^t Q d \leq -\frac{\Delta F_l(x; d)}{\alpha^*(x, d)}.$$

From (45) and (47), we deduce

$$\begin{aligned} \Delta F_q(x; \alpha^*(x, d)d) &= \alpha^*(x, d) \Delta F_l(x; d) + \frac{1}{2} \alpha^*(x, d)^2 d^t Q d \\ &\leq \alpha^*(x, d) \Delta F_l(x; d) - \frac{1}{2} \alpha^*(x, d) \Delta F_l(x; d) \\ &= \frac{1}{2} \alpha^*(x, d) \Delta F_l(x; d). \end{aligned}$$

If  $d^t Q d \leq 0$ , we have

$$\alpha^*(x, d) = \min \left\{ 1, \frac{\delta}{\|d\|} \right\}$$

and

$$\begin{aligned} \Delta F_q(x; \alpha^*(x, d)d) &= \alpha^*(x, d)\Delta F_l(x; d) + \frac{1}{2}\alpha^*(x, d)^2 d^t Q d \\ &\leq \frac{1}{2}\alpha^*(x, d)\Delta F_l(x; d). \end{aligned}$$

Therefore we proved (43) and (44).  $\square$

**THEOREM 3.** *Let an infinite sequence  $\{w_k\}$  be generated by Algorithm TR for fixed  $\mu > 0$  and  $\rho > 0$ . Then there exists an accumulation point that satisfies the KKT conditions (15).*

*Proof.* By Step 3 of Algorithm TR and by Lemma 7, we have

$$(48) \quad \Delta F_q(x_k, s_k) \leq \frac{1}{4}\Delta F_l(x_k, \Delta x_{SDk}) \min \left\{ \frac{\delta_k}{\|\Delta x_{SDk}\|}, -\frac{\Delta F_l(x_k, \Delta x_{SDk})}{\max \{\Delta x_{SDk}^t Q_k \Delta x_{SDk}, 0\}} \right\}.$$

We define subsequences  $K_1 \subset \{0, 1, \dots\}$  and  $K_2 \subset \{0, 1, \dots\}$  that satisfy  $K_1 \cup K_2 = \{0, 1, 2, \dots\}$  and  $K_1 \cap K_2 = \emptyset$  by

$$(49) \quad \Delta F(x_k, s_k) > \frac{1}{4}\Delta F_q(x_k, s_k), \quad k \in K_1,$$

$$(50) \quad \Delta F(x_k, s_k) \leq \frac{1}{4}\Delta F_q(x_k, s_k), \quad k \in K_2.$$

(i) Suppose that  $K_1$  is an infinite sequence.

(i-a) If  $\liminf_{k \rightarrow \infty, k \in K_1} \delta_k = 0$ , then there exists an infinite set  $K'_1 \subset K_1$  such that  $\delta_k \rightarrow 0, k \in K'_1$ . Then because  $\|s_k\| \leq \delta_k$ , we have  $\|s_k\| \rightarrow 0, k \in K'_1$ . Suppose  $\liminf_{k \rightarrow \infty} \|\Delta x_{SDk}\| > 0$ . Then Assumption GTR (6) and (38) yield

$$\liminf_{k \rightarrow \infty, k \in K'_1} |\Delta F_l(x_k, \Delta x_{SDk})| > 0.$$

On the other hand, we have

$$\begin{aligned} \Delta F(x_k; s_k) &= \Delta F_l(x_k, s_k) + O(\|s_k\|^2) \\ &= \Delta F_q(x_k, s_k) + O(\|s_k\|^2). \end{aligned}$$

From (49) and the above relation, we have

$$-\Delta F_q(x_k, s_k) < O(\|s_k\|^2).$$

However, this contradicts (48) because it gives the relation

$$-\Delta F_q(x_k, s_k) \geq \frac{|\Delta F_l(x_k, \Delta x_{SDk})|}{4\|\Delta x_{SDk}\|} \|s_k\| = O(\|s_k\|)$$

for sufficiently large  $k \in K'_1$ . Thus we obtain  $\liminf_{k \rightarrow \infty} \|\Delta x_{SDk}\| = 0$  in this case.

(i-b) If  $\liminf_{k \rightarrow \infty, k \in K_1} \delta_k > 0$ , the condition  $\Delta F(x_k, s_k) \leq \frac{3}{4} \Delta F_q(x_k, s_k)$  must be satisfied infinitely many times for  $k \notin K_1$ , and this case corresponds to (ii) below.

(ii) Suppose that  $K_2$  is an infinite sequence.

(ii-a) Suppose that there exists an infinite sequence  $K'_2 \subset K_2$  such that  $\liminf_{k \rightarrow \infty, k \in K'_2} \delta_k > 0$ . Since  $\{F(x_k, \mu)\}$  is bounded below and decreasing and  $\Delta F(x_k, s_k) \leq 0$  for  $k \in K_2$ , we have

$$F(x_{k+1}, \mu) - F(x_k, \mu) = \Delta F(x_k, s_k) \rightarrow 0, \quad k \in K_2,$$

and thus  $\Delta F_q(x_k, s_k) \rightarrow 0, k \in K_2$ , from (50). Therefore, we have  $\Delta F_l(x_k, \Delta x_{SDk}) \rightarrow 0, k \in K'_2$ , from (48). Then, by (38), we obtain  $\Delta x_{SDk} \rightarrow 0, k \in K'_2$ , and thus  $\liminf_{k \rightarrow \infty} \|\Delta x_{SDk}\| = 0$  in this case.

(ii-b) Suppose  $\lim_{k \rightarrow \infty, k \in K_2} \delta_k = 0$ . Then the condition  $\Delta F(x_k, s_k) > \frac{1}{4} \Delta F_q(x_k, s_k)$  must be satisfied infinitely many times. This case corresponds to (i) above. If the case (i-a) holds, then (51) is proved as above. Otherwise, we prove that the case (i-b) does not occur in this case. Suppose that we have the case in which (i-b) occurs. Then  $\liminf_{k \rightarrow \infty, k \in K_1} \delta_k > 0$  and  $\lim_{k \rightarrow \infty, k \in K_2} \delta_k = 0$ . This is a contradiction because  $\delta_{k+1} = \delta_k, \frac{1}{2}\delta_k$  or  $2\delta_k$  for any  $k$ . Therefore, the case (i-b) does not occur.

