An interior point method with a primal-dual l_2 barrier penalty function for nonlinear optimization

Hiroshi Yamashita* and Hiroshi Yabe†

March, 1999

Abstract

In this paper, we are concerned with a primal-dual interior point method for solving nonlinearly constrained optimization problems, in which Newton-like methods are applied to the shifted barrier KKT conditions. We propose a new primal-dual merit function, which is called the primal-dual l_2 barrier penalty function, within the framework of line search methods, and show the global convergence property of our method. Furthermore, by carefully controlling parameters in the algorithm, its superlinear convergence property is shown.

1 Introduction

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

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

where we assume that the functions $f : \mathbf{R}^n \to \mathbf{R}^1$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice continuously differentiable, and I_P is a subset of the index set $\{1, 2, ..., n\}$. Let $n' = |I_P| > 0$ and E be a $n' \times n$ matrix whose rows consist of $e_i^t,\ i \in I_P$, where $e_i \in \mathbf{R}^n$ denotes the i-th column vector of the identity matrix. Then problem (1) is written as:

$$(2)\qquad \begin{array}{lll} \text{minimize} & f(x), & x \in \mathbf{R}^n, \\ \text{subject to} & g(x) = 0, & Ex \geq 0. \end{array}$$

In the sequel, we use the notation

$$x' \equiv Ex \ \in \mathbf{R}^{n'}$$

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

†Department of Applied Mathematics, Faculty of Science, Science University of Tokyo, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, Japan. yabe@am.kagu.sut.ac.jp

for simplicity.

Let the Lagrangian function of the above problem be defined by

(3) $$L(w) = f(x) - y^t g(x) - z^t E x = f(x) - y^t g(x) - z^t x',$$

where $w = (x, y, z)^t$, and $y \in \mathbf{R}^m$ and $z \in \mathbf{R}^{n'}$ are the Lagrange multiplier vectors which correspond to the equality and inequality constraints respectively. Then Karush-Kuhn-Tucker (KKT) conditions for optimality of the above problem are given by

(4) $$r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X' Z e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and

(5) $$x' \geq 0, \qquad z \geq 0,$$

where

$$\begin{aligned} \nabla_x L(w) &= \nabla f(x) - A(x)^t y - E^t z, \\ A(x) &= \begin{pmatrix} \nabla g_1(x)^t \\ \vdots \\ \nabla g_m(x)^t \end{pmatrix}, \\ X' &= \operatorname{diag}(x'_1, \cdots, x'_{n'}), \\ Z &= \operatorname{diag}(z_1, \cdots, z_{n'}), \\ e &= (1, \cdots, 1)^t \in \mathbf{R}^{n'}. \end{aligned}$$

To solve the above problem by a primal-dual interior point method, Yamashita [15] introduces the barrier penalty function $F(\bullet, \mu) : S \to \mathbf{R}^1$ which is defined by

(6) $$F(x, \mu) = f(x) - \mu \sum_{i=1}^{n'} \log x'_i + \rho \sum_{i=1}^{m} |g_i(x)|,$$

where μ and ρ are given positive constants, and $S = \{x \in \mathbf{R}^n \,|x' > 0\}$. If ρ is sufficiently large, the necessary condition for the optimality of the barrier penalty function minimization problem for a given $\mu > 0$ is

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

and $x' > 0,\ z > 0$. The above conditions are called the barrier KKT conditions. The search direction of the proposed method is based on the Newton step for solving the equality part of the barrier KKT conditions. Let $\Delta w = (\Delta x, \Delta y, \Delta z)^t$ be defined by a solution of

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

and

$$J(w) = \begin{pmatrix} G & -A(x)^t & -E^t \\ A(x) & 0 & 0 \\ ZE & 0 & X' \end{pmatrix}, \tag{9}$$

where we use the relation $X'Ze = X'z = ZEx$. The matrix G is $\nabla_x^2 L(w)$ or a quasi-Newton approximation to the Hessian matrix.

Let $\Delta F_l(x, \mu; s)$ be a first order approximation to the quantity $F(x+s,\mu) - F(x,\mu)$, i.e.,

$$\Delta F_l(x,\mu;s) \equiv \nabla f(x)^t s - \mu e^t (X')^{-1} E s + \rho \sum_{i=1}^{m} \left| g_i(x) + \nabla g_i(x)^t s \right| - \rho \sum_{i=1}^{m} |g_i(x)| . \tag{10}$$

Then it is possible to prove that

$$\Delta F_l(x,\mu;\Delta x) \le -\Delta x^t (G + E^t (X')^{-1} Z E) \Delta x - \sum_{i=1}^{m} (\rho - |y_i + \Delta y_i|) |g_i(x)| .$$

The above inequality shows that the direction Δx which is derived from (8) can be a descent direction of the barrier penalty function $F(x,\mu)$ if G is positive definite and ρ is sufficiently large. Based on this observation, the line search algorithm and the trust region algorithm for the primal variable are proposed by Yamashita[15] and Yamashita et al.[16, 18] respectively. For the variable z, the step size is controlled by a box constraint. The step size for the variable y is usually taken equal to the one for z. Both algorithms are shown to be quite efficient. Some researchers have dealt with other primal merit functions within the framework of line search strategies or trust region strategies (See for example, Breitfield and Shanno[4], Dennis, Heinkenschloss and Vicente[8], Byrd, Gilbert and Nocedal[5], and Akrotirianakis and Rustem[1, 2]). Superlinear convergence properties of primal-dual methods based on solving the barrier KKT conditions have been also studied by several authors, for example, Martinez, Parada and Tapia[12], El-Bakry, Tapia, Tsuchiya and Zhang[9], Yamashita and Yabe[17], Yabe and Yamashita[13], Yamashita, Yabe and Tanabe[18], and Byrd, Liu and Nocedal[7].

In this paper, we consider a more conventional merit function:

$$F_0(x,\mu) = f(x) - \mu \sum_{i=1}^{n'} \log x_i' + \frac{1}{2\mu} \sum_{i=1}^{m} g_i(x)^2, \tag{11}$$

which is extensively described in a book by Fiacco and McCormick [10]. We also call this function the barrier penalty function. To discriminate this function from (6), we may call this the l_2 barrier penalty function. Whereas the function defined in (6) may be called the l_1 barrier penalty function.

