Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 11, No. 4, pp. 721–736, October 2015

A PRIMAL-DUAL EXTERIOR POINT METHOD WITH A PRIMAL-DUAL QUADRATIC PENALTY FUNCTION FOR NONLINEAR OPTIMIZATION*

Yu Igarashi and Hiroshi Yabe

Abstract: In this paper, we consider a primal-dual exterior point method for solving nonlinearly constrained optimization problems, in which Newton-like methods are applied to the shifted smoothing KKT conditions. We propose a new primal-dual merit function, which is called the primal-dual quadratic penalty function, within the framework of line search strategy, and prove global convergence property of our method. Some numerical experiments are presented to show the performance of our method.

Key words: *nonlinear optimization, constrained optimization, primal-dual exterior point method, primal-dual differentiable quadratic penalty function, global convergence*

Mathematics Subject Classification: *65K05, 90C30*

---

1 Introduction

This paper is concerned with the following constrained optimization problem:

$$\text{minimize } f(x) \quad \text{subject to} \quad g(x) = 0,\ x \geq 0, \tag{1.1}$$

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

It is well known that primal-dual interior point methods perform effectively for solving problem (1.1) (see [8, 12, 13]). One of the shortcomings of the primal-dual interior point method is that the iterates should be kept strictly inside the interior region. So it is not easy to utilize a warm start in the interior point method framework. Therefore, it is of interest to consider an algorithm that does not need an interior point requirement. On the other hand, exterior point methods have also been studied by some researchers. For example, Andrus and Schaferkotter [3] applied this method to linear programming by using projection procedure and proved that their method is a polynomial time algorithm. Andrus [2] and Polyak [9] considered exterior point methods for convex optimization. Andrus [2] showed convergence property within the framework of gradient projection algorithm. Polyak [9] dealt with primal-dual exterior point method based on the nonlinear rescaling multipliers method and showed that this method generates a primal-dual sequence, which globally converges to the primal-dual solution with asymptotic quadratic rate. Griva [6] introduced an algorithm for

*This research is supported in part by the Grant-in-Aid for Scientific Research (C) 25330030 of Japan Society for the Promotion of Science.

solving nonlinear optimization. This algorithm uses the exterior point method when the interior point method fails to achieve the solution with the high level of accuracy. Yang [16] proposed an exterior point method for computing points that satisfy second-order necessary conditions for a certain optimization problem. Al-Sultan and Murty [1] considered the exterior penalty method for finding the nearest points in a convex polyhedral cone to a given point. They proved convergence property of their algorithm and extended it to convex quadratic programs.

Recently, Yamashita and Tanabe [14] paid attention to the exterior point method in order to solve the issues of the primal-dual interior point method. They define the following problem:

$$\text{minimize } F_{0YT}(x,\rho) = f(x) + \rho\sum_{i=1}^{n} |x_i|_- \quad \text{subject to} \quad g(x) = 0, \tag{1.2}$$

where $\rho > 0$ is a penalty parameter and $|x_i|_- = \max\{-x_i, 0\} = \dfrac{|x_i| - x_i}{2}$. They approximated this nondifferentiable function by a smooth differentiable function and proposed the l_1 penalty merit function. They showed the global convergence properties within the framework of line search strategy and trust region strategy, respectively. However, their merit function is still nondifferentiable and includes only primal variables. In this paper, to solve these points at issue, we propose a differentiable merit function and construct a new exterior point method based on it.

This paper is organized as follows. We will first review the idea of Yamashita and Tanabe [14], and propose a differentiable primal-dual merit function in Section 2. Next we show global convergence property of our algorithm within the framework of the line search strategy in Section 3. Finally preliminary numerical results will be presented in Section 4.

2 Preliminaries

Let the Lagrangian function of problem (1.1) be defined by

$$L(w) \equiv f(x) - y^T g(x) - z^T x,$$

where $w = (x, y, z)^T \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ 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 (1.1) are given by

$$\begin{pmatrix} \nabla_x L(w) \\ g(x) \\ XZe \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad x \geq 0 \quad \text{and} \quad z \geq 0, \tag{2.1}$$

where

$$\nabla_x L(w) = \nabla f(x) - A(x)^T y - z, \quad A(x) = (\nabla g_1(x) \cdots \nabla g_m(x))^T,$$

$$X = \operatorname{diag}(x_1, \cdots, x_n),\ Z = \operatorname{diag}(z_1, \cdots, z_n) \text{ and } e = (1, \ldots, 1)^T \in \mathbf{R}^n.$$

The next lemma gives the necessary conditions for optimality of problem (1.2). This lemma follows from Section 14 of [4].

Lemma 2.1. *The necessary conditions for optimality of problem* (1.2) *are given by*

$$\nabla_x L(w) = 0,\ g(x) = 0,\ z \in -\partial\left\{\rho\sum_{i=1}^{n}|x_i|_-\right\}, \tag{2.2}$$

where the symbol ∂ means the subdifferential of the function in the braces with respect to x.

We note that for $i = 1, \ldots, n$, the ith element of z in (2.2) is given by

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

These conditions are written as

$$|x_i| z_i - \rho |x_i|_- = 0, \; 0 \le z_i \le \rho, \; i = 1, \ldots, n. \tag{2.3}$$

Therefore, by using (2.3), conditions (2.2) can be expressed as

$$r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ r_C(w) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad 0 \le z \le \rho e, \tag{2.4}$$

where the ith element of $r_C(w) \in \mathbf{R}^n$ is defined by

$$r_C(w)_i \;\; = \;\; |x_i| z_i - \rho |x_i|_-. \tag{2.5}$$

The next lemma motivates that we consider conditions (2.4) instead of conditions (2.1), and it can be easily shown.

Lemma 2.2. *If* $\|z\|_\infty < \rho$, *conditions* (2.4) *are equivalent to conditions* (2.1).