Thus we proved

$$(51) \quad \liminf_{k \rightarrow \infty} \|\Delta x_{SDk}\| = 0.$$

By the requirement (42), this means that we have

$$\liminf_{k \rightarrow \infty} \|\Delta x_k\| = 0.$$

Thus there exists an infinite sequence  $K \subset \{0, 1, \dots\}$  and an accumulation point  $\hat{x} \in \mathbb{R}_+^n$  such that

$$x_k \rightarrow \hat{x}, \quad s_k \rightarrow 0, \quad \Delta x_k \rightarrow 0, \quad x_{k+1} \rightarrow \hat{x}, \quad k \in K.$$

Since Assumption GTR ensures the boundedness of  $\{U(x_k, \mu)^{-1}V(w_k, \mu)\}$ , we have

$$\lim_{k \rightarrow \infty, k \in K} \|z_k + \Delta z_k + \rho H'(x_k, \mu)e\| = 0.$$

If we define  $\hat{z} = -\rho H'(\hat{x}, \mu)e \in (0, \rho)$ , then we have

$$z_k + \Delta z_k \rightarrow \hat{z} \in (0, \rho), \quad k \in K,$$

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

Since  $\alpha_{z_k} = 1$  is accepted for  $k \in K$  sufficiently large, so is  $\alpha_{y_k} = 1$ . Because the matrix  $A(\hat{x})$  is of full rank, the sequence  $\{y_k + \Delta y_k\}, k \in K$  converges to a point  $\hat{y} \in \mathbb{R}^m$ . Thus we proved that  $(x_{k+1}, y_{k+1}, z_{k+1}) \rightarrow (\hat{x}, \hat{y}, \hat{z})$  for  $k \in K$  and that

$$\begin{aligned} \nabla f(\hat{x}) - A(\hat{x})^t \hat{y} - \hat{z} &= 0, \\ g(\hat{x}) &= 0, \\ U(\hat{x}, \mu) \hat{z} &= \rho H(\hat{x}, \mu)e, \quad 0 < \hat{z} < \rho. \end{aligned}$$

This completes the proof.  $\square$

The actual trust region step calculation is similar to the one described in [14] and is not described here. However, we note that in our method, the step  $s_k$  is proportional to a vector which is a convex combination of  $\Delta x_k$  and  $\Delta x_{SDk}$  and satisfies the second condition in (42).

**5. Superlinear/quadratic convergence.** In this section, we extend the algorithm of this paper so that it is superlinearly/quadratically convergent in addition to the global convergence property proved in the above. For this purpose, we add a procedure called a trial Newton step (see below) that checks if the Newton step gives a point  $w_{k+1}$  that satisfies the condition  $\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k^\eta, \eta \in (0, 1]$  for a given  $\mu_k$  with a single step. If it is satisfied, then we accept the point as a next iterate. If not, the minimization of the merit function by the line search or trust region algorithm given above is executed to obtain a point that satisfies the condition  $\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k^\eta, \eta \in (0, 1]$ . We note that the condition for the approximate KKT point here is looser than the condition in Algorithm EP for  $\mu_k < 1$  and  $\eta < 1$ . The procedure is described as Steps 2 and 3 of the following algorithm.

**Algorithm superlinearEP**

**Step 0.** (Initialize) Choose parameters  $\rho > 0, M_c > 0, \tau > 0, \eta \in (0, 1]$ , and  $\varepsilon > 0$ . Select an initial point  $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n_\rho$ . Let  $k = 0$ .

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

**Step 2.** (Trial Newton step) If  $\|r_0(w_k)\|$  is sufficiently small ( $w_k$  is close to a KKT point), execute the following steps. Otherwise, choose  $\mu_k \in (0, \mu_{k-1})$ , and go to Step 3.

**Step 2.1** Choose  $\mu_k = \Theta(\|r_0(w_k)\|^{1+\tau})$ . Calculate the direction  $\Delta w_k$  by

$$J(w_k, \mu_k)\Delta w_k = -r(w_k, \mu_k),$$

where

$$J(w_k, \mu_k) = \begin{pmatrix} \nabla_x^2 L(w_k) & -A(x_k)^t & -I \\ A(x_k) & 0 & 0 \\ V(w_k, \mu_k) & 0 & U(x_k, \mu_k) \end{pmatrix}.$$

If  $J(w_k, \mu_k)$  is singular, go to Step 3.

**Step 2.2** (Step size) Calculate the step size  $\alpha_k \in (0, 1]$  such that  $0 \leq z_k + \alpha_k \Delta z_k \leq \rho$ . First calculate the maximum step  $\bar{\alpha}_k$  to the constraints  $0 \leq z_k + \alpha_k \Delta z_k \leq \rho$  by

$$(52) \quad \bar{\alpha}_k = \min \left\{ \min_i \left\{ \frac{\rho - (z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i > 0 \right\}, \min_i \left\{ -\frac{(z_k)_i}{(\Delta z_k)_i} \mid (\Delta z_k)_i < 0 \right\} \right\}.$$

Then determine the step  $\alpha_k$  by

$$(53) \quad \alpha_k = \min \{1, \bar{\alpha}_k\}.$$

**Step 2.3** If  $\|r(w_k + \alpha_k \Delta w_k, \mu_k)\| \leq M_c \mu_k^\eta$ , then set  $w_{k+1} = w_k + \alpha_k \Delta w_k$ , and go to Step 4. Otherwise, go to Step 3.

**Step 3.** (Line search/trust region procedure) By using Algorithm LS or TR, find a point  $w_{k+1}$  that satisfies the condition

$$(54) \quad \|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k^\eta.$$

**Step 4.** Set  $k := k + 1$ , and go to Step 1.  $\square$

The global convergence of Algorithm superlinearEP is apparent from the previous sections. So in the following, we confine our discussion to the rate of local convergence issue. Therefore, the initial point  $w_0$ , in particular, is assumed to be close to a KKT point  $w^*$ . And we will prove that if  $\mu_k$  is updated by the rule in Step 2.1, then the point  $w_k + \alpha_k \Delta w_k$  satisfies the condition  $\|r(w_k + \alpha_k \Delta w_k, \mu_k)\| \leq M_c \mu_k^\eta$ , Step 3 is skipped, and the convergence rate of the sequence  $\{w_k\}$  is superlinear/quadratic under appropriate conditions. We list a few definitions and assumptions that are necessary in the following discussion.

