A Primal-dual Exterior Point Method for Nonlinear Optimization

Hiroshi Yamashita* and Takahito Tanabe†

Abstract

In this paper, a primal dual method for general possible nonconvex nonlinear optimization problems is considered. The method is an exterior point type method which means that it permits primal variables violate inequality constraints during the iterations. The method is based on exact penalty type transformation of inequality constraints, and use smooth approximation of the problem to form primal-dual iteration based on Newton method as in usual primal-dual interior point method. The global convergence and local superlinear/quadratic convergence of the proposed methods are proved. For global convergence, methods using line search and trust region are proposed. The method is tested with CUTE problems, and is shown to have similar efficiency to the primal-dual interior point method proposed by Yamashita, Yabe and Tanabe. It is also shown that the method can enjoy warm start conditions easily unlike interior point methods.

1 Introduction

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

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

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

Let the Lagrangian function of the above problem be defined by

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

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 optimality of problem (1) are given by

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

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

and

$$x \geq 0, \qquad z \geq 0, \tag{4}$$

where

$$\begin{aligned}
\nabla_x L(w) &= \nabla f(x) - A(x)^t y - z, \\
A(x) &= \begin{pmatrix} \nabla g_1(x)^t \\ \vdots \\ \nabla g_m(x)^t \end{pmatrix}, \\
X &= \operatorname{diag}(x_1, \cdots, x_n), \quad Z = \operatorname{diag}(z_1, \cdots, z_n), \\
e &= (1, \cdots, 1)^t \in \mathbb{R}^n.
\end{aligned}$$

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

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

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

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

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

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

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. Iterates are kept in the interior region that satisfies $x > 0$ and $z > 0$ by definition.

Recent researches on interior point methods for nonlinear optimization problems ([5], [6], [8], [9], [1]) show good theoretical properties and practical performance for wide range of problems. One possible drawback of the method is that the iterates should be kept interior - the very basic nature of the algorithm. If the feasible region is "narrow", the iterates that starts from a point far from the solution may take many iterations to arrive at a region near the solution. If an iterate happens to be near the boundary of the feasible region which is not close to the 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 μ is small. Also it is known that the warm start condition is not easy to utilize in the interior point method framework despite several past researches on this topic ([4], [3]). Therefore it is of interest to consider an algorithm that does not need interior point requirement, and is able to utilize warm start conditions. In this paper we consider a primal-dual iteration that can lie outside the primal interior region. And we will show that the method is of similar numerical performance for various test problems, and it can in fact utilize the warm start condition, and it can be effective in parametric programming usage.

To this end we firstly define the following problem:

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

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). In this form the non-negativity restriction on the variable x in (1) is eliminated. Thus we consider solving problem (6) in the primal-dual space hereafter.

The necessary conditions for optimality of this problem are (see 14.2 of Fletcher [2])

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

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

$$\begin{array}{ll} 0 \le z_i \le \rho, & x_i = 0, \\ z_i = 0, & x_i > 0, \\ z_i = \rho, & x_i < 0, \end{array}$$

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

$$|x_i| \, z_i - \rho \, |x_i|_- = 0, \; 0 \le z_i \le \rho, \; i = 1, \cdots, n. \tag{8}$$

Therefore conditions (7) can be written as

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

and

$$0 \le z \le \rho \tag{10}$$

where

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

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:

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

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

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

$$h''(a,\mu) = \frac{\mu^2}{2(a^2+\mu^2)^{3/2}}, \tag{12}$$

and note that

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

for $\mu > 0$. By using the function $h(a,\mu)$, problem (6) is approximated by the following problem:

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

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

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

By introducing the auxiliary variable z by

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

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, in view of (8) and (11), we modify the third equation to

$$\sqrt{x_i^2+\mu^2} \cdot z_i - \rho h(x_i,\mu) = 0, \quad i = 1,...,n. \tag{13}$$

We note that (13) can be viewed as a smooth approximation of (8). Thus we express the KKT conditions as

$$r(w,\mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ U(x,\mu)z - \rho H(x,\mu)e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{14}$$

where

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

An algorithm of this paper approximately solves the sequence of conditions (14) with a decreasing sequence of the parameter μ that tends to 0, and thus obtains a solution to the KKT conditions. For definiteness, we describe a prototype of such algorithm as follows.