Next, we consider a smooth approximation of problem (1.2). Yamashita and Tanabe [14] approximated the nondifferentiable function $|a|_-$, $a \in \mathbf{R}$, by the differentiable function:

$$h(a, \mu) = \frac{1}{2}(\sqrt{a^2 + \mu^2} - a), \tag{2.6}$$

where $\mu > 0$ is a parameter that controls the accuracy of the approximation. It is clear that $h(a, \mu) \to (|a| - a)/2 = |a|_-$ with $\mu \to 0$. The first and second derivatives of $h(a, \mu)$ are given by

$$h'(a, \mu) = \frac{1}{2}\left(\frac{a}{\sqrt{a^2 + \mu^2}} - 1\right) = -\frac{h(a, \mu)}{\sqrt{a^2 + \mu^2}}, \quad h''(a, \mu) = \frac{\mu^2}{2(a^2 + \mu^2)^{3/2}} \tag{2.7}$$

and we note that the functions (2.6) and (2.7) have the following properties

$$h(a, \mu) > 0, \; -1 < h'(a, \mu) < 0, \; h''(a, \mu) > 0.$$

By using $h(x, \mu)$, Yamashita and Tanabe approximated problem (1.2) by the problem:

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

The KKT conditions for the above problem are

$$\nabla f(x) - A(x)^T y + \rho H'(x, \mu) e = 0 \quad \text{and} \quad g(x) = 0,$$

where

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

By introducing the auxiliary variable z as $z = -\rho H'(x,\mu)e$, the KKT conditions are rewritten as

$$\nabla_x L(w) = 0,\ g(x) = 0 \quad \text{and} \quad z + \rho H'(x,\mu)e = 0.$$

Furthermore, by using (2.7), the third equation of the above conditions reduces to

$$\sqrt{x_i^2 + \mu^2} \cdot z_i - \rho h(x_i,\mu) = 0,\ i = 1,\dots,n. \tag{2.8}$$

Then (2.8) can be viewed as a smoothing approximation of (2.5), and we express the KKT conditions as

$$r(w,\mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ U(x,\mu)z - \rho H(x,\mu)e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{2.9}$$

where

$$u(x_i,\mu) = \sqrt{x_i^2 + \mu^2}, \quad i = 1,\dots,n, \quad U(x,\mu) = \operatorname{diag}(u(x_1,\mu),\dots,u(x_n,\mu)). \tag{2.10}$$

We note that $H'(x,\mu) = -U(x,\mu)^{-1}H(x,\mu)$ holds.

To globalize the algorithm, Yamashita and Tanabe [14] introduced the l_1 penalty function $F_{YT}(\bullet,\mu) : \mathbf{R}^n \to \mathbf{R}$ defined by

$$F_{YT}(x,\mu) = f(x) + \rho \sum_{i=1}^{n} h(x_i,\mu) + \rho' \sum_{i=1}^{m} |g_i(x)|, \tag{2.11}$$

where μ, ρ and ρ' are given positive constants.

Yamashita and Tanabe [14] proposed to use the function (2.11) as a merit function, based on the fact that if ρ is sufficiently large, the necessary condition for the optimality of the penalty function minimization problem for a given $\mu > 0$ is the KKT conditions (2.9). They showed the global convergence properties of their primal-dual exterior point method within the framework of line search strategy and trust region strategy, respectively. However, in spite of approximating the function $|x|_-$, this merit function is nondifferentiable even now. Furthermore, the merit function only depends on primal variable x.

On the other hand, following Forsgren and Gill [5], Yamashita and Yabe [15] dealt with a quadratic barrier penalty function

$$F_{qb}(x,\tilde{\mu}) = f(x) - \tilde{\mu} \sum_{i=1}^{n} \log x_i + \frac{1}{2\tilde{\mu}} \sum_{i=1}^{m} g_i(x)^2 \tag{2.12}$$

for the primal-dual interior point method, where $\tilde{\mu} > 0$ is a barrier parameter. By using this function, they showed the global convergence property of their primal-dual interior point method within the framework of line search strategy.

In this paper, we apply this idea to the exterior point method and introduce a differentiable penalty function (called as the quadratic penalty function):

$$F_0(x,\mu) = f(x) + \rho \sum_{i=1}^{n} h(x_i,\mu) + \frac{1}{2\mu} \sum_{i=1}^{m} g_i(x)^2. \tag{2.13}$$

We note that the smoothing parameter μ is also used in the quadratic penalty term as well as the barrier parameter $\tilde{\mu}$ was used in the quadratic penalty term in (2.12).

The necessary condition for the optimality of the problem

$$\text{minimize} \quad F_0(x, \mu), \quad x \in \mathbf{R}^n$$

is represented by

$$\nabla F_0(x, \mu) = \nabla f(x) + \rho \sum_{i=1}^{n} h'(x_i, \mu) e_i + \frac{1}{\mu} \sum_{i=1}^{m} g_i(x) \nabla g_i(x) = 0. \tag{2.14}$$

As in [14] and [15], we define the variables y and z by $y = -g(x)/\mu$ and $z = -\rho H'(x, \mu)e$. Then by (2.8) and (2.10), the above conditions are written as

$$r(w, \mu) \equiv \begin{pmatrix} \nabla f(x) - A(x)^T y - z \\ g(x) + \mu y \\ U(x, \mu) z - \rho H(x, \mu) e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.15}$$

In what follows, we treat x, y and z as independent variables. We call conditions (2.15) and $0 \leq z \leq \rho e$ the shifted KKT (SKKT) conditions. It should be noted that these conditions become (2.4) as $\mu \to 0$.

Throughout this paper, 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 $\mathbf{R}^n_\rho = \{ z \in \mathbf{R}^n | \ 0 \leq z \leq \rho e \}$.

3 Algorithm and Its Global Convergence

In this section, we propose a primal-dual exterior point method. We first give the algorithm of the outer iteration following [14] in Section 3.1. Next we present a method that finds an SKKT point in Section 3.2 and show its global convergence in Section 3.3.

3.1 Outer iteration

The algorithm of the outer iteration of the primal-dual exterior point method is described as follows.

Algorithm EP

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

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

$$\| r(w_{k+1}, \mu_k) \| \leq M_c \mu_k \quad \text{and} \quad 0 \leq z_{k+1} \leq \rho e. \tag{3.1}$$

Step 2. (Termination) If $\| r_0(w_{k+1}) \| \leq \varepsilon$, then stop.

