Journal of the Operations Research
Society of Japan
Vol. 40, No. 3, September 1997

Q-SUPERLINEAR CONVERGENCE OF PRIMAL-DUAL INTERIOR POINT QUASI-NEWTON METHODS FOR CONSTRAINED OPTIMIZATION

Hiroshi Yabe
Science University of Tokyo

Hiroshi Yamashita
Mathematical Systems, Inc.

(Received July 9, 1996; Revised March 21, 1997)

Abstract This paper analyzes local convergence rates of primal-dual interior point methods for general nonlinearly constrained optimization problems. For this purpose, we first discuss modified Newton methods and modified quasi-Newton methods for solving a nonlinear system of equations, and show local and Q-quadratic/Q-superlinear convergence of these methods. These methods are characterized by a perturbation of the right-hand side of the Newton equation applied to the system, an approximation of the Jacobian matrix by some matrix, and component-wise dampings of the step. By applying these convergence results for the nonlinear system of equations to the primal-dual interior point methods for nonlinear optimization, we obtain convergence results of the primal-dual interior point Newton and quasi-Newton methods. A necessary and sufficient condition for Q-superlinear convergence of the latter methods corresponds to the Dennis-Moré condition. Furthermore, we present some quasi-Newton updating formulae. Finally, we give an analysis of the Q-rate in a part of variables for the primal-dual interior point quasi-Newton methods, and obtain a necessary and sufficient condition for the Q-rate. This condition is a generalization of the result given by Martinez, Parada and Tapia (1995), which was done independently.

1. Introduction

This paper is concerned with primal-dual interior point methods for solving the nonlinearly constrained optimization problem:

$$\text{(1.1)} \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad g(x)=0,\ x\geq 0,\ x\in R^n,$$

where $f: R^n \rightarrow R$ and $g: R^n \rightarrow R^m$. Many numerical methods have been studied for solving the problem. Among them, the augmented Lagrangian method and the SQP method (see for example [11]) have been regarded as representatives of general and effective methods. On the other hand, the excellent success of interior point methods for linear programming, especially the primal-dual method ([15], [17], [20]), has affected researches on numerical methods for nonlinear optimization and has aroused renewed interests between researchers about interior point methods applied to the problem.

Recently, primal-dual interior point methods for general nonlinearly constrained problems have been studied by several authors. Local convergence properties are discussed by McCormick and Falk [19], El-Bakry, Tapia, Tsuchiya and Zhang [8], Yamashita and Yabe [26], and Martinez, Parada and Tapia [18]. In [26], we studied primal-dual interior point methods based on the Newton method and the quasi-Newton method, and proved local

A preliminary version of this paper was presented at the conference "The State of the Art of Scientific Computing and its Prospect" held at Research Institute for Mathematical Sciences, Kyoto University, Kyoto, October 1993. ([24])

and quadratic/superlinear convergence of these methods. On the other hand, the global convergence property was studied by Yamashita [25]. Furthermore, Yamashita [25], Vial [23], and Lasdon, Plummer and Yu [16] presented some computational experience. In [25], Yamashita applied his globally convergent primal-dual interior point method to Hock and Schittkowski problems [14] and showed efficiency of the method.

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

$$L(x,y,z) = f(x) - y^t g(x) - z^t x, \tag{1.2}$$

where y and z are the multiplier vectors corresponding to the constraints $g(x) = 0$ and $x \geq 0$, respectively. In our previous paper [26], we dealt with the following modified Karush-Kuhn-Tucker (K-K-T) conditions:

$$r(x,y,z) \equiv \begin{pmatrix} r_L(x,y,z) \\ r_E(x) \\ r_C(x,z) \end{pmatrix} \equiv \begin{pmatrix} \nabla_x L(x,y,z) \\ g(x) \\ XZe - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{1.3}$$

and

$$x > 0, \qquad z > 0 \tag{1.4}$$

for a non-negative number μ, instead of the K-K-T conditions of problem (1.1):

$$r_0(x,y,z) \equiv \begin{pmatrix} \nabla_x L(x,y,z) \\ g(x) \\ XZe \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{1.5}$$

and

$$x \geq 0, \qquad z \geq 0, \tag{1.6}$$

where

$$\begin{aligned} X &= \mathrm{diag}(x_1, x_2, \cdots, x_n), \\ Z &= \mathrm{diag}(z_1, z_2, \cdots, z_n), \\ e &= (1, 1, \cdots, 1)^t \in \mathbf{R}^n, \end{aligned}$$