**Definition**

- (i) The active constraint set at  $x$  is defined by a set composed of all equality constraints and a set of variables with  $x_i = 0$ .
- (ii) The second order sufficient condition for optimality at  $w^*$  is  $v^t \nabla_x^2 L(w^*) v > 0$  for all  $v \neq 0$  satisfying  $A(x^*)v = 0$  and  $v_i = 0, i \in \{i | x_i^* = 0\}$ .
- (iii) The strict complementarity condition of the solution  $w^*$  is that  $z_i^* \in (0, \rho)$  if  $x_i^* = 0$ .  $\square$

**Assumption L**

- (L1) The initial point  $w_0$  is sufficiently close to  $w^*$ .
- (L2) The second derivatives of the functions  $f$  and  $g$  are Lipschitz continuous at  $x^*$ .
- (L3) The linear independence of active constraint gradients, the second order sufficient condition for optimality, and the strict complementarity condition hold at  $w^*$ .  $\square$

We note that the strict complementarity condition above means that there exists a constant  $\beta \in (0, \rho/2)$  such that  $\beta \leq z_i^* \leq \rho - \beta$  if  $x_i^* = 0$ , i.e.,

$$|z_i^* - \rho/2| \leq \rho/2 - \beta \text{ if } x_i^* = 0.$$

LEMMA 8. *There exists a constant  $\delta > 0$  such that, if  $\|w - w^*\| \leq \delta$ , then the following estimates are valid:*

(i) If  $x_i^* < 0$ ,

$$\begin{aligned} u(x_i, \mu) &= |x_i| + O(\mu^2), \\ v(w_i, \mu) &= \rho - z_i + O(\mu^2). \end{aligned}$$

(ii) If  $x_i^* = 0$ ,

$$\begin{aligned} u(x_i, \mu) &= O(|x_i|) + O(\mu), \\ \left| v(w_i, \mu) - \frac{\rho}{2} \right| &\leq O(\|w - w^*\|) + \left| z_i^* - \frac{\rho}{2} \right|. \end{aligned}$$

(iii) If  $x_i^* > 0$ ,

$$\begin{aligned} u(x_i, \mu) &= |x_i| + O(\mu^2), \\ v(w_i, \mu) &= z_i + O(\mu^2). \end{aligned}$$

(iv)  $r(w, \mu) = r_0(w) + O(\mu)$ .

*Proof.* (i) The first estimate is obvious. The second estimate is derived from

$$\begin{aligned} v(w_i, \mu) &= \frac{x_i(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} + \frac{\rho}{2} = (-1 + O(\mu^2)) \left( z_i - \frac{\rho}{2} \right) + \frac{\rho}{2} \\ &= \rho - z_i + O(\mu^2). \end{aligned}$$

(ii) Since  $|z_i - \rho/2| \leq |z_i - z_i^*| + |z_i^* - \rho/2|$  and  $x_i/\sqrt{x_i^2 + \mu^2} \in (-1, 1)$ , we have

$$\left| v(w_i, \mu) - \frac{\rho}{2} \right| \leq O(\|w - w^*\|) + \left| z_i^* - \frac{\rho}{2} \right|.$$

(iii) This proof is similar to that of (i).

(iv) From the definition of  $r(w, \mu)$  and  $r_0(w)$ , we have

$$\begin{aligned} \|r(w, \mu) - r_0(w)\| &= O\left(\sum_i |(u(x_i, \mu) - |x_i|)z_i - \rho(h(x_i, \mu) - |x_i|_-)|\right) \\ &= O\left(\sum_i |(u(x_i, \mu) - |x_i|)(z_i - \rho/2)|\right) \\ &= O(\mu) \end{aligned}$$

from the above.  $\square$

Let  $\hat{J}(w, \hat{v})$  be defined by

$$\hat{J}(w, \hat{v}) = \begin{pmatrix} \nabla_x^2 L(w) & -A(x)^t & -I \\ A(x) & 0 & 0 \\ \hat{V} & 0 & \hat{U}(x) \end{pmatrix},$$

where  $\hat{v} \in \mathbb{R}^n$ ,  $\hat{U}(x) = \text{diag}(|x_1|, \dots, |x_n|)$ ,  $\hat{V} = \text{diag}(\hat{v}_1, \dots, \hat{v}_n)$ , and

$$\begin{aligned} \hat{v}_i &= 0, & x_i &\neq 0, \\ \hat{v}_i &> 0, & x_i &= 0 \end{aligned}$$

for  $i = 1, \dots, n$ . Then by Assumption (L3), it can be shown that the matrix  $\hat{J}(w^*, \hat{v})$  is nonsingular as in a usual Jacobian uniqueness condition. This fact is stated more precisely in the next lemma.

LEMMA 9. *Let  $\theta \in (0, \rho]$  be an arbitrary given constant. Then there exists a positive constant  $\xi$  such that*

$$\left\| \hat{J}(w^*, \hat{v})^{-1} \right\| \leq \xi$$

for any  $\hat{v}_i \in [\theta, \rho]$ ,  $i \in \{i \mid x_i^* = 0\}$ .

LEMMA 10. *There exists a constant  $\delta > 0$  such that, if  $\|w - w^*\| \leq \delta$ , then, for sufficiently small  $\mu > 0$ , there exists a positive constant  $\xi'$  such that*

$$(55) \quad \|J(w, \mu)^{-1}\| \leq \xi'.$$

*Proof.* From Lemma 8 and Assumption (L3), there exists  $\theta > 0$  and  $\hat{v}_i \in [\theta, \rho]$ ,  $i \in \{i \mid x_i^* = 0\}$  such that

$$\left\| J(w, \mu) - \hat{J}(w^*, \hat{v}) \right\| \leq O(\|w - w^*\|) + O(\mu)$$

for sufficiently small  $\delta$ . Thus, by the Banach perturbation lemma and by Lemma 9, we proved the lemma.  $\square$

LEMMA 11. *There exists a constant  $\delta > 0$  such that, if  $\|w - w^*\| \leq \delta$ , then*

$$r(w, \mu) + J(w, \mu)(w^* - w) = O(\mu) + O(\|w - w^*\|^2).$$

*Proof.* To prove the lemma, we evaluate the value of vector  $r(w, \mu) + J(w, \mu)(w^* - w)$ . The first two components that arise from  $\nabla L(w)$  and  $g(x)$  need no specific proof. So we consider only the third part. Therefore, we will prove

$$\begin{aligned} p_i &\equiv u(x_i, \mu)z_i - \frac{\rho}{2}(u(x_i, \mu) - x_i) + v(w_i, \mu)(x_i^* - x_i) - u(x_i, \mu)(z_i - z_i^*) \\ &= u(x_i, \mu) \left( z_i^* - \frac{\rho}{2} \right) + \frac{\rho}{2}x_i + v(w_i, \mu)(x_i^* - x_i) \\ &= O(\mu) + O(\|w - w^*\|^2) \end{aligned}$$

for each  $i$ .