Algorithm EP

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

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

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

$$\|r(w_{k+1},\mu_k)\| \le M_c \mu_k, \tag{15}$$

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

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 (15) for each k, and μ_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

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

we have

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

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

We note that the parameter sequence $\{\mu_k\}$ in Algorithm EP need not be determined beforehand. The value of each μ_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 in this paper. Let $\Delta w = (\Delta x, \Delta y, \Delta z)^t$ be defined by a solution of

$$J(w,\mu)\Delta w = -r(w,\mu), \tag{16}$$

where

$$J(w) = \begin{pmatrix} G & -A(x)^t & -I \\ A(x) & 0 & 0 \\ V(w,\mu) & 0 & U(x,\mu) \end{pmatrix}, \tag{17}$$

$$\begin{aligned} v(w_i,\mu) &= z_i u'(x_i,\mu) - \rho h'(x_i,\mu) \\ &= \frac{x_i(z_i - \rho/2)}{\sqrt{x_i^2+\mu^2}} + \frac{\rho}{2}, \quad i = 1,...,n, \\ V(w,\mu) &= \operatorname{diag}\{v(w_1,\mu),...,v(w_n,\mu)\}, \end{aligned} \tag{18}$$

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 (16), or by solving

$$(G + U(x,\mu)^{-1}V(w,\mu))\Delta x - A^t\Delta y = -\nabla f(x) + A^t y - \rho H'(x,\mu)e, \tag{19}$$

$$A\Delta x = -g(x), \tag{20}$$

and

$$\Delta z = -z - \rho H'(x,\mu)e - U(x,\mu)^{-1}V(w,\mu)\Delta x. \tag{21}$$

Following lemma gives a basic property of the iteration vector $\Delta w = (\Delta x, \Delta y, \Delta z)^t$, and is apparent from (19) to (21).

Lemma 1 *Suppose that Δw satisfies (16) at a point w.*
(i) If $\Delta w = 0$, then the point w is a KKT point that satisfies (14).
(ii) If $\Delta x = 0$, then the point $(x, y + \Delta y, z + \Delta z)$ is a KKT point that satisfies (14). □

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

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 (18), 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 assure global convergence of the proposed algorithms that use the Newton iteration, we need a merit function. To this purpose the following form of the penalty function:

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

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 ρ, ρ' and μ 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.,

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

where

$$Q = \nabla^2 f(x) - \sum_{i=1}^{m} y_i \nabla^2 g_i(x) + \rho H''(x, \mu),$$
$$H''(x,\mu) = \mathrm{diag}\left\{ h''(x_1, \mu), ..., h''(x_n, \mu) \right\}.$$

We also need the difference of these quantity 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.

Lemma 3 *Suppose that Δw satisfies (16) at a point w.*
Then there holds

$$\Delta F_l(x, \Delta x) \leq -\Delta x^t (G + U(x,\mu)^{-1} V(w,\mu)) \Delta x - (\rho' - \|y + \Delta y\|_\infty) \sum_{i=1}^{m} |g_i(x)|. \tag{23}$$

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

Proof. From (19) and (20) 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 (23).

A proof of the second statement is easy because two terms in (23) are nonpositive by the assumption. □

Let $F'(x;s)$ be the 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 α.
(ii) There holds the relation

$$F(x) + F'(x;s) \leq F_l(x;s). \tag{24}$$

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

$$F(x+s) \leq F(x) + F'(x+\theta s, s). \tag{25}$$

Proof. See [6]. □

3 Line search algorithm

In this section, we describe an algorithm that uses line searches, and prove its global convergence. The algorithm is similar to the interior point method proposed by Yamashita [6]. The basic iteration of the line search algorithm may be described as

$$w_{k+1} = w_k + \Lambda_k \Delta w_k, \tag{26}$$

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.

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 method where the primal variable x should always satisfy positivity condition, there is no such restriction here. Therefore Armijo's step size rule is the same as in the unconstrained optimization. A 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

$$F(x_k + \beta^{l_k} \Delta x_k) - F(x_k) \leq \varepsilon_0 \beta^{l_k} \Delta F_l(x_k), \tag{27}$$

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

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

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

$$\alpha_{yk} = 1 \quad \text{or} \quad \alpha_{zk}. \tag{28}$$

The global convergence property given below holds for both choices.

The following algorithm describes the iteration for fixed $\mu > 0$, $\rho > 0$ and $\rho' > 0$. We note that this algorithm corresponds to Step 2 of Algorithm EP in Section 1.

Algorithm LS

Step 0. (Initialize) Let $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n_\rho$, where $\mathbb{R}^n_\rho = \{ z \in \mathbb{R}^n \,|0 \le z_i \le \rho, \ i = 1, ..., n \}$, 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 $\|r(w_k, \mu)\| \le \varepsilon'$, then stop.