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

and $A(x) \in R^{m\times n}$ is the Jacobian matrix of $g(x)$. Here we note that

$$r(x,y,z) = r_0(x,y,z) - \mu\hat{e}, \tag{1.7}$$

where

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

Denote the Jacobian matrix of $r(x,y,z)$ by

$$\nabla r(x,y,z) = \begin{pmatrix} \nabla_x^2 L(x,y,z) & -A(x)^t & -I \\ A(x) & O & O \\ Z & O & X \end{pmatrix}. \tag{1.8}$$

Note that $\nabla r(x,y,z)$ is equal to the Jacobian matrix $\nabla r_0(x,y,z)$. Denoting (x,y,z) by $w \in R^n \times R^m \times R^n$, we have the following prototype algorithm of our primal-dual interior point method, which was described in [26].

Algorithm I
For $k = 0, 1, 2, \cdots$, do
Step 1. Choose the parameter $\mu_k \geq 0$.
Step 2. Solve the following system for $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^t$

$$J_k \Delta w_k = -r(w_k), \tag{1.9}$$

where

$$J_k = \begin{pmatrix} G_k & -A(x_k)^t & -I \\ A(x_k) & O & O \\ Z_k & O & X_k \end{pmatrix}, \tag{1.10}$$

and G_k is the Hessian matrix $\nabla_x^2 L(w_k)$ of the Lagrangian function or its approximation.
Step 3. Compute the step sizes $\Lambda_k = \mathrm{diag}\,(\alpha_{xk} I_n, \alpha_{yk} I_m, \alpha_{zk} I_n)$, where I_n and I_m are n-th and m-th order identity matrices, respectively.
Step 4. Update:

$$w_{k+1} = w_k + \Lambda_k \Delta w_k. \quad \square \tag{1.11}$$

We note that the iteration defined by (1.9) is the Newton method for the solution of the system $r(w) = 0$ if the matrix G_k is the true Hessian matrix of the Lagrangian function. We also note that if the matrix G_k is an approximation to the Hessian matrix of the Lagrangian function, Algorithm I corresponds to the quasi-Newton method. Algorithm I has three characteristics. They are an approximation of the Hessian $\nabla_x^2 L(w_k)$ by some matrix G_k, a perturbation of the right-hand side of the Newton equation by a vector $\mu\hat{e}$, and componentwise dampings of the step Δw_k. In this paper, we call the Newton method and the quasi-Newton method that possess such characteristics as a *modified Newton method* and a *modified quasi-Newton method*, respectively.

In [26], we gave necessary and sufficient conditions for the point $\{(x_k, y_k, z_k)\}$ to converge Q-superlinearly to the K-K-T point (x^*, y^*, z^*) within the framework of the quasi-Newton method. In general, a Q-rate in (x, y, z) implies no more than the corresponding R-rate in (x, z). Thus it is interesting to analyze the Q-rate in (x, z). In the SQP method or the augmented Lagrangian method for solving constrained optimization, the analysis of the Q-rate in a part of variables is one of the main topics, and many studies have been done by, for example, Boggs, Tolle and Wang [1], Nocedal and Overton [21], Fontecilla [12], and Coleman [3]. They discussed conditions on the approximation to the projected Hessian matrix.

In this paper, we will consider similar conditions on the Q-rate in (x, y, z) and in (x, z) to those obtained in the SQP method and the augmented Lagrangian method. For this purpose, we will in general discuss modified Newton methods and modified quasi-Newton methods for solving a nonlinear system of equations, and will show local and Q-quadratic/Q-superlinear convergence of these methods. These results are in part related with the studies of inexact Newton methods by Dembo, Eisenstat and Steihaug [5] and inexact quasi-Newton methods by Eisenstat and Steihaug [7], because we have to deal with a perturbation of the right-hand side of the Newton equation. By applying these convergence results for the nonlinear system of equations to primal-dual interior point methods, we will obtain the main results in [26] as corollaries and, furthermore, some results for the Q-rate in (x, z) within the framework of primal-dual interior point quasi-Newton methods. The latter result is a generalization of the study by Martinez, Parada and Tapia [18], which was done independently.