The necessary condition for the optimality of the problem

$$\text{minimize} \quad F_0(x,\mu), \quad x \in S$$

is

$$\nabla F_0(x,\mu) = \nabla f(x) - \mu E^t (X')^{-1} e + \frac{1}{\mu} \sum_{i=1}^{m} g_i(x) \nabla g_i(x) = 0 \tag{12}$$

and $x' > 0$. As in [15], we introduce the variables y and z by $y = -g(x)/\mu$ and $z = \mu(X')^{-1}e$. Then the above conditions are written as

$$r(w,\mu) \equiv \begin{pmatrix} \nabla f(x) - A(x)^t y - E^t z \\ g(x) + \mu y \\ X'Ze - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{13}$$

and $x' > 0, z > 0$. We call these conditions the shifted barrier KKT (SBKKT) conditions. These conditions are also considered by Forsgren and Gill [11]. We do not consider the interior conditions $x' > 0, z > 0$ hereafter assuming these conditions are always satisfied.

In what follows, the subscript k denotes an iteration count in the inner iteration or in the outer iteration. Let $\|\cdot\|$ denote the l_2 norm for vectors and the operator norm induced from the l_2 vector norm for matrices. Let $\mathbf{R}_+^{n'} = \{z \in \mathbf{R}^{n'} |\ z > 0\}$.

2 Algorithm and its global convergence

2.1 Outer iteration

A prototype of the algorithm that uses the SBKKT conditions is described as follows.

Algorithm IP

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

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

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

$$\|r(w_{k+1}, \mu_k)\| \le M_c \mu_k. \tag{14}$$

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

We note that the barrier parameter sequence $\{\mu_k\}$ in Algorithm IP need not be determined beforehand. The value of each μ_k may be set adaptively as the iteration proceeds. We call condition (14) the approximate SBKKT condition, and call a point that satisfies this condition the approximate SBKKT point.

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

Theorem 1 *Let $\{w_k\}$ be an infinite sequence generated by Algorithm IP. Suppose that the sequences $\{x_k\}$ and $\{y_k\}$ are bounded. Then $\{z_k\}$ is bounded, and any accumulation point of $\{w_k\}$ satisfies KKT conditions (4) and (5).*

Proof. Assume that there exists an i such that $(E^t z_k)_i \to \infty$. Equation (14) yields

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

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

Let $\hat{w}$ be any accumulation point of $\{w_k\}$. Since the sequences $\{w_k\}$ and $\{\mu_k\}$ satisfy (14) for each k and μ_k approaches zero, $r_0(\hat{w}) = 0$ follows from the definition of $r(w, \mu)$. Therefore the proof is complete. □

2.2 Solving the shifted barrier KKT conditions

In this subsection we consider a method for solving the SBKKT conditions approximately for a given $\mu > 0$ (Step 2 of Algorithm IP). Therefore the index k denotes the inner iteration count for a given $\mu > 0$ in this subsection. The Newton-like iteration for solving (13) is defined by

(15) $$J_k \Delta w_k = -r(w_k, \mu),$$

where the Jacobian matrix J_k is given by

(16) $$J_k = \begin{pmatrix} G_k & -A(x_k)^t & -E^t \\ A(x_k) & \mu I & 0 \\ Z_k E & 0 & X_k' \end{pmatrix},$$

and the matrix G_k is $\nabla_x^2 L(w_k)$ or its approximation. The following lemma gives a sufficient condition for equation (15) to be solvable.

Lemma 1 *If G_k is positive definite, then the matrix J_k is nonsingular.*

Proof. Consider the equation

$$J_k \begin{pmatrix} \delta x \\ \delta y \\ \delta z \end{pmatrix} = 0,$$

for $(\delta x, \delta y, \delta z)^t \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^{n'}$. Then we have

$$\begin{aligned} (G_k + E^t (X_k')^{-1} Z_k E + \frac{1}{\mu} A(x_k)^t A(x_k)) \delta x &= 0, \\ \delta y &= -\mu^{-1} A(x_k) \delta x, \\ \delta z &= -(X_k')^{-1} Z_k E \delta x. \end{aligned}$$

By the assumption we obtain $\delta x = 0$, and therefore $\delta y = 0$ and $\delta z = 0$. This proves the lemma. □

We note that by eliminating Δy_k and Δz_k from the first set of equations (15):

(17) $$G_k \Delta x_k - A(x_k)^t \Delta y_k - E^t \Delta z_k = -\nabla_x L(w_k),$$

using the second and third sets of the equations:

$$A(x_k)\Delta x_k + \mu\Delta y_k = -g(x_k) - \mu y_k, \tag{18}$$

$$Z_k E\Delta x_k + X_k'\Delta z_k = \mu e - X_k' z_k, \tag{19}$$

we have

$$(G_k + E^t(X_k')^{-1}Z_k E + \frac{1}{\mu}A(x_k)^t A(x_k))\Delta x_k = -\nabla F_0(x_k, \mu). \tag{20}$$

Therefore it is easy to see that under appropriate assumptions the function $F_0(x, \mu)$ can be used as a merit function as in [15]. Because $F_0(x, \mu)$ depends only on the primal variables, we should use a method similar to the one which is given in [15] for controlling the step sizes for dual variables. Instead of following this possibility, we consider a merit function in the primal-dual space in this paper. Some primal-dual merit functions have been proposed (See for example, Argaez and Tapia[3], and El-Bakry, Tapia, Tsuchiya and Zhang[9] for solving the barrier KKT conditions (7), and Forsgren and Gill[11] for solving the SBKKT conditions (13)).