Step 2. (Compute direction) Calculate the direction Δw_k by (16).

Step 3. (Stepsize) Find the smallest nonnegative integer l_k that satisfies

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

Calculate

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

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

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

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

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 $g_i, i = 1, ..., m$, 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 $\{ x \in \mathbb{R}^n \,|F(x) \le F(x_0) \}$, 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.

(5) The penalty parameter ρ' satisfies $\rho' \geq \|y_k + \Delta y_k\|_\infty$ 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 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 Lemma 1 and Lemma 3. Therefore in the following, we may assume $\Delta F_l(x_k, \Delta x_k) < 0$ if an infinite sequence is generated by Algorithm LS. The following theorem gives a convergence of an infinite sequence generated by Algorithm LS.

Theorem 1 *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 $\Delta F_l(x_k, \Delta x_k) < 0$, 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

$$\frac{\|p\|^2}{M} \leq p^t(G_k + U(x_k, \mu)^{-1}V(w_k, \mu))p \leq M\|p\|^2, \quad \forall p \in \mathbf{R}^n, \tag{29}$$

by the assumption. From (27), (23) and (29), we have

$$F(x_{k+1}) - F(x_k) \leq \varepsilon_0 \beta^{l_k} \Delta F_l(x_k, \Delta x_k) \leq -\varepsilon_0 \beta^{l_k} \frac{\|\Delta x_k\|^2}{M} < 0. \tag{30}$$

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, then $\Delta x_k \to 0$ from (30). Now suppose that there exists a subsequence $K \subset \{0, 1, \cdots\}$ such that $l_k \to \infty, k \in K$. 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 (27), and we have

$$F(x_k + \alpha_{xk}\Delta x_k/\beta) - F(x_k) > \varepsilon_0 \alpha_{xk} \Delta F_l(x_k, \Delta x_k)/\beta. \tag{31}$$

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

$$\begin{aligned} F(x_k + \alpha_{xk}\Delta x_k/\beta) - F(x_k) &\leq \alpha_{xk} F'(x_k + \theta_k \alpha_{xk}\Delta x_k/\beta, \Delta x_k)/\beta \\ &\leq \alpha_{xk} \Delta F_l(x_k + \theta_k \alpha_{xk}\Delta x_k/\beta, \Delta x_k)/\beta, k \in K. \end{aligned} \tag{32}$$

Now from (31) and (32), 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

$$\begin{aligned} & \Delta F_l(x_k + \theta_k \alpha_{xk} \Delta x_k/\beta, \Delta x_k) - \Delta F_l(x_k, \Delta x_k) \\ > \; & (\varepsilon_0 - 1)\Delta F_l(x_k, \Delta x_k) > 0. \end{aligned} \tag{33}$$

Because Δx_k satisfies (19) and (20) and there holds (29), by Assumption GLS(3), $\|\Delta x_k\|$ is uniformly bounded above. Then by the assumption $l_k \to \infty, k \in K$, we have $\|\theta_k \alpha_{xk} \Delta x_k/\beta\| \to 0, k \in K$. Thus the left hand side of (33) and therefore $\Delta F_l(x_k, \Delta x_k)$ converges to zero when $k \to \infty, k \in K$. This yields $\Delta x_k \to 0, k \in K$ because we have

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

also from (23) and (29).

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

$$x_k \to \hat{x}, \quad \Delta x_k \to 0, \quad x_{k+1} \to \hat{x}, \quad k \in K. \tag{34}$$

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

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

from (21). If we define $\hat{z} = -\rho H'(\hat{x}, \mu)e$, then $0 < \hat{z} < \rho$, and

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

This shows that the point $z_k + \Delta z_k$ is always accepted as z_{k+1} (i.e., $\alpha_{zk} = 1$) 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

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

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

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

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. □

4 Trust region algorithm

In this section, we describe an algorithm that uses trust region type iterations. Basic algorithm is same as the primal-dual interior point trust region method proposed by Yamashita, Yabe and Tanabe [8].

As in [8], 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

$$\begin{pmatrix} D & -A(x)^t & -I \\ A(x) & 0 & 0 \\ V(w,\mu) & 0 & U(x,\mu) \end{pmatrix} \begin{pmatrix} \Delta x_{SD} \\ \Delta y_{SD} \\ \Delta z_{SD} \end{pmatrix} = -r(w,\mu), \tag{35}$$

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.