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

We call condition (3.1) the approximate SKKT condition, and call a point that satisfies these conditions an approximate SKKT point. The following theorem shows the global convergence property of Algorithm EP, which can be proved in the same way as the proof of Theorem 1 of Yamashita and Tanabe [14].

Theorem 3.1. *Let $\{w_k\}$ be an infinite sequence generated by Algorithm EP. Then any accumulation point of $\{w_k\}$ satisfies KKT conditions of problem* (1.2).

For a given ρ, Algorithm EP always imposes the condition $0 \le z_{k+1} \le \rho e$ on the variable z and Theorem 3.1 guarantees the existence of a KKT point of problem (1.2). However this does not imply the existence of a KKT point of problem (1.1), because Lemma 2.2 does not necessarily hold if the parameter ρ is not large. This phenomenon will be found in Table 2 in Section 4.

3.2 Finding an approximate SKKT point

In this subsection, we consider a method for finding an approximate SKKT point for a given $\mu > 0$ (Step 1 of Algorithm EP). In what follows, the index k denotes the inner iteration count and we note that $0 \le z_k \le \rho e$ holds for all k. To find an approximate SKKT point, we use the Newton-like method.

The Newton equations for solving (2.15) are given by

$$J_k \Delta w_k = -r(w_k, \mu), \tag{3.2}$$

where $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^T$ is a search direction, the matrix J_k is defined by

$$J_k = \begin{pmatrix} G_k & -A(x_k)^T & -I \\ A(x_k) & \mu I & 0 \\ V(w_k,\mu) & 0 & U(x_k,\mu) \end{pmatrix}, \tag{3.3}$$

the matrix G_k is $\nabla_x^2 L(w_k)$ or its approximation, and

$$\begin{aligned} v(w_i,\mu) &= z_i u'(x_i,\mu) - \rho h'(x_i,\mu) = \frac{x_i(z_i - \rho/2)}{\sqrt{x_i^2+\mu^2}} + \frac{\rho}{2},\ i = 1,\ldots n, \\ V(w,\mu) &= \operatorname{diag}(v(w_1,\mu),\ldots,v(w_n,\mu)). \end{aligned} \tag{3.4}$$

If $G_k = \nabla_x^2 L(w_k)$, then J_k becomes the Jacobian matrix of $r(w,\mu)$ at w_k. We note that equations (3.2) can be represented by the forms

$$\begin{aligned} &(G_k + U(x_k,\mu)^{-1} V(w_k,\mu))\Delta x_k - A(x_k)^T \Delta y_k \\ &\quad = -\nabla f(x_k) + A(x_k)^T y_k - \rho H'(x_k,\mu)e, \end{aligned} \tag{3.5}$$

$$A(x_k)\Delta x_k + \mu \Delta y_k = -g(x_k) - \mu y_k \tag{3.6}$$

for Δx_k and Δy_k, and

$$\Delta z_k = -z_k - \rho H'(x_k,\mu)e - U(x_k,\mu)^{-1} V(w_k,\mu)\Delta x_k. \tag{3.7}$$

It follows from (2.14), (3.5), (3.6) and (3.7) that

$$\left(G_k + U(x_k,\mu)^{-1} V(w_k,\mu) + \frac{1}{\mu} A(x_k)^T A(x_k)\right)\Delta x_k = -\nabla F_0(x_k,\mu). \tag{3.8}$$

The next lemma gives basic properties of finite termination.

Lemma 3.2. *Suppose that Δw_k satisfies* (3.2) *at a point w_k.*

(i) *If* $\Delta w_k = 0$*, then the point* w_k *is an SKKT point that satisfies* (2.15)*.*

(ii) *If* $\Delta x_k = 0$*, then the point* $(x_k, y_k + \Delta y_k, z_k + \Delta z_k)^T$ *is an SKKT point that satisfies* (2.15)*.*

Proof. (i) The result follows directly from (3.2).
(ii) If $\Delta x_k = 0$, from (3.5), (3.6) and (3.7), we have

$$\begin{aligned} \nabla f(x_k) - A(x_k)^T(y_k + \Delta y_k) + \rho H'(x_k,\mu)e &= 0, \\ g(x_k) + \mu(y_k + \Delta y_k) &= 0, \\ z_k + \Delta z_k + \rho H'(x_k,\mu)e &= 0. \end{aligned}$$

This can be represented by

$$\begin{pmatrix} \nabla f(x_k) - A(x_k)^T(y_k + \Delta y_k) - (z_k + \Delta z_k) \\ g(x_k) + \mu(y_k + \Delta y_k) \\ U(x_k,\mu)(z_k + \Delta z_k) - \rho H(x_k,\mu)e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

so the point $(x_k, y_k + \Delta y_k, z_k + \Delta z_k)^T$ is an SKKT point from (2.15). ☐

The following lemma gives a sufficient condition for equation (3.2) to be solvable.

Lemma 3.3. *If the matrix* $G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k)$ *is positive definite, then the matrix* J_k *in* (3.3) *is nonsingular.*

Proof. Consider the equations

$$G_k v_x - A(x_k)^T v_y - v_z = 0,\ A(x_k)v_x + \mu v_y = 0 \quad \text{and} \quad V(w_k,\mu)v_x + U(x_k,\mu)v_z = 0$$

for $(v_x, v_y, v_z)^T \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$. By the second and third equations, we have

$$v_y = -\frac{1}{\mu}A(x_k)v_x \quad \text{and} \quad v_z = -U(x_k,\mu)^{-1}V(w_k,\mu)v_x.$$

Substituting v_y and v_z into the first equation, we have

$$\left(G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k)\right)v_x = 0.$$

By the assumption, we obtain $v_x = 0$, and therefore $v_y = 0$ and $v_z = 0$. This proves the lemma. ☐