(i) If  $x_i^* < 0$  ( $z_i^* = \rho$ ), then, from Lemma 8(i), we have

$$\begin{aligned} p_i &= (|x_i| + O(\mu^2)) \left( z_i^* - \frac{\rho}{2} \right) + \frac{\rho}{2}x_i + (\rho - z_i + O(\mu^2))(x_i^* - x_i) \\ &= O(\mu^2) + O(\|w - w^*\|^2). \end{aligned}$$

(ii) If  $x_i^* = 0$  ( $z_i^* \in (0, \rho)$ ), then, from Lemma 8(ii), we have

$$\begin{aligned} p_i &= \sqrt{x_i^2 + \mu^2} \left( z_i^* - \frac{\rho}{2} \right) - \frac{x_i^2(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} \\ &= \sqrt{x_i^2 + \mu^2} \left( z_i - \frac{\rho}{2} \right) + \sqrt{x_i^2 + \mu^2}(z_i^* - z_i) - \frac{x_i^2(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} \\ &= \frac{\mu^2(z_i - \rho/2)}{\sqrt{x_i^2 + \mu^2}} + O(\mu)O(\|w - w^*\|) + O(\|w - w^*\|^2) \\ &= \mu \frac{(z_i - \rho/2)}{\sqrt{(x_i/\mu)^2 + 1}} + O(\mu)O(\|w - w^*\|) + O(\|w - w^*\|^2) \\ &= O(\mu) + O(\|w - w^*\|^2). \end{aligned}$$

(iii) If  $x_i^* > 0$  ( $z_i^* = 0$ ), then, from Lemma 8(iii), we have

$$\begin{aligned} p_i &= (|x_i| + O(\mu^2)) \left( z_i^* - \frac{\rho}{2} \right) + \frac{\rho}{2}x_i + (z_i + O(\mu^2))(x_i^* - x_i) \\ &= O(\mu^2) + O(\|w - w^*\|^2). \end{aligned}$$

Thus the lemma is proved.  $\square$

LEMMA 12. *There exists a constant  $\delta > 0$  such that, if  $\|w - w^*\| \leq \delta$ , then, for sufficiently small  $\mu > 0$  and for  $i$  such that  $x_i^* < 0$ ,*

$$(56) \quad 0 > \Delta z_i = O(\mu^2)$$

when  $z_i = \rho$ , and

$$(57) \quad \frac{\Delta z_i}{\rho - z_i} = 1 + O(\mu^2), \quad \frac{\Delta z_i}{z_i} = O(\|r(w, \mu)\|) + O(\|w - w^*\|) + O(\mu^2)$$

when  $z_i < \rho$ ;

for  $i$  such that  $x_i^* = 0$ ,

$$(58) \quad \frac{\Delta z_i}{\rho - z_i} = O(\|r(w, \mu)\|), \frac{\Delta z_i}{z_i} = O(\|r(w, \mu)\|);$$

and for  $i$  such that  $x_i^* > 0$ ,

$$(59) \quad 0 < \Delta z_i = O(\mu^2)$$

when  $z_i = 0$ , and

$$(60) \quad \frac{\Delta z_i}{\rho - z_i} = O(\|r(w, \mu)\|) + O(\|w - w^*\|) + O(\mu^2), -\frac{\Delta z_i}{z_i} = 1 + O(\mu^2)$$

when  $z_i > 0$ .

*Proof.* We note that, for each  $i$ ,

$$v(w_i, \mu)\Delta x_i + u(x_i, \mu)\Delta z_i = -u(x_i, \mu)z_i + \frac{\rho}{2}(u(x_i, \mu) - x_i)$$

and that  $\|\Delta w\| = O(\|r(w, \mu)\|)$  by Lemma 10.

(i) If  $x_i^* < 0$  and  $z_i = \rho$ , we have

$$\begin{aligned} \Delta z_i &= -\frac{v(w_i, \mu)}{u(x_i, \mu)}\Delta x_i - z_i + \frac{\rho}{2}\left(1 - \frac{x_i}{u(x_i, \mu)}\right) \\ &= -\frac{\rho}{2}\left(\frac{x_i}{\sqrt{x_i^2 + \mu^2}} + 1\right)\left(\frac{\Delta x_i}{\sqrt{x_i^2 + \mu^2}} + 1\right), \end{aligned}$$

and (56) follows.

If  $x_i^* < 0$  and  $z_i < \rho$ , we have, from Lemma 8(i),

$$\begin{aligned} \Delta z_i &= -\frac{\rho - z_i + O(\mu^2)}{u(x_i, \mu)}\Delta x_i - z_i + \frac{\rho}{2}\left(1 - \frac{x_i}{u(x_i, \mu)}\right) \\ &= \frac{\rho - z_i}{x_i}\Delta x_i + \rho - z_i + O(\mu^2), \end{aligned}$$

and (57) follows.

(ii) If  $x_i^* = 0$ , we have (58) because of the strict complementarity assumption.

(iii) If  $x_i^* > 0$  and  $z_i = 0$ , we have

$$\Delta z_i = \frac{\rho}{2}\left(-\frac{x_i}{\sqrt{x_i^2 + \mu^2}} + 1\right)\left(\frac{\Delta x_i}{\sqrt{x_i^2 + \mu^2}} + 1\right),$$

and (59) follows.

If  $x_i^* > 0$  and  $z_i > 0$ , we have

$$\begin{aligned} \Delta z_i &= -\frac{z_i + O(\mu^2)}{u(x_i, \mu)}\Delta x_i - z_i + \frac{\rho}{2}\left(1 - \frac{x_i}{u(x_i, \mu)}\right) \\ &= -\frac{z_i}{x_i}\Delta x_i - z_i + O(\mu^2), \end{aligned}$$

and (60) follows.  $\square$

LEMMA 13. *There exists a constant  $\delta > 0$  such that, if  $\|w - w^*\| \leq \delta$ , then*

$$(61) \quad 1 - \alpha = O(\mu^2).$$

*Proof.* If there exists an  $i$  such that  $x_i^* \neq 0$ , then

$$1 - \bar{\alpha} = O(\mu^2)$$

from Lemma 12 and the definition of  $\bar{\alpha}$  (52). If not, we have  $\bar{\alpha} > 1$ . Thus, we have (61) from the definition of  $\alpha$  (53).  $\square$

Now we prove the superlinear convergence of Algorithm superlinearEP.