Replacing G by D in Lemma 3, we have

$$\Delta F_l(x;\Delta x_{SD}) \leq -\Delta x_{SD}^t (D + U(x,\mu)^{-1}V(w,\mu))\Delta x_{SD} - (\rho' - \|y + \Delta y_{SD}\|_\infty) \sum_{i=1}^m |g_i(x)| \,. \tag{36}$$

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 Δx_{SD} is a descent direction of the merit function $F(x)$.

A trust region algorithm that finds a KKT point for a fixed μ may proceeds as follows. At iteration k, let us assume that the trust region radius $\delta_k > 0$ and the vectors Δw_k and Δw_{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

$$\Delta F_q(x_k; s_k) \leq \frac{1}{2} \Delta F_q(x_k; \alpha^*(x_k, \Delta x_{SDk})\Delta x_{SDk}), \tag{37}$$

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

$$\alpha^*(x,d) = \arg\min \left\{ F_q(x;\alpha d) \ \mid \|\alpha d\| \leq \delta \right\}. \tag{38}$$

for $x \in \mathbf{R}^n_+$, $d \in \mathbf{R}^n$. 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 δ. Therefore condition (37) is a sufficient decrease condition based on the steepest descent step.

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

Algorithm TR

Step 0. An initial point $w_0 \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n_\rho$ and positive parameters μ, ρ and ρ' 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 Δw_k and Δw_{SDk} that satisfy (16) and (35) respectively. If $G_k = \nabla_x^2 L(w_k)$ gives a too large vector that does not satisfy the first inequality of (40) given below, G_k is modified to satisfy (40).

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

$$\|s_k\| \le \delta_k, \tag{39}$$

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

Step 4. Update the trust region radius δ_{k+1} by

$$\begin{aligned} \text{If } \Delta F(x_k, s_k) &> \frac{1}{4}\Delta F_q(x_k, s_k), \text{ then } \delta_{k+1} = \frac{1}{2}\delta_k; \\ \text{If } \Delta F(x_k, s_k) &\le \frac{3}{4}\Delta F_q(x_k, s_k), \text{ then } \delta_{k+1} = 2\delta_k; \\ \text{Otherwise } \delta_{k+1} &= \delta_k. \end{aligned}$$

Step 5. If $\Delta F(x_k, s_k) \le 0$, then set $x_{k+1} = x_k + s_k$, compute α_{yk} and α_{zk}, 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 the global convergence of Algorithm TR, we list the necessary assumptions.

Assumption GTR

(1) The functions f and $g_i, i = 1, ..., m$, 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

$$\|\Delta x_k\| \le M \|\Delta x_{SDk}\|, \|s_k\| \le M \|\Delta x_{SDk}\|, \tag{40}$$

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

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

It follows from Assumption GTR that the linear system of equations (35) has a unique solution and that the direction Δx_{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 (16) and (35).
(2) Since $(0, \Delta y_k, \Delta z_k)^t$ satisfies (35) and the coefficient matrix of (35) is nonsingular, the uniqueness of the solution to (35) implies $\Delta x_{SDk} = 0$.
(3) This follows from GTR (5).
(4) If $\Delta x_k = 0$, then by (21) 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 (19)-(21) 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. □

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 $k = 0, 1, \cdots$ in the following.

We state the following simple lemma first.

Lemma 6 *If a vector* $d \in \mathbf{R}^n$ *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

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

Thus the proof is complete. □

Lemma 7 *Let* $x \in \mathbf{R}^n, 0 \neq d \in \mathbf{R}^n$ *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 (38) can be expressed as

$$\alpha^*(x,d) = \min\left\{ 1,\ \frac{\delta}{\|d\|}, -\frac{\Delta F_l(x;d)}{\max\{d^tGd, 0\}} \right\}, \tag{41}$$

where the last term in the braces in the right hand side is assumed to give the value ∞ *if the value of the denominator is* 0*. Further we have*

$$\Delta F_q(x;\alpha^*(x,d)d) \leq \frac{1}{2}\alpha^*(x,d)\Delta F_l(x;d). \tag{42}$$

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

$$F_q(x,\alpha d) = F(x,\mu) + \alpha\Delta F_l(x;d) + \frac{1}{2}\alpha^2 d^tQd, \quad \alpha \in [0,1]. \tag{43}$$