The present paper is organized as follows. In Section 2, we will discuss general properties of modified Newton methods and modified quasi-Newton methods for solving a nonlinear system of equations, and show their local convergence and the rate of convergence. In Section 3, by using the results of Section 2, we will derive the convergence results of the primal-dual interior point methods proposed in [26] for solving the nonlinear optimization problem (1.1). In addition, we will construct concrete updating formulae within the framework of primal-dual interior point quasi-Newton methods. Section 4 will be devoted to the analysis of Q-superlinear convergence of the sequence $\{(x_k, z_k)\}$. In the following, for simplicity, we omit the prefix "Q-" of Q-quadratic and Q-superlinear convergence.

2. General theory on modified Newton and modified quasi-Newton methods

In this section, we consider the nonlinear system of equations

$$r_0(w) = 0, \qquad r_0 : R^n \to R^n, \quad w \in R^n. \tag{2.1}$$

Let $w^* \in R^n$ be a solution to the equations and let r_0 be continuously differentiable. When the Newton method is applied to the nonlinear system of equations, we have the following Newton equation:

$$\nabla r_0(w_k)\Delta w_k = -r_0(w_k),$$

and the new iterate is given by

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

where w_k is a current approximation to the solution w^* and ∇r_0 denotes the Jacobian matrix of r_0.

We partition the vectors into

$$w = \begin{pmatrix} u \\ v \end{pmatrix}, \quad r_0 = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \quad u \in R^{n_1}, v \in R^{n_2}, r_1 : R^n \to R^{n_1}, r_2 : R^n \to R^{n_2},$$

where $n_1 + n_2 = n$. Then the Newton equation is written as

$$\begin{pmatrix} \nabla_u r_1(w_k) & \nabla_v r_1(w_k) \\ \nabla_u r_2(w_k) & \nabla_v r_2(w_k) \end{pmatrix} \begin{pmatrix} \Delta u_k \\ \Delta v_k \end{pmatrix} = - \begin{pmatrix} r_1(w_k) \\ r_2(w_k) \end{pmatrix}.$$

Now we consider the following:

(1) Approximation to $\nabla_u r_1(w_k)$ by some matrix G_k.

(2) Perturbation of the right-hand side of the Newton equation by a vector $p_k \in R^n$.

(3) Component-wise dampings of the Newton step Δw_k.

As methods based on the idea of (2), inexact Newton methods are well known. Dembo, Eisenstat and Steihaug [5] first proposed these methods for solving the nonlinear system of equations and analyzed local convergence and the rate of convergence. An application of inexact Newton methods to unconstrained optimization was also derived by Dembo and Steihaug [4]. As methods based on the ideas of (1) and (2), Eisenstat and Steihaug [7] derived inexact quasi-Newton methods for solving the system of nonlinear equations and studied local convergence and the rate of convergence. Also, Steihaug [22] and Fontecilla [12] discussed methods based on the idears of (1) and (2) for solving unconstrained optimization and constrained optimization, respectively.

Standard methods use a single step size for the Newton direction and analyze the rate of convergence for the case of a unit step size. However, since we would like to deal with the primal-dual interior point methods, we should consider component-wise dampings in the idea of (3) according to the current practice of implementations, and furthermore, we give sufficient conditions for a rapid convergence.

By taking the preceding ideas into account, we consider the following algorithm.