To have a merit function which has a minimum point at the SBKKT point, and which gives a descent direction with a Newton step, it is natural to consider

$$F_0(x, \mu) + \frac{\rho}{2}\left\|g(x) + \mu y\right\|^2 + \frac{\rho}{2}\left\|X'z - \mu e\right\|^2,$$

where ρ is a positive constant. We note that the second and third terms correspond to the second and third components in $r(w, \mu)$ respectively. However, this function does not prevent each component of the variable z tend to 0, and therefore cannot give a globally convergent algorithm unless an appropriate procedure is devised. Thus we need a sort of the barrier term for the variable z. In this paper we propose the following function which is called the primal-dual barrier penalty function:

$$F(w, \mu) = F_0(x, \mu) + \rho \log \frac{\{(x')^t z\}^\nu / n' + \left\|g(x) + \mu y\right\|^2 + \left\|X'z - \mu e\right\|^2}{\left(\prod_{i=1}^{n'} x_i' z_i\right)^{\nu/n'}}, \tag{21}$$

where $\rho > 0$ and $\nu \in (0, 2)$ are constants, which is a modification of the primal-dual merit function proposed by Yamashita[14]. The denominator in the second term is to prevent z_i tend to 0 for each i. For notational convenience we denote the expression in the last term in (21) by $\rho\phi(w)$, i.e.,

$$\begin{aligned}\phi(w) &\equiv \log \frac{\{(x')^t z\}^\nu / n' + \left\|g(x) + \mu y\right\|^2 + \left\|X'z - \mu e\right\|^2}{\left(\prod_{i=1}^{n'} x_i' z_i\right)^{\nu/n'}} \\ &= \log\left(\{(x')^t z\}^\nu / n' + \left\|g(x) + \mu y\right\|^2 + \left\|X'z - \mu e\right\|^2\right) - \frac{\nu}{n'}\sum_{i=1}^{n'} \log x_i' z_i.\end{aligned} \tag{22}$$

For later convenience we quote two well known relations

$$\frac{(x')^t z}{n'} \geq \left(\prod_{i=1}^{n'} x_i' z_i\right)^{1/n'}, \tag{23}$$

(24)
$$\sum_{i=1}^{n'} \frac{1}{n' x'_i z_i} \;\geq\; \frac{1}{\left(\prod_{i=1}^{n'} x'_i z_i\right)^{1/n'}}.$$

In the above inequalities, the equalities hold if and only if $x'_1 z_1 = \cdots = x'_{n'} z_{n'}$.

From (23), it is easy to prove the following lemma.

Lemma 2 *There hold:*

(i) $\phi(w) \geq 0$.

(ii) $\phi(w) = 0$ *if and only if* $g(x) + \mu y = 0$ *and* $X'z - \mu e = 0$. □

Now we calculate the derivatives of the merit function:

(25)
$$\nabla F(w,\mu) = \begin{pmatrix} \nabla F_0(x,\mu) + \rho \nabla_x \phi(w) \\ \rho \nabla_y \phi(w) \\ \rho \nabla_z \phi(w) \end{pmatrix},$$

where

$$\begin{aligned}
\nabla_x \phi(w) &= \frac{\nu\{(x')^t z\}^{\nu-1} E^t z/n' + 2A(x)^t (g(x)+\mu y) + 2E^t Z(X'z - \mu e)}{\{(x')^t z\}^\nu/n' + \|g(x)+\mu y\|^2 + \|X'z - \mu e\|^2} - \frac{\nu E^t (X')^{-1} e}{n'}, \\
\nabla_y \phi(w) &= \frac{2\mu(g(x)+\mu y)}{\{(x')^t z\}^\nu/n' + \|g(x)+\mu y\|^2 + \|X'z - \mu e\|^2}, \\
\nabla_z \phi(w) &= \frac{\nu\{(x')^t z\}^{\nu-1} x'/n' + 2X'(X'z - \mu e)}{\{(x')^t z\}^\nu/n' + \|g(x)+\mu y\|^2 + \|X'z - \mu e\|^2} - \frac{\nu Z^{-1} e}{n'}.
\end{aligned}$$

Lemma 3 *There hold the following relations:*

(26)
$$\begin{aligned}
r(w,\mu) = 0 \;&\Longleftrightarrow\; \nabla F_0(x,\mu) = 0,\; g(x) + \mu y = 0,\; X'z - \mu e = 0 \\
&\Longleftrightarrow\; \nabla F(w,\mu) = 0.
\end{aligned}$$

Proof. The first equivalence is obvious from (12).

The second relation comes from (25). If $\nabla F_0(x,\mu) = 0$, $g(x)+\mu y = 0$ and $X'z - \mu e = 0$, then we have $\nabla F(w,\mu) = 0$. Conversely assume that $\nabla F(w,\mu) = 0$. Then it follows from the relations $\nabla_y \phi(w) = 0$ and $\nabla_z \phi(w) = 0$ that

$$g(x) + \mu y = 0$$

and

(27)
$$\frac{\nu\{(x')^t z\}^{\nu-1} x'/n' + 2X'(X'z - \mu e)}{\{(x')^t z\}^\nu/n' + \|X'z - \mu e\|^2} - \frac{\nu Z^{-1} e}{n'} = 0.$$

Equation (27) yields

$$\frac{\nu\{(x')^t z\}^{\nu-1} z/n' + 2Z(X'z - \mu e)}{\{(x')^t z\}^\nu/n' + \|X'z - \mu e\|^2} - \frac{\nu (X')^{-1} e}{n'} = 0,$$

which implies $\nabla_x \phi(w) = 0$ and we have

$$\nabla F_0(x, \mu) = \nabla_x F(w, \mu) = 0.$$

Equation (27) also yields

$$2(X'z - \mu e) = \frac{\nu}{n'} \left(\frac{\{(x')^t z\}^\nu}{n'} + \|X'z - \mu e\|^2 \right) (X'Z)^{-1} e - \frac{\nu\{(x')^t z\}^{\nu-1}}{n'} e.$$

Multiplying $(X'z - \mu e)^t$ to both sides of the above equality, we have

$$\begin{aligned} 2\|X'z - \mu e\|^2 &= \nu \left(\frac{\{(x')^t z\}^\nu}{n'} + \|X'z - \mu e\|^2 \right) - \frac{\nu\{(x')^t z\}^\nu}{n'} \\ &\quad - \frac{\mu\nu}{n'} \left(\frac{\{(x')^t z\}^\nu}{n'} + \|X'z - \mu e\|^2 \right) e^t (X'Z)^{-1} e + \mu\nu\{(x')^t z\}^{\nu-1} \\ &= \nu\|X'z - \mu e\|^2 + \mu\nu\{(x')^t z\}^{\nu-1} - \frac{\mu\nu}{n'} \left(\frac{\{(x')^t z\}^\nu}{n'} + \|X'z - \mu e\|^2 \right) e^t (X'Z)^{-1} e. \end{aligned}$$