THEOREM 4. *If  $w_0$  is sufficiently close to  $w^*$  and  $\mu_k = \Theta(\|r_0(w_k)\|^{1+\tau})$  for  $\tau \in (0, 2/\eta - 1)$ , then the sequence  $\{w_k\}$  satisfies the condition  $\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k^\eta$  and converges to  $w^*$  superlinearly. If  $\tau \in [1, 2/\eta - 1)$ , then the convergence rate is quadratic.*

*Proof.* From Lemmas 13 and 8(iv), we have

$$\begin{aligned} \|r(w_k + \alpha_k \Delta w_k, \mu_k)\| &= \left\| r(w_k, \mu_k) + \alpha_k J(w_k, \mu_k) \Delta w_k + O(\|\alpha_k \Delta w_k\|^2) \right\| \\ &= \left\| (1 - \alpha_k) r(w_k, \mu_k) + O(\|r(w_k, \mu_k)\|^2) \right\| \\ &\leq O(\mu_k^2) O(\|r(w_k, \mu_k)\|) + O(\|r(w_k, \mu_k)\|^2) \\ &\leq O(\mu_k) O(\|r_0(w_k)\|) + O(\|r_0(w_k)\|^2) + O(\mu_k^2) \\ &= O(\mu_k^{1+1/(1+\tau)}) + O(\mu_k^{2/(1+\tau)}) + O(\mu_k^2) \\ &= O(\mu_k^{2/(1+\tau)}) \\ &\leq M_c \mu_k^\eta. \end{aligned}$$

The last inequality follows from  $2/(1 + \tau) > \eta$ .

Next we have

$$\begin{aligned} \|w_k + \alpha_k \Delta w_k - w^*\| &= \|w_k - w^* - \alpha_k J(w_k, \mu_k)^{-1} r(w_k, \mu_k)\| \\ &= \|J(w_k, \mu_k)^{-1} (J(w_k, \mu_k)(w_k - w^*) - \alpha_k r(w_k, \mu_k))\| \\ &\leq \|J(w_k, \mu_k)^{-1}\| \|J(w_k, \mu_k)(w_k - w^*) - \alpha_k r(w_k, \mu_k)\| \\ &= \|J(w_k, \mu_k)^{-1}\| \left\| (1 - \alpha_k) J(w_k, \mu_k)(w_k - w^*) + O(\mu_k) + O(\|w_k - w^*\|^2) \right\| \end{aligned}$$

from Lemma 11. Then we obtain from Lemmas 10 and 13

$$\begin{aligned} \|w_k + \alpha_k \Delta w_k - w^*\| &\leq |1 - \alpha_k| O(\|w_k - w^*\|) + O(\mu_k) + O(\|w_k - w^*\|^2) \\ &= O(\mu_k^2) O(\|w_k - w^*\|) + O(\mu_k) + O(\|w_k - w^*\|^2) \\ &= O(\|r_0(w_k)\|^{1+\tau}) + O(\|w_k - w^*\|^2) \\ &= O(\|w_k - w^*\|^{1+\tau}) + O(\|w_k - w^*\|^2). \end{aligned}$$

This proves the superlinear convergence if  $\tau > 0$  and the quadratic convergence if  $\tau \geq 1$ .

Therefore, the theorem is proved.  $\square$

**6. Numerical experiments.** In this section, we report results of various numerical experiments done with our implementation of the algorithm of this paper.

**6.1. Discussion on penalty parameters.** Before proceeding to the description of our numerical experiments, we first discuss issues related to penalty parameters. The most prominent feature of our problem (6) is the existence of the penalty parameter  $\rho$  which is not relevant to the original problem itself. Because of this parameter, problem (6) is not invariant under scalings of variables, for example. If  $\rho$  is not large enough, problem (6) may give infeasible or unbounded solutions even when problem (1) is feasible. These may be serious objections to our algorithm. However, we list a few arguments that support our research in the following.

(i) It is possible to extend our algorithm to solve problem (1) rather faithfully. To this end, we could modify Step 2 of Algorithm EP so that the value of  $\rho$  is increased for the next value of  $\mu$  if a component of  $x$  is negative or a component of  $z$  is close to  $\rho$  at the current approximate KKT point. With this modification, a feasible solution of problem (1) may be obtained if  $\rho$  remains finite, i.e., if it does not diverge to infinity. We did not include this modification to our algorithm to avoid too much complexity in our presentation, and we just performed numerical experiments on various values of  $\rho$ , as shown below.

(ii) It could be argued that the practical performance of our algorithm is seriously affected by the actual value of  $\rho$ . It is true that the performance varies with each value of  $\rho$ . But we will show in the following experiment that our algorithm is rather stable with parameter change and is not so sensitive to the parameter values.

(iii) Suppose that problem (1) does not have a feasible solution. Even in this case, it may be possible to find a “solution” to problem (6). And this “solution” may give valuable information about the original infeasible problem because our problem (6) is a “soft constraint” version of the original “hard constraint” problem. In practical optimization processes, it frequently occurs that some constraints should be relaxed in some way to find a feasible solution. Therefore, we believe it is of practical value to solve problems of the form (6), particularly with nonlinear constraints. In this respect, it is interesting to have an algorithm for problems that have a mix of constraint types from the problems (1) and (6), and we believe it is not difficult to have a mix of the exterior point type and the interior point type algorithms, for example.

There is one more penalty parameter  $\rho'$  which appears in our merit function:

$$F(x) = f(x) + \rho \sum_{i=1}^n h(x_i, \mu) + \rho' \sum_{i=1}^m |g_i(x)|.$$

Recent research pays attentions to updating methods of the parameters (see [2], [6]). In our algorithms, the value of  $\rho'$  is simply supposed to be sufficiently large as in Assumptions GLS(5) and GTR(6) for simplicity of description. However, it should be stated that in actual implementation, this value is adjusted with the progress of the optimization process. The value of  $\rho'$  is increased to satisfy the above assumption if necessary. If  $\rho'$  stays finite, we will obtain an approximate KKT point that corresponds to the current value of  $\mu$  under the assumptions of this paper. It is important to note that the value of  $\rho'$  can be decreased if it is judged to be too large compared with the value of  $\|y + \Delta y\|_\infty$  (line search algorithm) or  $\|y + \Delta y_{SD}\|_\infty$  (trust region algorithm). Theoretically this reduction should be done only finitely many times to ensure global convergence to a KKT point for fixed  $\mu$ . This method of penalty parameter control is explained in [14], so further explanation is omitted here. We believe our method of controlling the parameter  $\rho'$  is flexible and practical, as shown in the following and in [14].