Algorithm X
For $k = 0, 1, 2, \cdots$, do
Step 1. Solve the following system for $\Delta w_k = (\Delta u_k, \Delta v_k)^t$

$$J_k \Delta w_k = -r_0(w_k) + p_k, \tag{2.2}$$

where

$$J_k = \begin{pmatrix} G_k & \nabla_v r_1(w_k) \\ \nabla_u r_2(w_k) & \nabla_v r_2(w_k) \end{pmatrix}. \tag{2.3}$$

Step 2. Compute the step size

$$\Lambda_k = \mathrm{diag}(\alpha_{1k}, \alpha_{2k}, \cdots, \alpha_{nk}).$$

Step 3. Update

$$w_{k+1} = w_k + \Lambda_k \Delta w_k. \quad \blacksquare \tag{2.4}$$

Let $D(\subset R^n)$ be an open convex subset that contains w^*. Let $\|\cdot\|$ denote the l_2 norm for vectors and matrices, and for matrices, let $\|\cdot\|_M$ and $\|\cdot\|_F$ be a matrix norm and the Frobenius norm, respectively. Then, by the norm equivalence, there is a positive constant η such that, for any matrix C,

$$\frac{1}{\eta}\|C\|_F \le \|C\| \le \eta\|C\|_F \quad \text{and} \quad \|C\|_F \le \eta\|C\|_M. \tag{2.5}$$

In this section, we make the following assumptions:
(XA1) The function $r_0(w)$ is continuously differentiable and there exists a positive constant ξ such that

$$\begin{aligned} \|r_0(w) - r_0(w')\| &\le \xi\|w - w'\|, \\ \|\nabla r_0(w) - \nabla r_0(w')\| &\le \xi\|w - w'\| \end{aligned}$$

for any $w,\ w' \in D$.
(XA2) The Jacobian matrix $\nabla r_0(w)$ is nonsingular at w^*.
(XA3) The vector p_k and the matrix Λ_k satisfy

$$\|p_k\| = O(\|r_0(w_k)\|^{1+\tau_1})$$

and

$$\|\Lambda_k - I\| = O(\|r_0(w_k)\|^{\tau_2}), \qquad \|\Lambda_k\| \le \lambda$$

for positive constants τ_1, τ_2 and λ. $\square$
It follows from assumption (XA1) that

$$\|r_0(w) - r_0(w') - \nabla r_0(w'')(w - w')\| \le \frac{1}{2}\xi(\|w - w''\| + \|w' - w''\|)\|w - w'\| \quad \text{for } \forall w, w', w'' \in D. \tag{2.6}$$

The following lemma is useful for convergence analysis in the following subsections.

Lemma 1 *Suppose that assumptions (XA1) and (XA2) hold. Then there exist an $\varepsilon > 0$ and a $\delta > 0$ such that, if*

$$\|w_k - w^*\| \leq \varepsilon, \qquad \|G_k - \nabla_u r_1(w^*)\|_M \leq \delta, \qquad w_k \in D,$$

there hold

(2.7)
$$\|J_k - \nabla r_0(w^*)\| \leq \eta^2\sqrt{\delta^2 + \xi^2\varepsilon^2} \quad \text{and} \quad \|J_k^{-1}\| \leq \zeta$$

for some positive constant ζ.

Proof. Since

$$J_k - \nabla r_0(w^*) = \begin{pmatrix} G_k - \nabla_u r_1(w^*) & \nabla_v r_1(w_k) - \nabla_v r_1(w^*) \\ \nabla_u r_2(w_k) - \nabla_u r_2(w^*) & \nabla_v r_2(w_k) - \nabla_v r_2(w^*) \end{pmatrix},$$

we have

$$\begin{aligned} \|J_k - \nabla r_0(w^*)\|_F^2 &= \|G_k - \nabla_u r_1(w^*)\|_F^2 + \|\nabla_v r_1(w_k) - \nabla_v r_1(w^*)\|_F^2 \\ &\quad + \|\nabla_u r_2(w_k) - \nabla_u r_2(w^*)\|_F^2 + \|\nabla_v r_2(w_k) - \nabla_v r_2(w^*)\|_F^2 \\ &\leq \|G_k - \nabla_u r_1(w^*)\|_F^2 + \|\nabla r_0(w_k) - \nabla r_0(w^*)\|_F^2 \\ &\leq \eta^2(\|G_k - \nabla_u r_1(w^*)\|_M^2 + \|\nabla r_0(w_k) - \nabla r_0(w^*)\|^2) \\ &\leq \eta^2(\delta^2 + \xi^2\|w_k - w^*\|^2). \end{aligned}$$

Thus we see

$$\|J_k - \nabla r_0(w^*)\| \leq \eta\|J_k - \nabla r_0(w^*)\|_F \leq \eta^2\sqrt{\delta^2 + \xi^2\varepsilon^2},$$

which proves the first inequality of (2.7).