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

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

Therefore we obtain

$$\alpha^*(x,d) = \min\left\{ \frac{\delta}{\|d\|}, -\frac{\Delta F_l(x,d)}{d^tQd} \right\}, \tag{44}$$

in this case. From this relation we have

$$d^tQd \leq -\frac{\Delta F_l(x;d)}{\alpha^*(x,d)}. \tag{45}$$

From (43) and (45) we deduce

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

If $d^tQd \leq 0$, we have

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

and

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

Therefore we proved (41) and (42). □

Theorem 2 *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 (14).*

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

$$\Delta F_q(x_k, s_k) \leq \frac{1}{4}\Delta F_l(x_k, \Delta x_{SDk}) \min\left\{ \frac{\delta_k}{\|\Delta x_{SDk}\|}, -\frac{\Delta F_l(x_k, \Delta x_{SDk})}{\max\{\Delta x_{SDk}^t Q_k \Delta x_{SDk}, 0\}}\right\}. \tag{46}$$

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

$$\Delta F(x_k, s_k) > \frac{1}{4}\Delta F_q(x_k, s_k),\ k \in K_1, \tag{47}$$

$$\Delta F(x_k, s_k) \leq \frac{1}{4}\Delta F_q(x_k, s_k),\ k \in K_2. \tag{48}$$

(i) Suppose that K_1 is an infinite sequence.
(i-a) If $\liminf_{k\to\infty, k\in K_1} \delta_k = 0$, then there exists an infinite set $K_1' \subset K_1$ such that $\delta_k \to 0, k \in K_1'$. Then because $\|s_k\| \leq \delta_k$, we have $\|s_k\| \to 0,\ k \in K_1'$. Suppose $\liminf_{k\to\infty} \|\Delta x_{SDk}\| > 0$. Then Assumption (G6) and (36) yield

$$\liminf_{k\to\infty, k\in K_1'} |\Delta F_l(x_k, \Delta x_{SDk})| > 0.$$

On the other hand, we have

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

From (47) and the above relation, we have

$$-\Delta F_q(x_k, s_k) < \mathrm{O}\left(\|s_k\|^2\right).$$

However this contradicts (46), because it gives the relation

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

for sufficiently large $k \in K_1'$. Thus we obtain $\liminf_{k\to\infty} \|\Delta x_{SDk}\| = 0$ in this case.
(i-b) If $\liminf_{k\to\infty, k\in K_1} \delta_k > 0$, 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 $\liminf_{k\to\infty, k\in K_2'} \delta_k > 0$. 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

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

and thus $\Delta F_q(x_k, s_k) \to 0,\ k \in K_2$, from (48). Therefore we have $\Delta F_l(x_k, \Delta x_{SDk}) \to 0,\ k \in K_2'$, from (46). Then, by (36) we obtain $\Delta x_{SDk} \to 0,\ k \in K_2'$, and thus $\liminf_{k\to\infty} \|\Delta x_{SDk}\| = 0$ in this case.

(ii-b) Suppose $\lim_{k\to\infty, k\in K_2} \delta_k = 0$. 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 (49) 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 $\liminf_{k\to\infty, k\in K_1} \delta_k > 0$ and $\lim_{k\to\infty, k\in K_2} \delta_k = 0$. 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

$$\liminf_{k\to\infty} \|\Delta x_{SDk}\| = 0. \tag{49}$$

By the requirement (40), this means that we have

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

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

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

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

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

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

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

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

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

We note that actual trust region step calculation is similar to the one described in [8], and is not described here.

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 Trial Newton step (see below) that checks if the Newton step gives a point w_{k+1} that satisfies the condition $\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k^\eta, \eta \in (0,1]$ for a given μ_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 $\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k^\eta, \eta \in (0,1]$. We note that the condition for approximate KKT point here is looser than the condition in Algorithm EP for $\mu_k < 1$ and $\eta < 1$. The procedure is described as Step 2 and 3 of the algorithm below.

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 $\|r_0(w_k)\| \le \varepsilon$, 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 Δw_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 $0 \le z_k + \alpha_k \Delta z_k \le \rho$: Firstly calculate the maximum step $\bar{\alpha}_k$ to the constraints $0 \le z_k + \alpha_k \Delta z_k \le \rho$ by

$$\bar{\alpha}_k = \min\left\{ \min_i \left\{ \frac{\rho - (z_k)_i}{(\Delta z_k)_i} \,|(\Delta z_k)_i > 0 \right\}, \min_i \left\{ -\frac{(z_k)_i}{(\Delta z_k)_i} \,|(\Delta z_k)_i < 0 \right\} \right\}. \tag{50}$$