**6.2. CUTER problems.** The proposed methods are programmed and tested. Adopted parameter values are  $M_c = 7.5$  (for an approximate KKT condition), the initial value of  $\rho' = 100$ ,  $\mu_0 = 1.0$ ,  $\eta = 1$ , and  $\tau = 0.6$  (for superlinear convergence). After an approximate KKT condition is obtained in Algorithm superlinearEP, i.e.,  $w_k$  satisfies

$$(62) \quad \|r(w_k, \mu_{k-1})\| \leq M_c \mu_{k-1},$$

we calculate  $\mu_k$  by

$$\mu_k = \max \left\{ \frac{\|r_0(w_k)\|}{M_\mu}, \frac{\mu_{k-1}}{10} \right\}, M_\mu = 10,$$

if  $\|r_0(w_k)\| \geq 10^{-2}$ , and execute Step 3 in Algorithm superlinearEP. If  $\|r_0(w_k)\| < 10^{-2}$ , we calculate  $\mu_k$  by

$$\mu_k = \max \left\{ \|r_0(w_k)\|^{1.6}, \frac{\mu_{k-1}}{100} \right\}$$

and try Steps 2.1–2.3 in Algorithm superlinear EP in order to obtain fast convergence in the final stage of iterations.

We report the results of numerical experiments on CUTER problems [7] with the trust region algorithm (superlinear version) here. From the CUTER problems, we select 23 problems with  $n + m > 1000$ , where  $n$  is the number of variables and  $m$  is the number of constraints. The number  $m$  does not count bound constraints. We exclude bound constrained problems, problems that are too easy, similar problems, and problems that are too difficult for the purpose of valuable comparative study.

We first show the results obtained by the interior point line search methods—IPOPT in [9]—for comparison in Table 6.1. An indication of OPT in the **stat** column denotes a feasible optimal solution has been found. Numbers in the **nitr** column denote total iteration counts executed. The numbers in the **res** column denote the final KKT condition residuals obtained.

In Table 6.2, results obtained by the present exterior point method (superlinear trust region method) with various values of  $\rho$  are summarized. EPM(10) to EPM(100000) denote the method with the corresponding value of  $\rho$ . EXT denotes the optimal but exterior solution obtained. **ext** in **nitr(ext)** denotes the number of iterations during which to compute an exterior point (violation of a bound larger than  $10^{-8}$ ).

In Table 6.2, eight exterior solutions (EXT) are obtained with  $\rho = 10$ , five EXTs are obtained with  $\rho = 100$ , two EXTs with  $\rho = 1000$ , and one EXT with  $\rho = 10000$ , and all other solutions are feasible. From the table, we see that the performance of EPM is comparable with that of IPOPT in our implementation. We see that at least in this experiment, fairly large values of  $\rho$  do not give serious performance deterioration, although the number of total iterations increases gradually with the penalty parameter value increase.

**6.3. Warm start and parametric programming.** As noted in the Introduction, it is not easy to utilize the warm/hot start of a given initial point with the interior point methods. However, the algorithm of this paper can enjoy this condition easily. To this end, we modify our basic algorithm to utilize a warm start initial point. We allow the parameter  $\mu$  to have “components”  $\mu_i, i = 1, \dots, n$ . In condition (62),  $\mu_{k-1}$  on the right-hand side is common to all components and is updated as stated above. We denote this parameter as  $\bar{\mu}$  here. On the other hand, we allow  $\mu_{k-1}$  on the

TABLE 6.1  
*CUTEr* results by IPOPT.

Problem	Problem		IPOPT		
	$n$	$m$	stat	nitr	res
AUG2DCQP	3280	1600	OPT	23	2.9e-09
AUG3DCQP	3873	1000	OPT	21	3.1e-09
BLOCKQP2	10010	5001	OPT	19	2.5e-09
BLOWEYC	2002	1002	OPT	9	2.5e-09
BRIDGEND	2734	2727	OPT	57	2.5e-09
CLNLBEAM	3003	2000	OPT	378	4.8e-09
CVXQP1	1000	500	OPT	19	2.9e-09
DTOC2	5998	3996	OPT	10	1.3e-09
HELSEBY	1408	1399	OPT	34	2.5e-09
HUESTIS	5000	2	OPT	27	7.8e-10
HYDROELL	1009	1008	OPT	235	2.7e-09
JANNSON4	10000	2	OPT	13	2.5e-09
NCVXQP1	1000	500	OPT	188	1.0e-09
ORTHDM2	4003	2000	OPT	6	9.1e-14
ORTHGDM	4003	2000	OPT	6	1.0e-09
PRIMAL4	1489	75	OPT	13	6.2e-09
READING7	1002	500	OPT	126	9.1e-10
SOSQP1	5000	2501	OPT	7	2.5e-09
SSNLBEAM	3003	2000	OPT	165	9.1e-10
STCQP1	4097	2052	OPT	14	4.7e-09
STNQP1	4097	2052	OPT	17	2.5e-09
YAO	2002	2000	OPT	27	4.3e-10
ZAMB2	1326	480	OPT	33	2.5e-09
TOTAL				1447	

left-hand side to differ with each component  $i$ . It is easy to extend our algorithm to this case and is not explained further.

Now we describe how to calculate  $\mu_i$  while the warm start optimization is performed. If  $(x_i, z_i)$  satisfy  $x_i > 0$  and  $z_i \in (0, \rho/2)$ , or  $x_i < 0$  and  $z_i \in (\rho/2, \rho)$  at an initial warm start point or at a later approximate KKT point, we can calculate  $\hat{\mu}_i$  that satisfies (14) analytically. For those components, we define  $\mu_i$  by

$$\mu_i = \min \{ \hat{\mu}_i, 10^3 \bar{\mu} \}.$$

Naturally we avoid too small values (e.g.,  $10^{-12}$ ) of  $\mu_i$  in our program. For other components, we simply set  $\mu_i = \bar{\mu}$ . The initial value of  $\bar{\mu}$  is set equal to  $10^{-8}$ .

In the experiment reported in Table 6.3, we perturb all the primal and dual solutions obtained from the cold start of the algorithm (above experiment with  $\rho = 10000$ ) with the maximum relative amounts from  $10^{-5}$  to  $10^{-1}$  by uniform random numbers, respectively, to create warm starts for various problems. More specifically, we set the initial point of the warm start  $x_{wi}$  of the variable  $x_i$  as

$$x_{wi} = x_i^* + \alpha r_i |x_{ci} - x_i^*|,$$

where  $x_i^*$  is an optimal value of  $x_i$  obtained by the cold start optimization from the cold start initial point  $x_{ci}$ ,  $r_i$  is a uniform random number between  $-1$  and  $1$ , and  $\alpha$  changes from  $10^{-5}$  to  $10^{-1}$ . Similar conditions are set to the dual variable  $z$ .