Thus there holds

$$(2 - \nu)\|X'z - \mu e\|^2 = \mu\nu\{(x')^t z\}^{\nu-1} - \frac{\mu\nu}{n'} \left(\frac{\{(x')^t z\}^\nu}{n'} + \|X'z - \mu e\|^2 \right) e^t (X'Z)^{-1} e.$$

By (23) and (24), we have

$$\begin{aligned} \left(2 - \nu + \frac{\mu\nu}{n'} e^t (X'Z)^{-1} e \right) \|X'z - \mu e\|^2 &= \mu\nu\{(x')^t z\}^{\nu-1} - \mu\nu \frac{\{(x')^t z\}^\nu}{n'} \frac{e^t (X'Z)^{-1} e}{n'} \\ &\leq \mu\nu\{(x')^t z\}^{\nu-1} - \mu\nu\{(x')^t z\}^{\nu-1} \frac{\left(\prod_{i=1}^{n'} x_i' z_i\right)^{1/n'}}{\left(\prod_{i=1}^{n'} x_i' z_i\right)^{1/n'}} \\ &= 0, \end{aligned}$$

which implies $X'z - \mu e = 0$.

Therefore the proof is complete. □

In the following, we set $\Delta x' = E\Delta x$. To derive an upper bound on the directional derivative of F, we first calculate the one for ϕ.

(28)

$$\begin{aligned} &\nabla\phi(w)^t \Delta w \\ = &\frac{\nu\{(x')^t z\}^{\nu-1} \left(z^t \Delta x' + (x')^t \Delta z \right) / n' + 2(A(x)\Delta x + \mu\Delta y)^t (g(x) + \mu y) + 2\left(Z\Delta x' + X'\Delta z \right)^t (X'z - \mu}{\{(x')^t z\}^\nu / n' + \|g(x) + \mu y\|^2 + \|X'z - \mu e\|^2} \\ &- \frac{\nu}{n'} \sum_{i=1}^{n'} \frac{z_i \Delta x_i' + x_i' \Delta z_i}{x_i' z_i}. \end{aligned}$$

Lemma 4 *If Δw_k solves (15), then we have*

$$\nabla\phi(w_k)^t\Delta w_k \le -(2-\nu)\frac{\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}{\{(x'_k)^t z_k\}^\nu/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}. \tag{29}$$

Proof. From (28), we have

$$\begin{aligned}\nabla\phi(w_k)^t\Delta w_k &= \frac{\nu\{(x')^t z\}^{\nu-1}(\mu-(x'_k)^t z_k/n')-2\|g(x_k)+\mu y_k\|^2-2\|X'_k z_k-\mu e\|^2}{\{(x')^t z\}^\nu/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}\\ &\quad -\nu\sum_{i=1}^{n'}\frac{\mu-(x'_k)_i(z_k)_i}{n'(x'_k)_i(z_k)_i}\\ &= \frac{\mu\nu\{(x')^t z\}^{\nu-1}-(2-\nu)\|g(x_k)+\mu y_k\|^2-(2-\nu)\|X'_k z_k-\mu e\|^2}{\{(x')^t z\}^{1+\nu}/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}\\ &\quad -\sum_{i=1}^{n'}\frac{\mu\nu}{n'(x'_k)_i(z_k)_i}.\end{aligned}$$

From relations (23) and (24), we obtain

$$\begin{aligned}&\frac{\mu\nu\{(x')^t z\}^{\nu-1}-(2-\nu)\|g(x_k)+\mu y_k\|^2-(2-\nu)\|X'_k z_k-\mu e\|^2}{\{(x')^t z\}^{\nu}/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}-\sum_{i=1}^{n'}\frac{\mu\nu}{n'(x'_k)_i(z_k)_i}\\ \le\ &\frac{n'\mu\nu}{(x'_k)^t z_k}-\frac{\mu\nu}{\left(\prod_{i=1}^{n'}(x'_k)_i(z_k)_i\right)^{1/n'}}-(2-\nu)\frac{\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}{(x'_k)^t z_k/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}\\ \le\ &-(2-\nu)\frac{\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}{(x'_k)^t z_k/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}.\end{aligned}$$

This proves the lemma. □

Lemma 5 *If Δw_k solves (15), then we have*

$$\begin{aligned}\nabla F(w_k,\mu)^t\Delta w_k &\le -\Delta x_k^t(G_k+E^t(X'_k)^{-1}Z_kE+\frac{1}{\mu}A(x_k)^tA(x_k))\Delta x_k\\ &\quad -\rho(2-\nu)\frac{\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}{(x'_k)^t z_k/n'+\|g(x_k)+\mu y_k\|^2+\|X'_k z_k-\mu e\|^2}.\end{aligned}$$

Proof. From (20) and (25), we obtain

$$\begin{aligned}\nabla F(w_k,\mu)^t\Delta w_k &= -\Delta x_k^t(G_k+E^t(X'_k)^{-1}Z_kE+\frac{1}{\mu}A(x_k)^tA(x_k))\Delta x_k\\ &\quad +\rho\nabla\phi(w_k)^t\Delta w_k\end{aligned}$$

which proves the lemma from (29). □

Lemma 6 *Assume that* Δw_k *solves (15). If* $\Delta x_k = 0$, $g(x_k)+\mu y_k = 0$ *and* $X_k' z_k - \mu e = 0$, *then* w_k *is an SBKKT point.*

Proof. $\Delta x_k = 0$ means $\nabla F_0(x_k, \mu) = 0$ from (20). Thus from (26), $r(w_k, \mu) = 0$ follows. □

We note that this lemma shows that if G_k is positive definite and w_k is not an SBKKT point, then the direction Δw_k is a descent direction for the primal-dual barrier penalty function from Lemma 5.

2.3 Line search algorithm

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

$$w_{k+1} = w_k + \alpha_k \Delta w_k,$$

where α_k is a step size determined by the line search procedure described below.