We have presented a differentiable penalty function (2.13) instead of (2.11). Though the function (2.13) may be used as a merit function, $F_0(x,\mu)$ depends only on primal variables. Since we have the Newton direction in the primal-dual space, it is significant to consider a merit function in the primal-dual space. In order to obtain a merit function whose stationary point becomes an SKKT point, we propose the following function

$$F(p,\mu) = F_0(x,\mu) + \frac{\sigma}{2}\left\|g(x) + \mu y\right\|^2, \tag{3.9}$$

where $p = (x,y)^T \in \mathbf{R}^n \times \mathbf{R}^m$, and σ is a positive constant. We call this the shifted penalty function. We note that the second term in (3.9) corresponds to the second equation in (2.15). Similarly to the merit functions proposed by Forsgren and Gill [5], Vanderbei and

Shanno [10], and Yamashita and Yabe [15] for interior-point methods, it may be possible to consider the additional term $\|U(x,\mu)z-\rho H(x,\mu)e\|^2$ that corresponds to the third equation in (2.15). However, to deal with the box constraints $0 \leq z \leq \rho e$ separately, we propose the merit function above. For notational convenience, we define

$$\phi(p) \equiv \frac{1}{2}\left\|g(x)+\mu y\right\|^2. \tag{3.10}$$

Then (3.9) becomes $F(p,\mu) = F_0(x,\mu) + \sigma\phi(p)$. Now we calculate the derivatives of the merit function:

$$\nabla F(p,\mu) = \begin{pmatrix} \nabla F_0(x,\mu) + \sigma\nabla_x\phi(p) \\ \sigma\nabla_y\phi(p) \end{pmatrix}, \tag{3.11}$$

where

$$\nabla_x\phi(p) = A(x)^T(g(x)+\mu y) \quad \text{and} \quad \nabla_y\phi(p) = \mu(g(x)+\mu y).$$

Let $\Delta p = (\Delta x, \Delta y)^T \in \mathbf{R}^n \times \mathbf{R}^m$. The following lemma evaluates an upper bound on the directional derivative of F along Δw_k at w_k.

Lemma 3.4. *If Δw_k solves* (3.2), *then*

$$\begin{aligned} \nabla F(p_k,\mu)^T\Delta p_k \quad \leq \quad & -\Delta x_k^T(G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k))\Delta x_k \\ & -\sigma\left\|g(x_k)+\mu y_k\right\|^2. \end{aligned} \tag{3.12}$$

Furthermore, if $\Delta x_k \neq 0$ and $G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k)$ is positive definite, then $\nabla F(p_k,\mu)^T\Delta p_k < 0$ holds.

Proof. From (3.6) and (3.10), we have

$$\begin{aligned} \nabla\phi(p_k)^T\Delta p_k &= (A(x_k)\Delta x_k + \mu\Delta y_k)^T(g(x_k)+\mu y_k) \\ &= -\left\|g(x_k)+\mu y_k\right\|^2 \leq 0. \end{aligned} \tag{3.13}$$

Since $\nabla F(p_k,\mu)^T\Delta p_k = \nabla F_0(x_k,\mu)^T\Delta x_k + \sigma\nabla\phi(p_k)^T\Delta p_k$ by (3.11), it follows from (3.8) and (3.13) that

$$\begin{aligned} \nabla F(p_k,\mu)^T\Delta p_k \quad = \quad & -\Delta x_k^T(G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k))\Delta x_k \\ & -\sigma\left\|g(x_k)+\mu y_k\right\|^2, \end{aligned}$$

which proves the lemma. □

To obtain a global convergence to an SKKT point for fixed μ, we adopt Armijo's rule as the line search for the variables x and y. For each component of variable z, we adopt the box constraints rule such that

$$\alpha_{z_i k} = \max_{\alpha}\left\{\alpha | 0 \leq (z_k)_i + \alpha(\Delta z_k)_i \leq \rho,\ 0 \leq \alpha \leq 1\right\}, i = 1,\ldots,n. \tag{3.14}$$

This rule is the essential feature of exterior point method and means that each step size $\alpha_{z_i k}$ is the maximal step that satisfies the box constraints.

Now we present the algorithm of our method for finding an approximate SKKT point.

Algorithm LS

Step 0. (Initialize) Let $w_0 \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n_\rho$, and $\mu > 0$, $\rho > 0$, $\sigma > 0$. 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 (3.2).

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

$$F(p_k + \beta^{l_k} \Delta p_k, \mu) - F(p_k, \mu) \le \varepsilon_0 \beta^{l_k} \nabla F(p_k, \mu)^T \Delta p_k. \tag{3.15}$$

Calculate $\alpha_k = \beta^{l_k}$, $\alpha_{z_i k}$ in (3.14) for $i = 1, \ldots, n$, and $\Lambda_k = \mathrm{diag}(\alpha_k I_n, \alpha_k I_m, \alpha_{z_1 k}, \ldots, \alpha_{z_n k})$.

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. □

3.3 Global convergence of Algorithm LS

In this subsection, we prove the global convergence of Algorithm LS. If there exists an index k such that $\Delta x_k = 0$, the point $(x_k, y_k + \Delta y_k, z_k + \Delta z_k)^T$ is an SKKT point that satisfies (2.15) by Lemma 3.2. Hence, we assume that $\Delta x_k \ne 0$ for each $k = 0, 1, \ldots$. To prove the global convergence of Algorithm LS, we make the following assumptions.

Assumption G

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

(G2) The sequence $\{x_k\}$ is bounded.

(G3) The matrix G_k is uniformly bounded and the matrix $G_k + U(x_k, \mu)^{-1} V(w_k, \mu) + \frac{1}{\mu} A(x_k)^T A(x_k)$ is uniformly positive definite. □

The compactness of the generated sequence is derived if we assume the compactness of the level set of the function $F(p, \mu)$ at the initial point, for example, because the iterates give decreasing function values. We note that if a quasi-Newton approximation is used for computing the matrix G_k, then Assumption (G1) can be replaced by the assumption of the continuous differentiability. Assumption G is more relaxed than Assumption GLS of Yamashita and Tanabe [14].

In order to generate a decent search direction, we need to have a positive definite $V(w, \mu)$. The next lemma shows a condition for this property to hold (see Lemma 2 of Yamashita and Tanabe [14].).