Then determine the step α_k by

$$\alpha_k = \min\{1, \bar{\alpha}_k\}. \tag{51}$$

Step 2.3 If $\|r(w_k + \alpha_k \Delta w_k, \mu_k)\| \le M_c \mu_k^\eta$, 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

$$\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k^\eta. \tag{52}$$

Step 4. Set $k := k + 1$ and go to Step 1. □

The global convergence of Algorithm superlinearEP is apparent from the previous sections. So in the following we confine the discussion to the local convergence property. Therefore the initial point w_0 in particular is assumed to be close to a KKT point w^*. And we will prove that if μ_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) Active constraint set at x is defined by a set composed of all equality constraints and 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) Strict complementarity condition of the solution w^* is that $z_i^* \in (0, \rho)$ if $x_i^* = 0$.□

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^*. □

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 following estimates are valid:*

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

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

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

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

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

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

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

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

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

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

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

(iii) Proof is similar to (i).
(iv) From the definition of $r(w,\mu)$ and $r_0(w)$, we have

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

from the above. □

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

$$\begin{aligned} \hat{v}_i = 0, \quad x_i \ne 0, \\ \hat{v}_i > 0, \quad x_i = 0, \end{aligned}$$

for $i=1,...,n$. Then by the Assumption (L3), it can be shown that the matrix $\hat{J}(w^*,\hat{v})$ is nonsingular as in 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 ξ such that*

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

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

Lemma 10 *There exists a constant $\delta>0$ such that, if $\|w-w^*\|\le\delta$, then for sufficiently small $\mu>0$ there exists a positive constant ξ' such that*

$$\left\|J(w,\mu)^{-1}\right\|\le\xi'. \tag{53}$$

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

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

for sufficiently small δ. Thus by Banach perturbation lemma and Lemma 9, we proved the lemma. □

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

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

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

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

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

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

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

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

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

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

Thus the lemma is proved. □

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

$$0 > \Delta z_i = \mathrm{O}(\mu^2), \tag{54}$$

when $z_i = \rho$, and

$$\frac{\Delta z_i}{\rho - z_i} = 1 + \mathrm{O}(\mu^2), \frac{\Delta z_i}{z_i} = \mathrm{O}(\|r(w,\mu)\|) + \mathrm{O}(\|w - w^*\|) + \mathrm{O}(\mu^2) \tag{55}$$

when $z_i < \rho$;

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

$$\frac{\Delta z_i}{\rho - z_i} = \mathrm{O}(\|r(w,\mu)\|), \frac{\Delta z_i}{z_i} = \mathrm{O}(\|r(w,\mu)\|); \tag{56}$$

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

$$0 < \Delta z_i = \mathrm{O}(\mu^2), \tag{57}$$

when $z_i = 0$, and

$$\frac{\Delta z_i}{\rho - z_i} = \mathrm{O}(\|r(w,\mu)\|) + \mathrm{O}(\|w - w^*\|) + \mathrm{O}(\mu^2), -\frac{\Delta z_i}{z_i} = 1 + \mathrm{O}(\mu^2), \tag{58}$$

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 $\|\Delta w\| = \mathrm{O}(\|r(w,\mu)\|)$ by Lemma 10.

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

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

and (54) follows.

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

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

and (55) follows.

(ii) If $x_i^* = 0$, we have (56) 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 (57) follows.

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

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

and (58) follows. □

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

$$1 - \alpha = \mathrm{O}(\mu^2). \tag{59}$$

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

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

from Lemma 12 and the definition of $\bar{\alpha}$ (50). If not $\bar{\alpha} > 1$. Thus we have (59) from the definition of α (51). □

Now we prove the superlinear convergence of Algorithm EPlocal.

Theorem 2 *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 $\|r(w_{k+1},\mu_k)\| \le M_c\mu_k^\eta$, and converges to w^* superlinearly. If $\tau \in [1, 2/\eta - 1)$, then the convergence rate is quadratic.*

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

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

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

Next we have

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

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

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

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

Therefore the theorem is proved. □

6 Numerical Experiment

6.1 CUTE problems

The proposed method is programmed and tested. We report the results of numerical experiment for CUTE problems with the trust region algorithm here. In this experiment, we set $M_c = 1.1, \rho = 10^4$ and $\mu_0 = x_0^t z_0 / n\rho$. From CUTE problem (the version around 1997), we select 42 problems with $n + m > 2000$, where n is the number of variables and m is the number of constraints. The number m does not count bound constraints. In the following table, `IPM` means the result of the interior point method in [8], and `EPM` means the present method. `OPT` in stat column denotes the optimal solution, and `EXT` denotes optimal but exterior solution. `F` means a failure of the methods due to iteration count over. Numbers in `nitr` column denote total trust region iteration counts executed. The numbers in `res` column denote the final KKT condition residuals obtained.