The main iteration is to decrease the value of the primal-dual barrier penalty function $F(w, \mu)$ for fixed μ. Thus the step size is determined by the sufficient decrease rule of the merit function. We adopt Armijo's rule. At the point w_k, we calculate the maximum allowed step to the boundary of the feasible region by

$$\alpha_{k\max} = \min \left\{ \min_i \left\{ -\frac{(x_k')_i}{(\Delta x_k')_i} \middle| (\Delta x_k')_i < 0 \right\}, \min_i \left\{ -\frac{(z_k)_i}{(\Delta z_k)_i} \middle| (\Delta z_k)_i < 0 \right\} \right\}.$$

A step to the next iterate is given by

$$\alpha_k = \bar{\alpha}_k \beta^{l_k}, \qquad \bar{\alpha}_k = \min \left\{ \gamma \alpha_{k\max}, 1 \right\},$$

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

$$F(w_k + \bar{\alpha}_k \beta^{l_k} \Delta w_k, \mu) - F(w_k, \mu) \le \varepsilon_0 \bar{\alpha}_k \beta^{l_k} \nabla F(w_k, \mu)^t \Delta w_k,$$

where $\varepsilon_0 \in (0,1)$.

Now we give the line search algorithm, which is called Algorithm LS. This algorithm can be regarded as the inner iteration of Algorithm IP (see Step 2 of Algorithm IP).
Algorithm LS

Step 0. (Initialize) Let $w_0 \in S \times \mathbf{R}^m \times \mathbf{R}_+^{n'}$, and $\mu > 0$, $\rho > 0$. Set $\varepsilon' > 0$, $\gamma \in (0,1)$, $\beta \in (0,1)$, $\varepsilon_0 \in (0,1)$. 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 (15).

Step 3. (Step size) Calculate

(30)

$$\alpha_{k\max} = \min\left\{\min_i\left\{-\frac{(x'_k)_i}{(\Delta x'_k)_i}\middle|(\Delta x'_k)_i<0\right\},\min_i\left\{-\frac{(z_k)_i}{(\Delta z_k)_i}\middle|(\Delta z_k)_i<0\right\}\right\},$$

(31) $$\bar{\alpha}_k = \min\{\gamma\alpha_{k\max}, 1\}.$$

Find the smallest nonnegative integer l_k that satisfies

(32) $$F(w_k+\bar{\alpha}_k\beta^{l_k}\Delta w_k,\mu) - F(w_k,\mu) \le \varepsilon_0\bar{\alpha}_k\beta^{l_k}\nabla F(w_k,\mu)^t\Delta w_k.$$

Calculate

$$\alpha_k = \bar{\alpha}_k\beta^{l_k}.$$

Step 4. (Update variables) Set

$$w_{k+1} = w_k + \alpha_k\Delta w_k.$$

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

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

Assumption G

(G1) The functions f and $g_i, i = 1, ..., m$, are twice continuously differentiable.

(G2) The level set of the primal-dual barrier penalty function $F(w,\mu)$ at an initial point $w_0 \in S\times\mathbf{R}^m\times\mathbf{R}_+^{n'}$, which is defined by $\left\{w\in S\times\mathbf{R}^m\times\mathbf{R}_+^{n'} \mid F(w,\mu)\le F(w_0,\mu)\right\}$, is compact for given $\mu>0$.

(G3) The matrix G_k is uniformly positive definite and uniformly bounded. □

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 (G1). We also note that for the case of $n' = n$, Assumption (G3) can be replaced by the following weaker condition:

(G3)′ The matrix G_k is positive semi-definite and uniformly bounded.

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 an SBKKT point.*

Proof. Since the sequence $\{F(w_k,\mu)\}$ is decreasing, each component of the sequence $\{x'_k\}$ is bounded away from zero and bounded above by the existence of the log barrier term and the assumption. The sequence $\{z_k\}$ also has these properties. Thus there exists a positive number M such that

$$\frac{\|v\|^2}{M} \le v^t(G_k + E^t(X'_k)^{-1}Z_kE)v \le M\|v\|^2, \quad \forall v \in \mathbf{R}^n, \tag{33}$$

by the assumption. From Lemma 5 and (33), we have

$$\nabla F(w_k,\mu)^t \Delta w_k \le -\frac{\|\Delta x_k\|^2}{M} + \rho\nabla\phi(w_k)^t\Delta w_k < 0, \tag{34}$$

and from (32),

$$\begin{aligned} F(w_{k+1},\mu) - F(w_k,\mu) &\le \varepsilon_0\bar{\alpha}_k\beta^{l_k}\nabla F(w_k,\mu)^t\Delta w_k \qquad (35)\\ &\le -\varepsilon_0\bar{\alpha}_k\beta^{l_k}\left(\frac{\|\Delta x_k\|^2}{M} - \rho\nabla\phi(w_k)^t\Delta w_k\right)\\ &< 0. \qquad (36)\end{aligned}$$

Because the sequence $\{F(w_k,\mu)\}$ is decreasing and bounded below, the left hand side of (35) converges to 0. From (33) and (20), $\|\Delta x_k\|$ is uniformly bounded above. Then from (18) and (19) for Δy_k and Δz_k, we conclude that $\|\Delta w_k\|$ is uniformly bounded above. Since $\liminf_{k\to\infty}(x'_k)_i > 0, \liminf_{k\to\infty}(z_k)_i > 0, i = 1,\cdots,p$, we have $\liminf_{k\to\infty}\bar{\alpha}_k > 0$.

Suppose that there exists a subsequence $K \subset \{0,1,\cdots\}$ and a δ such that

$$\liminf_{k\to\infty}\left|\nabla F(w_k,\mu)^t\Delta w_k\right| \ge \delta > 0, \quad k \in K. \tag{37}$$

Then we have $l_k \to \infty, k \in K$ from (35) because the left most expression tends to zero, and therefore we can assume $l_k > 0$ for sufficiently large $k \in K$ without loss of generality. If $l_k > 0$ then the point $w_k + \alpha_k\Delta w_k/\beta$ does not satisfy condition (32). Thus, we have

$$F(w_k + \alpha_k\Delta w_k/\beta,\mu) - F(w_k,\mu) > \varepsilon_0\alpha_k\nabla F(w_k,\mu)^t\Delta w_k/\beta. \tag{38}$$

By the mean value theorem, there exists a $\theta_k \in (0,1)$ such that

$$F(w_k + \alpha_k\Delta w_k/\beta,\mu) - F(w_k,\mu) = \alpha_k\nabla F(w_k + \theta_k\alpha_k\Delta w_k/\beta,\mu)^t\Delta w_k/\beta. \tag{39}$$