Lemma 3.5. *If $\mu \ne 0$ and $0 \le z_i \le \rho$, then $v(w_i, \mu) \in (0, \rho)$ for $i = i, \ldots, n$, i.e., the diagonal matrix $V(w, \mu)$ is positive definite.*

Lemma 3.6. *Let an infinite sequence $\{w_k\}$ be generated by Algorithm LS with $\mu > 0$. If $\{x_k\}$ is bounded, then the sequence $\left\{U(x_k, \mu)^{-1} V(w_k, \mu)\right\}$ is uniformly positive definite and bounded.*

Proof. It follows from (2.10) and (3.4) that the ith diagonal component of $U(x_k,\mu)^{-1}V(w_k,\mu)$ is given by

$$\frac{1}{\sqrt{(x_k)_i^2+\mu^2}}\left(\frac{(x_k)_i((z_k)_i-\frac{\rho}{2})}{\sqrt{(x_k)_i^2+\mu^2}}+\frac{\rho}{2}\right)=\frac{2(x_k)_i((z_k)_i-\frac{\rho}{2})+\rho\sqrt{(x_k)_i^2+\mu^2}}{2((x_k)_i^2+\mu^2)}. \tag{3.16}$$

Since there exists a positive constant ξ such that $|(x_k)_i|\le\xi$ and $0\le(z_k)_i\le\rho$ holds, we have

$$\begin{aligned}\left|\frac{1}{\sqrt{(x_k)_i^2+\mu^2}}\left(\frac{(x_k)_i((z_k)_i-\frac{\rho}{2})}{\sqrt{(x_k)_i^2+\mu^2}}+\frac{\rho}{2}\right)\right| &\le \frac{2|(x_k)_i|(|(z_k)_i|+\frac{\rho}{2})+\rho\sqrt{(x_k)_i^2+\mu^2}}{2\mu^2}\\ &\le \frac{3\xi\rho+\rho\sqrt{\xi^2+\mu^2}}{2\mu^2}\end{aligned}$$

for $i=1,\dots,n$, which implies the boundedness of the sequence $\{U(x_k,\mu)^{-1}V(w_k,\mu)\}$. Furthermore, equation (3.16) yields

$$\begin{aligned}\frac{1}{\sqrt{(x_k)_i^2+\mu^2}}\left(\frac{(x_k)_i((z_k)_i-\frac{\rho}{2})}{\sqrt{(x_k)_i^2+\mu^2}}+\frac{\rho}{2}\right) &\ge \frac{-2|(x_k)_i||((z_k)_i-\frac{\rho}{2})|+\rho\sqrt{(x_k)_i^2+\mu^2}}{2(\xi^2+\mu^2)}\\ &\ge \frac{-\rho|(x_k)_i|+\rho\sqrt{(x_k)_i^2+\mu^2}}{2(\xi^2+\mu^2)}\\ &\ge \frac{\rho\mu^2}{2(\xi^2+\mu^2)(\sqrt{\xi^2+\mu^2}+\xi)},\end{aligned}$$

where the second inequality follows from $|(z_k)_i-\frac{\rho}{2}|\le\frac{\rho}{2}$ because of the condition $0\le(z_k)_i\le\rho$. This implies the uniformly positive definiteness of the sequence $\{U(x_k,\mu)^{-1}V(w_k,\mu)\}$.

Therefore, this lemma is proved. □

Now we obtain the global convergence theorem.

Theorem 3.7. *Suppose that Assumption G holds. 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 SKKT point.*

Proof. First, we prove the boundedness of the sequence $\{w_k\}$. From (3.9) and Armijo rule, we obtain

$$F(p_0,\mu)\ge F(p_k,\mu)=F_0(x_k,\mu)+\frac{\sigma}{2}\|g(x_k)+\mu y_k\|^2\ge F_0(x_k,\mu).$$

Moreover, by Assumptions (G1), (G2) and (2.6), the sequences $\{F_0(x_k,\mu)\}$ and $\{F(p_k,\mu)\}$ are bounded. Therefore, the sequence $\{y_k\}$ is bounded. From the above, Assumption (G2) and the boundedness of sequences $\{y_k\}$ and $\{z_k\}$, the sequence $\{w_k\}$ is bounded, and thus has at least one accumulation point.

By using this boundedness property, Assumption (G3) and Lemma 3.6, there exists a positive number M such that

$$\frac{\|v\|^2}{M}\le v^T(G_k+U(x_k,\mu)^{-1}V(w_k,\mu)+\frac{1}{\mu}A(x_k)^TA(x_k))v\le M\|v\|^2,\ \forall v\in\mathbf{R}^n, \tag{3.17}$$

for all k. From (3.12) and (3.17), we have

$$\nabla F(p_k,\mu)^T\Delta p_k \quad \leq \quad -\frac{\|\Delta x_k\|^2}{M} - \sigma\,\|g(x_k)+\mu y_k\|^2 < 0, \tag{3.18}$$

and from (3.15),

$$\begin{aligned} F(p_{k+1},\mu) - F(p_k,\mu) \quad &\leq \quad \varepsilon_0\beta^{l_k}\nabla F(p_k,\mu)^T\Delta p_k \qquad\qquad (3.19)\\ &\leq \quad -\varepsilon_0\beta^{l_k}\left(\frac{\|\Delta x_k\|^2}{M} + \sigma\,\|g(x_k)+\mu y_k\|^2\right) < 0. \end{aligned}$$

Since the sequence $\{F(p_k,\mu)\}$ is decreasing and bounded below, the left-hand side of (3.19) converges to 0. Therefore, $\beta^{l_k}\left(\frac{\|\Delta x_k\|^2}{M} + \sigma\,\|g(x_k)+\mu y_k\|^2\right)$ also converges to 0. Then, we can consider the following two cases:

(i) If there exists a number $N > 0$ such that $l_k < N$ for all k in a subsequence $K_1 \subset \{0,1,\cdots\}$, then $\Delta x_k \to 0$ and $g(x_k)+\mu y_k \to 0$ in this subsequence from (3.19).