By choosing ε and δ such that

$$\eta^2\sqrt{\delta^2 + \xi^2\varepsilon^2}\|\nabla r_0(w^*)^{-1}\| < 1,$$

it follows from Banach perturbation lemma that J_k is nonsingular and

$$\|J_k^{-1}\| \leq \zeta \equiv \frac{\|\nabla r_0(w^*)^{-1}\|}{1 - \eta^2\sqrt{\delta^2 + \xi^2\varepsilon^2}\|\nabla r_0(w^*)^{-1}\|},$$

which proves the second inequality of (2.7). □

2.1. Local and quadratic convergence of modified Newton methods

In this subsection, we give a sufficient condition for local and quadratic convergence of modified Newton methods. Letting $G_k = \nabla_u r_1(w_k)$ in Algorithm X, we have the following property.

Theorem 1 *Suppose that assumptions (XA1), (XA2) and (XA3) hold. Let $G_k = \nabla_u r_1(w_k)$ and $\tau_1 = \tau_2 = 1$ in Algorithm X. Let $\{w_k\}$ be generated by Algorithm X. Then there exists a positive constant ε such that if*

$$\|w_0 - w^*\| < \varepsilon,$$

then the sequence $\{w_k\}$ is well defined and satisfies

(2.8)
$$\|w_{k+1} - w^*\| \leq \nu\|w_k - w^*\|^2$$

for each $k \geq 0$, where ν is a positive constant, and the sequence $\{w_k\}$ converges to w^*.

Proof. We choose ε such that $\{w \mid \|w-w^*\|<\varepsilon\} \subset D$. Assume that

$$\|w_k-w^*\|<\varepsilon.$$

Since, by Lemma 1, $\nabla r_0(w_k)$ is nonsingular and $\|\nabla r_0(w_k)^{-1}\| \le \zeta$ for ε sufficiently small, we have

$$\begin{aligned} w_{k+1}-w^* &= (w_k-w^*)+\Lambda_k\Delta w_k \\ &= (I-\Lambda_k)(w_k-w^*)+\Lambda_k\nabla r_0(w_k)^{-1}(-r_0(w_k)+\nabla r_0(w_k)(w_k-w^*)+p_k), \end{aligned}$$

and hence, using (2.6)

$$\begin{aligned} &\|w_{k+1}-w^*\| \\ &\quad \le \|I-\Lambda_k\|\|w_k-w^*\|+\|\Lambda_k\|\|\nabla r_0(w_k)^{-1}\|\|r_0(w_k)-r_0(w^*)-\nabla r_0(w_k)(w_k-w^*)\| \\ &\qquad +\|\Lambda_k\|\|\nabla r_0(w_k)^{-1}\|\|p_k\| \\ &\quad = \|\Lambda_k-I\|\|w_k-w^*\|+O(\|w_k-w^*\|^2)+O(\|p_k\|). \end{aligned}$$

Since the assumptions give $\|p_k\|=O(\|r_0(w_k)\|^2)$ and $\|\Lambda_k-I\|=O(\|r_0(w_k)\|)$, there exists a positive constant ν such that

$$\|w_{k+1}-w^*\| \le \nu\|w_k-w^*\|^2,$$

which implies (2.8). Furthermore, we see

$$\|w_{k+1}-w^*\|<\varepsilon,$$

if ε is sufficiently small. Thus, by using mathematical induction, the proof is complete. □

2.2. Local and superlinear convergence of modified quasi-Newton methods

In this subsection, we show the local and superlinear convergence property of modified quasi-Newton methods. The following theorem gives local and linear convergence of modified quasi-Newton methods, and corresponds to the bounded deterioration theorem for unconstrained optimization by Broyden, Dennis and Moré [2].