Table 6.3 shows that the warm start is clearly effective in solving these problems. Therefore, we can expect that the present algorithm can be used efficiently in solving parametric programming problems. To confirm this possibility, we solve a sequence

TABLE 6.2  
*CUTEr* results for various values of  $\rho$ .

Problem	EPM(10)			EPM(100)		
	stat	nitr(ext)	res	stat	nitr(ext)	res
AUG2DCQP	EXT	18(17)	4.0e-09	EXT	18(17)	1.6e-09
AUG3DCQP	OPT	14(7)	3.3e-09	OPT	17(5)	2.6e-09
BLOCKQP2	OPT	11(0)	4.6e-09	OPT	16(5)	2.7e-10
BLOWEYC	OPT	18(0)	1.0e-09	OPT	8(0)	9.5e-09
BRIDGEND	OPT	42(14)	7.8e-09	OPT	27(8)	6.2e-09
CLNLBEAM	EXT	36(30)	5.1e-09	EXT	38(25)	1.1e-09
CVXQP1	OPT	16(7)	2.8e-10	OPT	18(5)	7.6e-10
DTOC2	OPT	6(0)	5.8e-09	OPT	6(0)	5.8e-09
HELSEBY	EXT	33(28)	4.7e-09	OPT	47(26)	6.1e-10
HUESTIS	EXT	10(9)	5.2e-09	EXT	12(11)	7.2e-09
HYDROELL	OPT	60(35)	8.2e-10	OPT	56(7)	6.0e-09
JANNSON4	EXT	10(8)	3.4e-9	OPT	11(0)	2.3e-09
NCVXQP1	EXT	49(45)	1.9e-09	OPT	25(14)	8.7e-10
ORTHRDM2	OPT	5(0)	2.1e-12	OPT	5(0)	2.1e-12
ORTHRGDM	OPT	6(0)	5.0e-14	OPT	6(0)	5.0e-14
PRIMAL4	OPT	10(4)	9.4e-09	OPT	13(4)	1.1e-09
READING7	OPT	78(1)	1.8e-09	OPT	71(1)	1.1e-09
SOSQP1	OPT	5(3)	7.5e-10	OPT	5(2)	8.3e-09
SSNLBEAM	EXT	119(118)	4.8e-09	EXT	34(33)	9.9e-10
STCQP1	OPT	13(3)	7.5e-11	OPT	12(1)	7.2e-09
STNQP1	OPT	9(5)	9.6e+10	OPT	9(6)	2.8e-09
YAO	EXT	36(36)	2.3e-09	EXT	45(45)	7.1e-09
ZAMB2	OPT	32(8)	2.2e-09	OPT	56(8)	4.6e-10
TOTAL		626(378)			555(223)	
Problem	EPM(1000)			EPM(10000)		
	stat	nitr(ext)	res	stat	nitr(ext)	res
AUG2DCQP	OPT	19(8)	8.6e-09	OPT	30(12)	1.4e-09
AUG3DCQP	OPT	23(8)	1.4e-09	OPT	25(7)	5.5e-10
BLOCKQP2	OPT	11(0)	6.1e-10	OPT	30(3)	1.0e-09
BLOWEYC	OPT	9(0)	1.0e-09	OPT	10(0)	1.0e-09
BRIDGEND	OPT	64(21)	1.0e-09	OPT	46(13)	8.5e-10
CLNLBEAM	OPT	58(36)	6.1e-09	OPT	165(118)	2.9e-09
CVXQP1	OPT	29(11)	4.8e-10	OPT	20(4)	6.9e-09
DTOC2	OPT	6(0)	5.8e-09	OPT	6(0)	5.8e-09
HELSEBY	OPT	41(16)	1.1e-10	OPT	28(9)	1.1e-09
HUESTIS	OPT	15(14)	1.7e-09	OPT	17(4)	1.4e-09
HYDROELL	OPT	70(1)	1.1e-09	OPT	74(3)	1.3e-09
JANNSON4	OPT	10(0)	4.5e-09	OPT	11(0)	2.3e-09
NCVXQP1	OPT	25(12)	1.1e-10	OPT	28(7)	6.2e-10
ORTHRDM2	OPT	5(0)	2.1e-12	OPT	5(0)	2.1e-12
ORTHRGDM	OPT	6(0)	5.0e-14	OPT	6(0)	5.0e-14
PRIMAL4	OPT	15(1)	6.1e-09	OPT	20(1)	1.0e-09
READING7	OPT	61(0)	9.6e-10	OPT	76(0)	3.3e-10
SOSQP1	OPT	8(2)	4.7e-09	OPT	8(3)	1.7e-09
SSNLBEAM	EXT	23(13)	4.6e-10	OPT	54(31)	6.3e-10
STCQP1	OPT	13(1)	4.6e-09	OPT	14(1)	2.9e-09
STNQP1	OPT	11(1)	3.9e-09	OPT	25(8)	6.5e-10
YAO	EXT	135(135)	2.9e-09	EXT	107(107)	2.8e-09
ZAMB2	OPT	67(12)	5.2e-09	OPT	68(17)	6.3e-09
TOTAL		724(294)			873(348)	

of problems that arise from portfolio optimization by the Markowitz model (see [8]) as an example. In the following model, the variables are  $x$  which denote the weight vector and auxiliary vector  $s$ . The data are composed of the return rate matrix  $R$ , the average return rate vector  $r$ , and the lower bound of the expected return rate  $r_p$ .  $n$  denotes the number of assets, and  $T$  denotes the length of the period.

$$\begin{aligned}
 & \text{mimimize} && s^t s / T, s \in \mathbb{R}^n, \\
 & && e^t x = 1, x \geq 0, x \in \mathbb{R}^n, \\
 & \text{subject to} && Rx = s, R \in \mathbb{R}^{T \times n}, \\
 & && r^t x \geq r_p.
 \end{aligned}$$

TABLE 6.2  
(cont'd.) CUTEr results for various values of  $\rho$ .