(ii) If there exists a subsequence $K_2 \subset \{0,1,\cdots\}$ such that $l_k \to \infty, k \in K_2$, then we can assume $l_k > 0$ for sufficiently large $k \in K_2$ without loss of generality. Since the point $p_k + \alpha_k\Delta p_k/\beta$ does not satisfy condition (3.15), we have

$$F(p_k+\alpha_k\Delta p_k/\beta,\mu) - F(p_k,\mu) > \varepsilon_0\alpha_k\nabla F(p_k,\mu)^T\Delta p_k/\beta. \tag{3.20}$$

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

$$F(p_k+\alpha_k\Delta p_k/\beta,\mu) - F(p_k,\mu) = \alpha_k\nabla F(p_k+\theta_k\alpha_k\Delta p_k/\beta,\mu)^T\Delta p_k/\beta. \tag{3.21}$$

Then, from (3.20) and (3.21), we have

$$\nabla F(p_k+\theta_k\alpha_k\Delta p_k/\beta,\mu)^T\Delta p_k > \varepsilon_0\nabla F(p_k,\mu)^T\Delta p_k.$$

This inequality yields

$$\begin{aligned} &\nabla F(p_k+\theta_k\alpha_k\Delta p_k/\beta,\mu)^T\Delta p_k - \nabla F(p_k,\mu)^T\Delta p_k \qquad (3.22)\\ &\quad > (\varepsilon_0-1)\nabla F(p_k,\mu)^T\Delta p_k > 0. \end{aligned}$$

Now, since the inverse of the coefficient matrix of (3.8) is uniformly bounded by (3.17), $\|\Delta x_k\|$ is uniformly bounded. Then it follows from (3.6) that $\|\Delta y_k\|$ is also uniformly bounded, so $\|\Delta p_k\|$ is uniformly bounded. Thus by the boundedness of $\|\Delta p_k\|$ and property $l_k \to \infty, k \in K_2$, we have $\|\theta_k\alpha_k\Delta p_k/\beta\| \to 0, k \in K_2$. Hence the left-hand side of (3.22) and therefore $\nabla F(p_k,\mu)^T\Delta p_k$ converges to zero when $k \to \infty, k \in K_2$. Inequality (3.18) yields $\Delta x_k \to 0$, $g(x_k)+\mu y_k \to 0$, $k \in K_2$.

From (i) and (ii), we have proved $\Delta x_k \to 0$ and $g(x_k)+\mu y_k \to 0$. Let an arbitrary accumulation point of the sequence $\{p_k\}$ be $\hat{p} = (\hat{x},\hat{y}) \in \mathbf{R}^n \times \mathbf{R}^m$, and let $x_k \to \hat{x}, y_k \to \hat{y}, k \in K$ for a subsequence $K \subset \{0,1,\cdots\}$. Thus we have

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

By $\Delta x_k \to 0$, $g(x_k)+\mu y_k \to 0$ and (3.6), $\Delta y_k \to 0$ holds and the sequence $\{y_k+\alpha_k\Delta y_k\}$, $k \in K$ converges to a point $\hat{y} \in \mathbf{R}^m$ which satisfies $g(\hat{x})+\mu\hat{y} = 0$.

On the other hand, since $\{U(x_k,\mu)^{-1}V(w_k,\mu)\}$ is bounded by Lemma 3.6, we have

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

from (3.7), which implies that $z_k + \Delta z_k \to \rho H'(\hat{x},\mu)e, k \in K$. If we define $\hat{z} = -\rho H'(\hat{x},\mu)e$, then $0 < \hat{z} < \rho e$ because of $-1 < h'(\hat{x},\mu) < 0$ and

$$U(\hat{x},\mu)\hat{z} = \rho H(\hat{x},\mu)e,\ z_k + \Delta z_k \to \hat{z},\ k \in K.$$

This means that the point $z_k + \Delta z_k$ is always accepted as z_{k+1} for sufficiently large $k \in K$. Therefore, it follows from (3.5) that there exists at least one accumulation point of $\{w_k\}$ that satisfies

$$\nabla_x L(\hat{x},\hat{y},\hat{z}) = 0, \quad g(\hat{x}) + \mu\hat{y} = 0,$$
$$U(\hat{x},\mu)\hat{z} = \rho H(\hat{x},\mu)e, \quad 0 < \hat{z} < \rho e$$

and for an arbitrary accumulation point $\hat{x}$ of $\{x_k\}$, there exist $\hat{y}$ and $\hat{z}$ that satisfy the SKKT condition (2.15). □

4 Preliminary Numerical Experiment

In this section, we report numerical experiments of an implementation of the algorithm given in this paper. Following Yamashita and Tanabe [14], the parameter values are $\varepsilon = 10^{-6}, \beta = 0.5, \sigma = 100, \varepsilon_0 = 10^{-6}, M_c = 7.5$. The initial value of $\mu_0 = 1.0$ and the update rule of this parameter is given by the following to obtain fast convergence in the final stage of iterations:

If $\|r_0(w_k)\| \geq 10^{-2}$, we update μ_k by $\mu_k = \max\left\{\frac{\|r_0(w_k)\|}{M_\mu}, \frac{\mu_{k-1}}{10}\right\}, M_\mu = 10$.

If $\|r_0(w_k)\| < 10^{-2}$, we update μ_k by $\mu_k = \max\left\{\|r_0(w_k)\|^{1.6}, \frac{\mu_{k-1}}{100}\right\}$.

The matrix G_k in (3.3) is $\nabla_x^2 L(w_k)$. We utilize Algorithm IC given in [11] in order to ensure the matrix $G_k + U(x_k,\mu)^{-1}V(w_k,\mu) + \frac{1}{\mu}A(x_k)^T A(x_k)$ is positive definite. The test problems are chosen from Hock and Schittkowski test set [7]. First, we fixed $\rho = 10$ and this result is shown in Table 1. We note that `n` means the number of variables, `m` means the number of constraints, `objective` means the final objective function value, `residual` means the final value of $\|r_0(w)\|$, `itr` means the number of total iterations of Algorithm EP, and `ext` means the number of iterations during which to compute an exterior point (violation of a bound larger than 10^{-8}). The mark `LOC` shows that a local optimum was obtained, the mark `EXT` shows that the obtained point was an exterior point and the mark `MAX` shows that the algorithm stopped because of the iteration limits. From these experiments, we see that the algorithm given in this paper achieved an exterior point occasionally and this phenomenon was caused by the value of ρ.