Theorem 2 *Let $\{w_k\}$ be generated by Algorithm X. Suppose that assumptions (XA1), (XA2) and (XA3) hold. Assume that the sequence of matrices $\{G_k\}$ satisfies the bounded deterioration property*

$$\|G_{k+1}-\nabla_u r_1(w^*)\|_M \le (1+\beta_1\sigma_k)\|G_k-\nabla_u r_1(w^*)\|_M+\beta_2\sigma_k, \tag{2.9}$$

where β_1 and β_2 are positive constants, and

$$\sigma_k=\max(\|w_{k+1}-w^*\|,\|w_k-w^*\|). \tag{2.10}$$

Then for any $\nu\in(0,1)$, there exist positive constants $\varepsilon=\varepsilon(\nu)$ and $\delta=\delta(\nu)$ such that if

$$\|w_0-w^*\|<\varepsilon$$

and

$$\|G_0-\nabla_u r_1(w^*)\|_M<\frac{\delta}{2},$$

then, the sequence $\{w_k\}$ is well defined and converges to w^* with

$$\|w_{k+1}-w^*\| \le \nu\|w_k-w^*\| \tag{2.11}$$

for each $k\ge 0$.

Proof. We choose ε such that $\{w \mid \|w - w^*\| < \varepsilon\} \subset D$. By mathematical induction with respect to k, we will prove that if, for $l = 0, 1, \cdots, k$,

$$\|w_l - w^*\| \leq \nu \|w_{l-1} - w^*\| < \varepsilon \qquad \text{and} \qquad \|G_l - \nabla_u r_1(w^*)\|_M < \delta, \tag{2.12}$$

then

$$\|w_{k+1} - w^*\| \leq \nu \|w_k - w^*\| < \varepsilon \qquad \text{and} \qquad \|G_{k+1} - \nabla_u r_1(w^*)\|_M < \delta. \tag{2.13}$$

For $k = 0$, the conditions (2.12) clearly hold. Note that, in this case, the first inequality in (2.12) means $\|w_0 - w^*\| < \varepsilon$. Now we show that (2.12) implies (2.13) for any $k \geq 0$. For ε and δ sufficiently small, Lemma 1 guarantees that J_k is nonsingular and $\|J_k^{-1}\| \leq \zeta$. It follows from the linear system (2.2) that

$$\begin{aligned} w_{k+1} - w^* &= (w_k - w^*) + \Lambda_k \Delta w_k \\ &= (I - \Lambda_k)(w_k - w^*) + \Lambda_k J_k^{-1} p_k + \Lambda_k J_k^{-1}(J_k - \nabla r_0(w^*))\nabla r_0(w^*)^{-1} r_0(w_k) \\ &\quad -\Lambda_k \nabla r_0(w^*)^{-1}(r_0(w_k) - r_0(w^*) - \nabla r_0(w^*)(w_k - w^*)). \end{aligned}$$

Hence

$$\begin{aligned} \|w_{k+1} - w^*\| &\leq \|I - \Lambda_k\|\|w_k - w^*\| + O(\|p_k\|) + \|J_k - \nabla r_0(w^*)\| O(\|r_0(w_k)\|) \\ &\quad + O(\|w_k - w^*\|^2) \\ &\leq \zeta'(\varepsilon^{\min(1,\tau_1,\tau_2)} + \sqrt{\delta^2 + \xi^2\varepsilon^2}\,)\|w_k - w^*\|, \end{aligned}$$

where ζ' is some positive constant. If we choose ε and δ such that

$$\zeta'(\varepsilon^{\min(1,\tau_1,\tau_2)} + \sqrt{\delta^2 + \xi^2\varepsilon^2}\,) < \nu,$$

then

$$\|w_{k+1} - w^*\| \leq \nu\|w_k - w^*\| < \varepsilon.$$

By using the same technique of Broyden, Dennis and Moré [2], we can show that

$$\|G_{k+1} - \nabla_u r_1(w^*)\|_M < \delta.$$

Therefore, the theorem is proved. □

Next we give a necessary and sufficient condition for superlinear convergence of these methods. We note that expression (2.15) given below corresponds to the Dennis-Moré condition [6] in the case of unconstrained optimization.