Problem	n	m	IPM			EPM		
			stat	nitr	res	stat	nitr	res
AUG2D	3280	1600	OPT	1	6.55E-16	OPT	1	6.55E-16
AUG2DC	3280	1600	OPT	1	2.03E-14	OPT	1	2.03E-14
AUG2DCQP	3280	1600	OPT	20	5.46E-08	EXT	27	1.66E-05
AUG2DQP	3280	1600	OPT	20	1.31E-06	EXT	34	2.26E-06
AUG3D	3873	1000	OPT	1	1.93E-16	OPT	1	1.93E-16
AUG3DC	3873	1000	OPT	1	1.70E-15	OPT	1	1.70E-15
AUG3DCQP	3873	1000	OPT	14	5.35E-07	OPT	29	9.44E-08
AUG3DQP	3873	1000	OPT	15	4.50E-07	OPT	28	1.22E-06
BLOCKQP1	2005	1001	OPT	9	1.69E-07	OPT	11	8.56E-11
BLOCKQP2	2005	1001	OPT	9	1.09E-06	OPT	23	4.28E-12
BLOCKQP3	2005	1001	OPT	9	1.42E-07	OPT	13	5.11E-12
BLOCKQP4	2005	1001	OPT	21	7.43E-07	OPT	21	8.41E-07
BLOCKQP5	2005	1001	OPT	9	1.42E-07	OPT	8	4.83E-07
BLOWEYA	2002	1002	OPT	7	4.16E-08	OPT	9	7.43E-07
BLOWEYB	2002	1002	OPT	5	2.27E-07	OPT	10	1.18E-07
BLOWEYC	2002	1002	OPT	8	6.19E-08	OPT	9	2.12E-12
BRIDGEND	2734	2727	OPT	41	2.09E-07	F	201	3.49E+03
CLNLBEAM	3003	2000	OPT	5	1.10E-06	OPT	174	2.71E-08
DTOC2	5998	3996	OPT	5	8.83E-08	OPT	5	4.38E-07
DTOC5	9999	4999	OPT	3	5.36E-09	OPT	3	5.36E-09
DTOC6	10001	5000	OPT	11	3.99E-10	OPT	11	3.99E-10
HELSBY	1408	1399	OPT	77	9.63E-10	F	201	2.33E+03
HYDROELL	1009	1008	OPT	37	9.23E-10	OPT	95	2.52E-07
JANNSON3	20000	3	OPT	8	1.22E-06	OPT	36	1.05E-06
JANNSON4	10000	2	OPT	11	9.48E-07	OPT	13	1.08E-06
MINC44	1113	1032	OPT	122	3.47E-08	F	201	1.16E+03
ORTHRDM2	4003	2000	OPT	4	8.85E-12	OPT	5	1.94E-11
ORTHRGDM	4003	2000	OPT	5	2.05E-12	OPT	6	1.17E-10
SARO	4754	4015	F	201	9.31E-05	F	201	6.18E+02
SAROMM	5120	5110	F	201	8.56E+00	F	201	6.18E+03
SIPOW1	2	2000	F	201	2.23E-02	OPT	16	1.02E-08
SIPOW1M	2	2000	F	201	1.65E-02	EXT	28	5.06E-05
SIPOW2	2	2000	F	201	9.26E-03	OPT	16	3.71E-08
SIPOW2M	2	2000	F	201	3.12E-03	OPT	25	1.73E-08
SOSQP1	2000	1001	OPT	27	9.36E-08	OPT	3	2.76E-11
SOSQP2	2000	1001	OPT	13	2.89E-07	OPT	16	7.95E-07
SSNLBEAM	3003	2000	OPT	5	5.22E-09	F	201	2.18E+00
STCQP1	4097	2052	OPT	9	3.93E-07	OPT	13	5.68E-07
STCQP2	4097	2052	OPT	9	4.28E-07	OPT	13	7.47E-07
STNQP1	4097	2052	OPT	11	1.47E-09	OPT	15	1.17E-06
STNQP2	4097	2052	OPT	10	4.56E-11	OPT	52	5.77E-08
YAO	2002	2000	OPT	58	8.60E-08	EXT	11	6.68E-13

From the above table, we see that IPM and EPM solves problems with the similar failure rates. EPM needs more iterations for convergence. The following table shows the summary of iteration counts.