`Table 1:` *Hock and Schittkowski test set results.*

problem	n	m	objective	residual	itr	ext	
HS001	2	1	5.8946259e-16	6.5e-07	40	24	
HS002	2	1	4.9412293	5.5e-07	9	3	
HS003	2	1	7.4940964e-09	1.5e-08	8	0	
HS004	2	1	2.6666666	5.2e-07	6	2	
HS005	2	2	-1.9132230	2.1e-09	5	0	
HS006	2	2	0	3.6e-07	10	2	

problem	n	m	objective	residual	itr	ext	
HS007	2	2	1.7320508	5.1e-08	7	0	
HS008	2	3	-1	1.1e-08	6	1	
HS009	2	2	-0.4999998	4.9e-07	6	0	
HS010	2	2	-0.999999996	6.7e-08	9	2	
HS011	2	2	-8.498961	6.6e-07	7	4	
HS012	2	2	-30	3.8e-08	10	1	
HS013	2	2	0.83186486	1.7e-03	100	0	MAX
HS014	2	3	1.3995	1.7e-07	5	0	
HS015	2	3	306.5	1.2e-07	9	2	
HS016	2	4	0.24999994	3.6e-07	15	6	
HS017	2	4	1	5.1e-07	13	2	
HS018	2	4	5	1.1e-08	14	9	
HS019	2	4	-6961.814	1.1e-07	11	2	
HS020	2	4	3.61606679	9.6e-08	7	7	EXT
HS021	2	3	-99.96	5.4e-07	6	1	
HS022	2	2	0.999999992	1.8e-07	6	0	
HS023	2	5	-1.99999998	1.5e-08	15	14	
HS024	2	3	-1	8.0e-08	8	0	
HS025	3	3	0	1.7e-08	49	32	
HS026	5	2	0	9.8e-07	4	0	
HS027	3	2	0.04001437	1.0e-00	100	0	MAX
HS028	3	1	0.1	8.2e-12	10	0	
HS029	3	2	0	1.6e-10	31	0	LOC
HS030	3	1	1.00000021	1.0e-08	7	2	
HS031	3	1	5.99999921	7.5e-07	15	14	
HS032	3	2	0.99999997	2.6e-08	7	0	
HS033	4	3	0.38490018	1.9e-08	8	7	
HS034	3	2	-0.82246661	4.5e-07	6	0	
HS035	3	1	0.111111111	1.4e-10	7	0	
HS036	3	4	-3300	3.6e-07	10	5	
HS037	3	3	-3456	5.1e-08	9	0	
HS038	4	1	0	4.2e-07	10	8	
HS039	4	2	-0.99999996	1.1e-08	11	0	
HS040	4	3	-0.24489792	9.3e-07	5	0	
HS041	4	3	1.9259312	6.6e-07	8	2	
HS042	4	2	13.85786409	6.0e-07	6	0	
HS043	4	3	-40.96328666	4.2e-07	10	0	
HS044	4	6	-15.000436	2.4e-07	23	0	
HS045	5	1	1.00001	9.2e-07	17	4	
HS046	5	1	0	1.0e-08	11	0	
HS047	5	3	0	3.1e-07	14	0	
HS048	5	2	0	5.9e-07	3	0	
HS049	5	2	0	1.4e-07	4	0	
HS050	5	3	0	5.0e-09	2	0	
HS051	5	3	0	5.8e-33	1	0	
HS052	5	3	5.32666	1.5e-08	5	0	
HS053	5	4	4.093019	5.7e-08	7	0	
HS054	6	1	-0.903488	1.2e-06	36	12	
HS055	6	6	6.33333	5.5e-09	7	1	
HS056	7	4	-3.456253	1.3e-08	6	0	
HS057	2	2	0.0306463	1.2e-07	13	3	
HS059	2	4	-7.802792	7.1e-08	20	11	
HS060	3	2	0.03256820	2.7e-07	4	0	
HS061	3	2	-143.081	6.9e-07	9	0	
HS062	3	2	-26272.515	1.3e-07	16	10	
HS063	3	3	961.7152	6.9e-08	13	12	
HS064	3	2	6299.851	2.7e-07	30	18	
HS065	3	2	0.95352928	5.8e-07	23	11	
HS066	3	3	0.5181633	9.0e-08	5	0	
HS067	3	15	-1001.125	3.4e-03	100	0	MAX

problem	n	m	objective	residual	itr	ext	
HS068	4	3	-0.9204251	5.5e-08	20	12	
HS069	4	3	-956.7239	4.3e-07	10	2	
HS070	4	2	0.0074985	1.0e-07	25	10	
HS071	4	3	0.63794166	2.6e-07	5	5	EXT
HS072	4	3	1.00316637	1.8e-03	100	100	MAX
HS073	4	4	29.89437	2.0e-08	10	7	
HS074	4	6	5126.4982	6.9e-08	13	7	
HS075	4	6	5174.413	1.4e-08	13	7	
HS076	4	4	-4.68181805	1.2e-07	6	0	
HS077	5	2	0.43810596	9.4e-07	5	0	LOC
HS078	5	3	-2.9197001	3.0e-07	5	0	
HS079	5	3	0.07880816	4.1e-07	4	0	
HS080	5	4	0.053949848	8.7e-09	6	3	
HS081	5	4	0.053949848	6.0e-08	7	3	
HS083	5	4	-30665.5387	5.8e-08	13	8	
HS084	5	7	-5280335.2	1.5e-07	23	0	
HS085	5	39	0.9535475	7.1e-08	31	0	LOC
HS086	5	11	-32.348679	8,6e-08	10	0	
HS087	6	5	8927.598	6.9e-07	22	4	
HS088	2	1	1.3626462	3.8e-07	19	6	
HS089	3	1	1.3626462	1.5e-08	21	6	
HS090	4	1	1.3626462	1.9e-07	15	7	
HS091	5	1	1.3626462	3.5e-07	15	8	
HS092	6	1	1.3626462	2.7e-07	20	9	
HS093	6	3	135.07596	4.4e-08	18	0	
HS095	6	5	0.0156195	7.0e-07	9	2	
HS096	6	5	0.0156194	8.1e-07	10	3	
HS097	6	5	3.1358092	5.1e-08	14	1	
HS098	6	5	3.1358091	2.8e-07	12	2	
HS099	7	3	-8.3107989e+08	3.1e-08	8	0	
HS100	7	4	680.6301	7.5e-08	11	0	
HS101	7	6	1809.7648	1.1e-07	20	16	
HS102	7	6	911.880571	2.7e-08	23	5	
HS103	7	6	543.667961	1.9e-07	27	20	
HS104	8	6	3.9511634	1.3e-08	22	4	
HS105	8	2	1136.3610	8.1e-07	25	20	
HS106	8	7	7049.33092	5.6e-07	40	25	
HS107	9	7	5055.0118	1.3e-09	21	9	
HS108	9	14	-0.67498143	2.1e-10	65	47	
HS109	9	11	5362.0693	2.5e-08	16	8	
HS110	10	1	-45.77847	3.9e-08	8	0	
HS111	10	4	-47.701091	2.0e-08	20	7	
HS112	10	4	-47.707569	7.9e-08	15	13	
HS113	10	8	24.30621	1.4e-09	27	0	
HS114	10	12	-0.89364	7.5e-08	50	38	EXT
HS116	15	15	-2.9197	2.6e-07	85	11	
HS117	15	6	0.0787768	4.4e-07	41	8	
HS118	15	18	0.0539498	1.4e-08	37	20	
HS119	16	9	1.14716	5.6e-07	30	22	EXT