Then, from (38) and (39), we have

$$\varepsilon_0\nabla F(w_k,\mu)^t\Delta w_k < \nabla F(w_k + \theta_k\alpha_k\Delta w_k/\beta,\mu)^t\Delta w_k.$$

This inequality yields

$$\begin{aligned} &\nabla F(w_k + \theta_k\alpha_k\Delta w_k/\beta,\mu)^t\Delta w_k - \nabla F(w_k,\mu)^t\Delta w_k \qquad (40)\\ > \;&(\varepsilon_0 - 1)\nabla F(w_k,\mu)^t\Delta w_k > 0.\end{aligned}$$

Thus by the property $l_k \to \infty$, we have $\|\theta_k\alpha_k\Delta w_k/\beta\| \to 0, k \in K$, because $\|\Delta w_k\|$ is uniformly bounded above. Thus the left hand side of (40) and therefore $\nabla F(w_k,\mu)^t\Delta w_k$

converges to zero when $k \to \infty, k \in K$. This contradicts assumption (37). Therefore we proved

(41) $$\lim_{k\to\infty} \nabla F(w_k, \mu)^t \Delta w_k = 0.$$

This implies that

(42) $$\Delta x_k \to 0, \quad g(x_k) + \mu y_k \to 0, \quad X'_k z_k - \mu e \to 0,$$

from (34). We should note that the existence of an accumulation point of the sequence $\{w_k\}$ is assured by Assumption (G2). Let an arbitrary accumulation point of the sequence $\{w_k\}$ be $\hat{w} = (\hat{x}, \hat{y}, \hat{z})^t \in S \times \mathbf{R}^m \times \mathbf{R}^{n'}_+$. Then from (42), we have

$$\hat{y} = -\frac{g(\hat{x})}{\mu}, \ \hat{z} = \mu(\hat{X}')^{-1}e,$$

where $\hat{X}' = \text{diag}(\hat{x}'_1, \cdots, \hat{x}'_{n'})$. Because $\Delta x_k \to 0$ implies $\nabla F_0(\hat{x}, \mu) = 0$ from (20), we have $r(\hat{w}, \mu) = 0$ from (26). □

3 Superlinear Convergence

In the previous section, we have proved the global convergence property of Algorithm IP. In this section, we discuss under which condition Algorithm IP can possess the superlinear convergence property. For this purpose, we first consider the following local algorithm, which is called Algorithm IPlocal. By appropriately controlling the parameters μ_k and γ_k at each step near a KKT point, we can show that the unit Newton-like step from an approximate SBKKT point yields a next approximate SBKKT point that corresponds to the new updated barrier parameter, and that the sequence $\{w_k\}$ generated by Algorithm IPlocal converges superlinearly to the KKT point.

Algorithm IPlocal

Step 0. (Initialize) Set $w_0 \in S \times \mathbf{R}^m \times \mathbf{R}^{n'}_+$, $\mu_0 > 0$, $0 < \gamma_0 < 1$ and $\varepsilon > 0$. Let $k = 0$.

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

Step 2. (Update the parameters) Choose the parameters $\mu_k > 0$ and $0 < \gamma_k < 1$.

Step 3. (Compute direction) Calculate the direction Δw_k by the linear system of equations

(43) $$J_k \Delta w_k = -r(w_k, \mu_k),$$

where the matrix J_k is given by

(44) $$J_k = \begin{pmatrix} G_k & -A(x_k)^t & -E^t \\ A(x_k) & \mu_k I & 0 \\ Z_k E & 0 & X'_k \end{pmatrix}.$$

Step 4. (Step size) Set

$$\begin{aligned}\alpha_{k\max} &= \min\left\{\min_i\left\{-\frac{(x'_k)_i}{(\Delta x'_k)_i}\middle|(\Delta x'_k)_i<0\right\},\min_i\left\{-\frac{(z_k)_i}{(\Delta z_k)_i}\middle|(\Delta z_k)_i<0\right\}\right\},\\ \alpha_k &= \min\{\gamma_k\alpha_{k\max},1\}.\end{aligned}$$

Step 5. (Update variables) Set

$$w_{k+1}=w_k+\alpha_k\Delta w_k.$$

Step 6. Set $k := k+1$ and go to Step 1. □

Denote the Jacobian matrix of $r(w,\mu)$ by

$$\nabla r(w,\mu)=\begin{pmatrix}\nabla_x^2 L(w) & -A(x)^t & -E^t\\ A(x) & \mu I & 0\\ ZE & 0 & X'\end{pmatrix}.$$

Let $w^*=(x^*,y^*,z^*)^t$ be a KKT point of (1). In the following, we assume that k is sufficiently large and μ_k is sufficiently close to 0. In order to prove superlinear convergence, we need Assumption L.

Assumption L

(L1) The sequence $\{w_k\}$ converges 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^*.

(L4) μ_k and γ_k are updated by the rules

$$\mu_k=\xi_k\|r_0(w_k)\|^{1+\tau_1}\quad\text{and}\quad 1-\gamma_k=\kappa\xi_k\|r_0(w_k)\|^{\tau_2}$$

for positive constants τ_1, τ_2 and κ such that $\min(1,\tau_2)>\tau_1$ and $0<\kappa<1$, and for a positive number ξ_k such that $\frac{1}{M'}\le\xi_k\le M'$, where M' is a positive constant.

(L5) The matrix G_k satisfies at each k,

$$\|G_k-\nabla_x^2 L(w^*)\|<\delta$$

for sufficiently small $\delta>0$, and

$$\|(G_k-\nabla_x^2 L(w_k))\Delta x_k\|=\mathrm{O}(\|\Delta w_k\|^{1+\tau_3})$$

for some positive constant τ_3 such that $\tau_3>\tau_1$. □

First we should note that by (L3), the Jacobian matrix $\nabla r_0(w^*)$ is nonsingular. Then by (L2), (L4) and (L5), we have

$$\begin{aligned}\|J_k - \nabla r_0(w^*)\| &\le \|\nabla r_0(w_k) - \nabla r_0(w^*)\| + \|G_k - \nabla_x^2 L(w^*)\| + \mu_k \\ &\le \|\nabla r_0(w_k) - \nabla r_0(w^*)\| + \delta + M'\|r_0(w_k)\|^{1+\tau_1}.\end{aligned}$$

Since for sufficiently large k and sufficiently small δ, there holds

$$\|\nabla r_0(w^*)^{-1}\|\|J_k - \nabla r_0(w^*)\| < 1,$$

J_k is nonsingular and we have

$$\|J_k^{-1}\| \le \nu$$

for a positive constant ν by Banach perturbation lemma. Thus the linear system of equations (43) has a unique solution.