Problem	EPM(100000)		
	stat	nitr	res
AUG2DCQP	OPT	33(12)	3.5e-09
AUG3DCQP	OPT	30(6)	2.5e-09
BLOCKQP2	OPT	10(0)	1.1e-09
BLOWEYC	OPT	10(0)	1.6e-09
BRIDGEND	OPT	43(16)	1.1e-09
CLNLBEAM	OPT	56(4)	8.5e-10
CVXQP1	OPT	24(0)	3.1e-10
DTOC2	OPT	6(0)	5.8e-09
HELSEBY	OPT	42(1)	7.0e-10
HUESTIS	OPT	19(3)	1.2e-09
HYDROELL	OPT	101(2)	1.6e-09
JANNSON4	OPT	15(1)	5.8e-10
NCVXQP1	OPT	27(6)	4.3e-10
ORTHRDM2	OPT	5(0)	2.1e-12
ORTHRGDM	OPT	6(0)	5.0e-14
PRIMAL4	OPT	36(1)	1.8e-09
READING7	OPT	100(0)	4.7e-10
SOSQP1	OPT	9(4)	2.0e-10
SSNLBEAM	OPT	171(119)	2.5e-09
STCQP1	OPT	16(1)	9.5e-10
STNQP1	OPT	15(0)	4.7e-10
YAO	OPT	23(22)	4.0e-09
ZAMB2	OPT	70(15)	6.5e-10
TOTAL		867(213)	

TABLE 6.3  
CUTEr results for hot/warm start initial points.

Problem	nitr(cold)	nitr(warm)				
		(10 <sup>-5</sup> )	(10 <sup>-4</sup> )	(10 <sup>-3</sup> )	(10 <sup>-2</sup> )	(10 <sup>-1</sup> )
AUG2DCQP	30	2	3	3	4	5
AUG3DCQP	25	2	2	3	3	6
BLOCKQP2	30	2	2	4	5	8
BLOWEYC	10	3	3	2	3	6
BRIDGEND	46	3	4	5	5	15
CLNLBEAM	165	3	4	63	107	128
CVXQP1	20	3	3	4	4	10
DTOC2	6	1	1	1	3	4
HELSEBY	28	2	3	5	6	18
HUESTIS	17	3	3	5	4	7
HYDROELL	74	2	3	12	15	30
JANNSON4	11	2	2	2	3	4
NCVXQP1	28	3	11	6	15	28
ORTHRDM2	5	1	2	2	2	3
ORTHRGDM	6	1	2	2	3	3
PRIMAL4	20	2	2	3	4	10
READING7	76	2	2	2	3	5
SOSQP1	8	2	2	2	4	6
SSNLBEAM	54	2	5	34	33	80
STCQP1	14	2	2	3	4	8
STNQP1	25	2	2	2	3	11
YAO	107	35	17	41	66	73
ZAMB2	68	5	4	4	19	63

We increment  $r_p$  with equal steps, starting from the smallest value which gives condition  $r^t x \geq r_p$  active and ending at the largest value that gives condition  $r^t x \geq r_p$  feasible. Table 6.4 shows the objective function values (variance) and iteration counts

TABLE 6.4  
*Parametric programming.*

variance	70.77509	70.77509	70.77509	70.77509	70.77509	70.77509	70.77509
nitr	2	2	2	2	2	2	2
variance	70.77509	70.77509	70.86993	71.54036	72.84926	76.29329	83.15537
nitr	2	2	3	3	7	3	8
variance	93.65315	106.88078	123.44398	145.32500	173.76275	209.46575	
nitr	10	4	7	3	7	4	

TABLE 6.5  
*Parametric programming in reverse order.*

variance	173.76275	145.32500	123.44398	106.88078	93.65315	83.15537	
nitr	9	11	11	9	7	10	
variance	76.29329	72.84926	71.54036	70.86993	70.77509	70.77509	70.77509
nitr	7	6	9	5	5	2	2
variance	70.77509	70.77509	70.77509	70.77509	70.77509	70.77509	70.77509
nitr	2	2	2	2	2	2	2

for warm start optimizations. The first iteration with a cold start which is not listed here gives 70.77509 by 20 iterations.

Next Table 6.5 shows the results done in reverse order, i.e., the parameter  $r_p$  is decreased from the largest value above with equal steps. In this case, the number of active constraints is decreased as the parameter  $r_p$  is decreased. The first iteration with a cold start gives the objective function value 209.46575 by 26 iterations.

From these experiments, we see that the present algorithm can be effectively used as an algorithm for the parametric programming problems as well as for general nonlinear programming.

**Acknowledgment.** The authors would like to thank the associate editor and referees for their valuable comments which improved our paper greatly.

#### REFERENCES

- [1] R. H. BYRD, M. B. HRIBAR, AND J. NOCEDAL, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.
- [2] L. CHEN AND D. GOLDFARB, *Interior point  $\ell_2$ -penalty methods for nonlinear programming with strong global convergence properties*, Math. Program., 108 (2006), pp. 1–36.
- [3] A. FISCHER, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.
- [4] R. FLETCHER, *Practical Methods of Optimization*, Wiley, London, 1987.
- [5] J. GONDZIO, *Warm start of the primal-dual method applied in the cutting-plane scheme*, Math. Program., 83 (1998), pp. 125–143.
- [6] N. I. M. GOULD, D. ORBAN, AND PH. L. TOINT, *An interior-point  $\ell_1$ -penalty method for nonlinear optimization*, Technical report RAL-TR-2003-022, Rutherford Appleton Laboratory, Oxfordshire, UK, 2003.
- [7] N. I. M. GOULD, D. ORBAN, AND PH. L. TOINT, *CUTEr (and SifDec), a Constrained and Unconstrained Testing Environment, Revisited*, Technical report, The Science and Technology Facilities Council, Swindon, UK, 2001; also available online from <http://www.hsl.rl.ac.uk/cuter-www/Doc/cuter.pdf>.
- [8] H. KONNO AND K. SUZUKI, *A fast algorithm for solving large scale mean-variance models by compact factorization of covariance matrices*, J. Oper. Res. S Japan, 35 (1992), pp. 93–104.
- [9] A. WÄCHTER AND L. T. BIEGLER, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57.
- [10] E. A. YILDIRIM AND S. J. WRIGHT, *Warm-start strategies in interior-point methods for linear programming*, SIAM J. Optim., 12 (2002), pp. 782–810.
- [11] H. YAMASHITA AND H. YABE, *Superlinear and quadratic convergence of some primal-dual inte-*

- rior point methods for constrained optimization*, Math. Program., 75 (1996), pp. 377–397.
- [12] H. YAMASHITA, *A globally convergent primal-dual interior point method for constrained optimization*, Optim. Methods Softw., 10 (1998), pp. 443–469.
- [13] H. YAMASHITA AND H. YABE, *An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization*, SIAM J. Optim., 14 (2003), pp. 479–499.
- [14] H. YAMASHITA, H. YABE, AND T. TANABE, *A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization*, Math. Program., 102 (2005), pp. 111–151.
- [15] H. YAMASHITA AND H. YABE, *Quadratic convergence of a primal-dual interior point methods for degenerate nonlinear optimization problems*, Comput. Optim. Appl., 31 (2005), pp. 123–143.