Next, we investigate influence of penalty parameter ρ. We chose the values $\rho = 10, 100, 1000$ and 10000, and applied our method to five problems HS114, HS116, HS117, HS118 and HS119. These results are given in Table 2. In this table, `OPT` shows that an optimum was obtained and `EXT` shows that the obtained point was an exterior point. From the table, it can be observed that our algorithm obtained exterior points with small values of ρ. However, as the value of ρ increases, we can obtain optimal solutions. Therefore, our algorithm performs well for nonlinear optimization problems by choosing a suitable penalty parameter.

Table 2: *Results for several values of ρ.*

problem	$\rho = 10$			$\rho = 100$		
	status	itr(ext)	residual	status	itr(ext)	residual
HS114	EXT	50(38)	7.5e-08	EXT	53(39)	4.2e-08
HS116	OPT	85(11)	2.6e-07	OPT	80(10)	3.0e-07
HS117	OPT	41(8)	4.4e-07	OPT	43(8)	5.1e-07
HS118	OPT	37(20)	1.4e-08	OPT	37(21)	3.1e-08
HS119	EXT	30(22)	5.6e-07	OPT	32(18)	5.9e-07
problem	$\rho = 1000$			$\rho = 10000$		
	status	itr(ext)	residual	status	itr(ext)	residual
HS114	OPT	55(23)	1.2e-08	OPT	59(26)	3.2e-08
HS116	OPT	86(15)	2.2e-07	OPT	84(17)	9.1e-07
HS117	OPT	48(11)	7.2e-07	OPT	49(9)	1.8e-07
HS118	OPT	33(21)	8.3e-09	OPT	38(25)	5.5e-08
HS119	OPT	30(13)	6.0e-07	OPT	33(14)	4.8e-07

5 Conclusions

In this paper, we have proposed a new primal-dual differentiable merit function and proved the global convergence property of our method within the line search framework. We have investigated numerical performance of our algorithm by using preliminary numerical experiments. Additional numerical experiments for large scale problems are our further work.

Acknowledgments. The authors would like to thank two anonymous referees for their valuable comments.

References

[1] K.S. Al-Sultan and K.G. Murty, Exterior point algorithms for nearest points and convex quadratic programs, *Math. Program.* 57 (1992) 145-161.

[2] J.F. Andrus, An exterior point method for the convex programming problem, *J. Optim. Theory Appl.* 72 (1992) 37-63.

[3] J.F. Andrus and M.R. Schaferkotter, An exterior-point method for linear programming problems, *J. Optim. Theory Appl.* 91 (1996) 561-583.

[4] R. Fletcher, *Practical Methods of Optimization (Second Edition)*, John Wiley & Sons, Chichester, 1987.

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

[6] I. Griva, Numerical experiments with an interior-exterior point method for nonlinear programming, *Comput. Optim. Appl.* 29 (2004) 173-195.

[7] W. Hock and K. Schittkowski, Test Examples for Nonlinear Programming Codes, *Lecture Notes in Economics and Mathematical Systems 187,* Springer, Berlin, 1981.

[8] J. Nocedal and S.J. Wright, *Numerical Optimization (Second Edition)*, Springer, New York, 2006.

[9] R.A. Polyak, Primal-dual exterior point method for convex optimization, *Optim. Methods Softw.* 23 (2008) 141-160.

[10] R.J. Vanderbei and D.F. Shanno, An interior-point algorithm for nonconvex nonlinear programming, *Comput. Optim. Appl.* 13 (1999) 231-252.

[11] A. Wächter and L.T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Math. Program.* 106 (2006) 25-57.

[12] S.J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

[13] H. Yamashita, A globally convergent primal-dual interior point method for constrained optimization, *Optim. Methods Softw.* 10 (1998) 443-469.

[14] H. Yamashita and T. Tanabe, A primal-dual exterior point method for nonlinear optimization, *SIAM J. Optim.* 20 (2010) 3335-3363.

[15] H. Yamashita and H. Yabe, An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization, *SIAM J. Optim.* 14 (2003) 479-499.

[16] X.Q. Yang, An exterior point method for computing points that satisfy second-order necessary conditions for a $C^{l,1}$ optimization problem, *J. Math. Anal. Appl.* 187 (1994) 118-133.

Manuscript received 30 July 2014
revised 24 March 2015
accepted for publication 15 May 2015

Yu Igarashi
Mizuho Information & Research Institute, Inc.
2-3, Kandanishikicho, Chiyoda-ku, Tokyo, 101-0054, Japan

Hiroshi Yabe
Department of Mathematical Information Science
Faculty of Science, Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601, Japan
E-mail address: `yabe@rs.kagu.tus.ac.jp`