Now we give the following theorem, which is very important for proving the superlinear convergence property of Algorithm IPlocal. This theorem shows that if w_k satisfies the approximate SBKKT condition for μ_{k-1}, then α_k is set to be unit in Step 4 of Algorithm IPlocal and $w_k + \Delta w_k$ also satisfies the approximate SBKKT condition for μ_k.

Theorem 3 *Let M_c be a constant such that $0 < M_c < \sqrt{n'}$.*
(1) If a point $\hat{w} = (\hat{x}, \hat{y}, \hat{z})^t \in S \times \mathbf{R}^m \times \mathbf{R}_+^{n'}$ satisfies $\|r(\hat{w}, \mu_k)\| \le M_c\mu_k$, then

$$\nu_1\|r_0(w_k)\|^{1+\tau_1} \le \|r_0(\hat{w})\| \le \nu_2\|r_0(w_k)\|^{1+\tau_1} \tag{45}$$

for positive constants ν_1 and ν_2.
(2) If $\|r(w_k, \mu_{k-1})\| \le M_c\mu_{k-1}$, then $\alpha_k = 1$.
(3) There holds

$$\|r(w_k + \Delta w_k, \mu_k)\| \le M_c\mu_k. \tag{46}$$

Proof. (1) Since $\|r(\hat{w}, \mu_k)\| \le M_c\mu_k$, we have

$$\|r_0(\hat{w})\| = \left\| r(\hat{w}, \mu_k) + \mu_k \begin{pmatrix} 0 \\ -\hat{y} \\ e \end{pmatrix} \right\| = \mathrm{O}(\mu_k) = \mathrm{O}(\|r_0(w_k)\|^{1+\tau_1}).$$

Furthermore we obtain

$$\begin{aligned}\|r_0(\hat{w})\| &= \left\| r(\hat{w}, \mu_k) + \mu_k \begin{pmatrix} 0 \\ -\hat{y} \\ e \end{pmatrix} \right\| \ge \mu_k \left\| \begin{pmatrix} 0 \\ -\hat{y} \\ e \end{pmatrix} \right\| - \|r(\hat{w}, \mu_k)\| \\ &= \mu_k\sqrt{\|\hat{y}\|^2 + \|e\|^2} - \|r(\hat{w}, \mu_k)\| \ge (\sqrt{n'} - M_c)\mu_k \\ &\ge \frac{\sqrt{n'} - M_c}{M'}\|r_0(w_k)\|^{1+\tau_1}.\end{aligned}$$

(2) We will show that

$$\gamma_k \min_i \left\{ -\frac{(x'_k)_i}{(\Delta x'_k)_i} \,\middle|\, (\Delta x'_k)_i < 0 \right\} \ge 1. \tag{47}$$

For i such that $(Ex^*)_i > 0$, it follows from $(\Delta x'_k)_i \to 0$ and $\gamma_k \to 1$ that

$$-\gamma_k \frac{(x'_k)_i}{(\Delta x'_k)_i} > 1 \quad \text{for} \quad (\Delta x'_k)_i < 0.$$

Now we consider an index i such that $(Ex^*)_i = 0$. In this case we note that $(z^*)_i > 0$ by Assumption (L3). By (43), we have

$$(x'_k)_i + (\Delta x'_k)_i = \frac{\mu_k}{(z_k)_i} - \frac{(x'_k)_i(\Delta z_k)_i}{(z_k)_i}. \tag{48}$$

Since $\|r(w_k, \mu_{k-1})\| \le M_c \mu_{k-1}$, we have

$$\mu_k \ge \frac{1}{M'} \|r_0(w_k)\|^{1+\tau_1} \ge \frac{\nu_1^{1+\tau_1}}{M'} \|r_0(w_{k-1})\|^{(1+\tau_1)^2} \tag{49}$$

by result (1), and

$$|(x'_k)_i(z_k)_i - \mu_{k-1}| \le M_c \mu_{k-1}.$$

The latter yields

$$(x'_k)_i \le \frac{(1+M_c)\mu_{k-1}}{(z_k)_i} = \frac{1+M_c}{(z_k)_i} \xi_{k-1} \|r_0(w_{k-1})\|^{1+\tau_1}.$$

Since

$$(\Delta z_k)_i \le \|\Delta w_k\| = \mathrm{O}(\|r(w_k, \mu_k)\|) = \mathrm{O}(\|r_0(w_k)\|) = \mathrm{O}(\|r_0(w_{k-1})\|^{1+\tau_1}),$$

we have

$$(x'_k)_i(\Delta z_k)_i = \mathrm{O}\left(\|r_0(w_{k-1})\|^{2(1+\tau_1)}\right). \tag{50}$$

Assumption (L4) implies $(1+\tau_1)^2 < 2(1+\tau_1)$. Thus it follows from (48), (49) and (50) that

$$(x'_k)_i + (\Delta x'_k)_i > \kappa \frac{\mu_k}{(z_k)_i}, \tag{51}$$

where κ is given by (L4). Since $(x'_k)_i(z_k)_i \le \|r_0(w_k)\|$, Assumption (L4) guarantees

$$\begin{aligned} \frac{\mu_k}{(z_k)_i} &= \frac{\xi_k \|r_0(w_k)\|^{1+\tau_1}}{(z_k)_i} \ge \xi_k (x'_k)_i \|r_0(w_k)\|^{\tau_1} \\ &\ge \xi_k (x'_k)_i \|r_0(w_k)\|^{\tau_2} = \frac{1}{\kappa}(x'_k)_i(1-\gamma_k), \end{aligned}$$

then we have

$$\kappa \frac{\mu_k}{(z_k)_i} \ge (x'_k)_i(1-\gamma_k). \tag{52}$$

Thus by (51) and (52) we obtain

$$(x'_k)_i + (\Delta x'_k)_i > (1-\gamma_k)(x'_k)_i,$$

which implies

$$\gamma_k \left(-\frac{(x'_k)_i}{(\Delta x'_k)_i}\right) > 1 \quad \text{for} \quad (\Delta x'_k)_i < 0.$$

Hence (47) holds.