	IPM	EPM
OPT/EXT count	36	36
average of required iterations (OPT/EXT)	17.3	21.7

6.2 Warm start condition and parametric programming

As noted in Introduction, it is not easy to utilize warm/hot start condition of a given initial point with the interior point methods. However, the algorithm of this paper can enjoy this condition easily. In the following experiment, we perturb all the primal and dual solutions obtained from the cold start condition (the above experiment) with the maximum relative amount from 10^{-5} to 10^{-2} by uniform random numbers. The initial value of μ is set to 10^{-6} for all cases. This value should be set adaptively for further improvement. `nitr` with `200` means iteration count over. `nitr` with `*` means convergence to an exterior point.

Problem	n	m	nitr(cold)	nitr(warm)			
				(10^{-5})	(10^{-4})	(10^{-3})	(10^{-2})
AUG2DCQP	3280	1600	27*	1	2*	3	5
AUG2DQP	3280	1600	34*	1	2	3	6
AUG3DCQP	3873	1000	29	1	4	6	24
AUG3DQP	3873	1000	28	2	3	5	106
BLOCKQP1	2005	1001	11	11	11	10	14
BLOCKQP2	2005	1001	23	3	3	4	11
BLOCKQP3	2005	1001	13	10	10	9	77
BLOCKQP4	2005	1001	21	3	3	4	11
BLOCKQP5	2005	1001	8	10	10	10	77
BLOWEYA	2002	1002	9	1	2	3	200
BLOWEYB	2002	1002	10	1	1	3	5
BLOWEYC	2002	1002	9	1	2	3	19
CLNLBEAM	3003	2000	174	13	10	25	97
DTOC2	5998	3996	5	1	1	1	2
DTOC5	9999	4999	3	1	1	1	1
DTOC6	10001	5000	11	1	1	2	2
HYDROELL	1009	1008	95	3	3	4	14
JANNSON3	20000	3	36	1	2	2	4
JANNSON4	10000	2	13	1	2	2	4
ORTHRDM2	4003	2000	5	1	1	2	2
ORTHRGDM	4003	2000	6	1	1	2	2
SIPOW1	2	2000	16	3	11	20	22
SIPOW1M	2	2000	28*	6*	34*	20*	25*
SIPOW2	2	2000	16	3	4	13	51
SIPOW2M	2	2000	25	2	12	16	39
SOSQP1	2000	1001	3	3	3	9	6
SOSQP2	2000	1001	16	3	8	15	200
STCQP1	4097	2052	13	2	2	2	3
STCQP2	4097	2052	13	1	2	3	27
STNQP1	4097	2052	15	2	2	3	5
STNQP2	4097	2052	52	2	3	4	4
YAO	2002	2000	11*	200	115*	200	50*

The above table shows that warm start condition is clearly effective in solving the problems except for a few problems. The reason of the failures for these is not unknown for now.

From the above experiment, we can expect that the present algorithm can be used effectively in parametric programming problems. To show this possibility, we calculate a sequence of problems that arise from portfolio optimization by Markowitz model with the interior point method and the exterior point method for comparison. 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 expected return rate r_p. n denotes the number of assets and T denotes the length of the period.

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

We increment r_p from 2.0% to 4.0% by 0.2% step for efficient frontier calculation. The following table shows the iteration counts of `IPM` and `EPM`.

r_p	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
IPM	28	22	22	22	21	23	21	23	24	24	25
EPM	24	3	3	3	3	8	4	8	9	8	4

From this experiment, we see that the present algorithm can be effectively used as an algorithm for the parametric programming problems.

References

[1] R. Byrd, M. B. Hribar and J. Nocedal, An interior point method for large scale nonlinear programming, *SIAM Journal on Optimization*, 9 (1999), pp. 877-900.

[2] R. Fletcher, *Practical methods of optimization*, Wiley, 1987.

[3] J. Gondzio, Warm start of the primal-dual method applied in the cutting-plane scheme, *Mathematical Programming*, 83 (1998), pp. 125-143.

[4] E. A. Yildirim and S. J. Wright, Warm start strategies in interior-point methods for linear programming, *SIAM Journal on Optimization,* 12 (2002), pp. 782–810.

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

[6] H. Yamashita, A globally convergent primal-dual interior point method for constrained optimization, *Optimization Methods and Software*, 10 (1998), pp.443-469.

[7] H.Yamashita and Y.Yabe, An iterior point method wih a primal-dual quadratic barrier penalty function for nonlinear optimization, *SIAM J. Optimization*, 14 (2003), pp.479-499.

[8] H.Yamashita, H.Yabe and T.Tanabe, A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization, *Mathematical Programming*, 102 (2005), pp.111-151.

[9] H.Yamashita and H.Yabe, Quadratic convergence of a primal-dual interior point methods for degenerate nonlinear optimization problems, *Computational Optimization and Applications*, 31 (2005), pp.123-143.