In the same way as above, we can prove that

$$\gamma_k \min_i \left\{ -\frac{(z_k)_i}{(\Delta z_k)_i} \middle| (\Delta z_k)_i < 0 \right\} \geq 1.$$

Therefore the result follows.
(3) From Assumptions (L4) and (L5), we directly obtain

$$\begin{aligned}
\|r(w_k + \Delta w_k, \mu_k)\| &= \|r(w_k, \mu_k) + \nabla r(w_k, \mu_k)\Delta w_k + \mathrm{O}(\|\Delta w_k\|^2)\| \\
&\leq \|r(w_k, \mu_k) + J_k \Delta w_k\| + \mathrm{O}(\|\Delta w_k\|^2) \\
&\qquad + \|(J_k - \nabla r(w_k, \mu_k))\Delta w_k\| \\
&= \|(G_k - \nabla_x^2 L(w_k))\Delta x_k\| + \mathrm{O}(\|\Delta w_k\|^2) \\
&= \mathrm{O}(\|\Delta w_k\|^{\min(1+\tau_3,\ 2)}) \\
&= \mathrm{O}(\|r(w_k, \mu_k)\|^{\min(1+\tau_3,\ 2)}) \\
&= \mathrm{O}(\|r_0(w_k)\|^{\min(1+\tau_3,\ 2)}) \\
&= \mathrm{o}(\|r_0(w_k)\|^{1+\tau_1}) \\
&= \mathrm{o}(\mu_k) \\
&\leq M_c \mu_k.
\end{aligned}$$

This proves (46).

Therefore the proof of this theorem is complete. □

4 Global and superlinear convergence

Theorem 2 assures the global convergence of Algorithm LS to an SBKKT point for a fixed μ and therefore the global convergence of Algorithm IP to a KKT point of problem (1), while Theorem 3 implies the superlinear convergence of Algorithm IPlocal to a KKT point of problem (1). Specifically Theorem 3 shows that if w_k satisfies the approximate SBKKT condition for μ_{k-1}, then α_k is set to be unit in Step 4 of Algorithm IPlocal and $w_k + \Delta w_k$ also satisfies the approximate SBKKT condition for μ_k. Thus by result (1) of Theorem 3, the superlinear convergence property of Algorithm IPlocal can be obtained if we choose an approximate SBKKT point for μ_0 as an initial point.

However this does not necessarily imply the superlinear convergence of Algorithm IP, because the Armijo line search criterion required in the inner iteration (Algorithm LS) may prevent from choosing a unit step size even if the iterates are near a KKT point. This phenomenon is known as the Maratos effect. However if we adopt a unit step size when the current point w_k (the initial point for the k-th inner iteration) satisfies the approximate SBKKT condition for sufficiently small μ_{k-1}, and $w_k + \Delta w_k$ (the first step for the k-th inner iteration) satisfies the approximate SBKKT condition for μ_{k-1}, even when the merit function value does not satisfy the Armijo rule, then Theorems 2 and 3 assure that we can have the global and superlinear convergence of Algorithm IP by appropriately controlling the parameters μ_k and γ_k at the final stage of iterations.

We could devise an algorithm for avoiding the Maratos effect explicitly. For this purpose, we could use a nonmonotone strategy like the primal-dual interior point trust region method given by Yamashita, Yabe and Tanabe[18] for example. However we did not adopt the technique just for simplicity of exposition of the algorithm in this paper.

References

[1] I.Akrotirianakis and B.Rustem, *A globally convergent interior point algorithm for general nonlinear programming problems*, Technical Report 97-14, Department of Computing, Imperial College of Science, Technology and Medicine, revised July 1998.

[2] I.Akrotirianakis and B.Rustem, *A primal-dual interior point algorithm with an exact and differentiable function for general nonlinear programming problems*, Technical Report 98-09, Department of Computing, Imperial College of Science, Technology and Medicine, 1998.

[3] M.Argaez and R.A.Tapia, *On the global convergence of a modified augmented Lagrangian linesearch interior point Newton method for nonlinear programming*, Technical Report TR95-38, Department of Computational and Applied Mathematics, Rice University, October 1995 (revised February 1997).

[4] M.G.Breitfield and D.F.Shanno, Preliminary computational experience with modified log-barrier functions for large-scale nonlinear programming, in *Large Scale Optimization*, Kluwer academic publishers, Dordrecht, Boston, London, 1994.

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

[6] R.H.Byrd, M.E.Hribar and J.Nocedal, *An interior point algorithm for large scale nonlinear programming*, Technical Report OTC 97/05, Optimization Technology Center, Argonne National Laboratory, August, 1997.

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

[8] J.E.Dennis, Jr., M.Heinkenschloss and L.N.Vicente, *Trust-region interior-point SQP algorithms for a class of nonlinear programming problems*, TR94-45, Dept. of Computational and Applied Mathematics, Rice University, Houston, Texas, USA, 1994 (revised November 1995).

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

[10] A.V.Fiacco and G.P.McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, SIAM, Philadelphia, 1990.

[11] A.Forsgren and P.E.Gill, Primal-dual interior methods for nonconvex nonlinear programming, *SIAM J. on Optimization*, 8 (1998) pp.1132-1152.

[12] H.J.Martinez, Z.Parada and R.A.Tapia, On the characterization of Q-superlinear convergence of quasi-Newton interior-point methods for nonlinear programming, *Bol. Soc. Mat. Mexicana*, Vol.1 (1995), pp.137-148.

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

[14] H.Yamashita, A primal-dual exact merit function for constrained optimization, *Optimization – Modeling and Algorithms 8*, Cooperative Research Report 84, The Institute of Statistical Mathematics, March (1996), pp.119-127.

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

[16] H.Yamashita and T.Tanabe, A primal-dual interior point trust region method for large scale constrained optimization, *Optimization – Modeling and Algorithms 6*, Cooperative Research Report 73, The Institute of Statistical Mathematics, March (1995), pp.1-25.

[17] 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.

[18] H.Yamashita, H.Yabe and T.Tanabe, *A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization*, Technical Report, 1997 July (revised 1998 July).