Theorem 3 *Assume that the sequence $\{w_k\}$ converges linearly to w^*. Suppose that assumptions (XA1), (XA2) and (XA3) hold. Then the following conditions are equivalent.*
(a) The sequence $\{G_k\}$ satisfies

$$\lim_{k\to\infty} \frac{\|(G_k - \nabla_u r_1(w^*))(u_{k+1} - u_k)\|}{\|w_{k+1} - w_k\|} = 0. \tag{2.14}$$

(b) The sequence $\{J_k\}$ satisfies

$$\lim_{k\to\infty} \frac{\|(J_k - \nabla r_0(w^*))(w_{k+1} - w_k)\|}{\|w_{k+1} - w_k\|} = 0. \tag{2.15}$$

(c) The sequence $\{r_0(w_k)\}$ *satisfies*

$$\lim_{k\to\infty} \frac{\|r_0(w_{k+1})\|}{\|w_{k+1}-w_k\|} = 0. \tag{2.16}$$

(d) The sequence $\{w_k\}$ *converges superlinearly to* w^**, i.e.*

$$\lim_{k\to\infty} \frac{\|w_{k+1}-w^*\|}{\|w_k-w^*\|} = 0. \tag{2.17}$$

Proof. First we note that linear convergence (see expression (2.11)) implies

$$\frac{\|w_k-w^*\|}{\|w_{k+1}-w_k\|} \le \frac{1}{1-\nu}. \tag{2.18}$$

(a) $\Longleftrightarrow$ (b): Since

(2.19)

$$\begin{aligned}&(J_k-\nabla r_0(w^*))(w_{k+1}-w_k)\\ &=\begin{pmatrix}(G_k-\nabla_u r_1(w^*))(u_{k+1}-u_k)+(\nabla_v r_1(w_k)-\nabla_v r_1(w^*))(v_{k+1}-v_k)\\ (\nabla_u r_2(w_k)-\nabla_u r_2(w^*))(u_{k+1}-u_k)+(\nabla_v r_2(w_k)-\nabla_v r_2(w^*))(v_{k+1}-v_k)\end{pmatrix},\end{aligned}$$

the l_2 norm for a vector implies

$$\begin{aligned}&\|(G_k-\nabla_u r_1(w^*))(u_{k+1}-u_k)\|\\ &\le \|(G_k-\nabla_u r_1(w^*))(u_{k+1}-u_k)+(\nabla_v r_1(w_k)-\nabla_v r_1(w^*))(v_{k+1}-v_k)\|\\ &\quad +\|(\nabla_v r_1(w_k)-\nabla_v r_1(w^*))(v_{k+1}-v_k)\|\\ &\le \|(J_k-\nabla r_0(w^*))(w_{k+1}-w_k)\|+\|\nabla_v r_1(w_k)-\nabla_v r_1(w^*)\|\|w_{k+1}-w_k\|.\end{aligned}$$

Then, by the continuity of $\nabla_v r_1(w)$, it is clear that (b) implies (a). Conversely, by (2.19), we have

$$\begin{aligned}&\|(J_k-\nabla r_0(w^*))(w_{k+1}-w_k)\|^2\\ &\le (\|(G_k-\nabla_u r_1(w^*))(u_{k+1}-u_k)\|+\|(\nabla_v r_1(w_k)-\nabla_v r_1(w^*))(v_{k+1}-v_k)\|)^2\\ &\quad +(\|(\nabla_u r_2(w_k)-\nabla_u r_2(w^*))(u_{k+1}-u_k)\|+\|(\nabla_v r_2(w_k)-\nabla_v r_2(w^*))(v_{k+1}-v_k)\|)^2\\ &\le \left(\frac{\|(G_k-\nabla_u r_1(w^*))(u_{k+1}-u_k)\|^2}{\|w_{k+1}-w_k\|^2}+O(\|w_k-w^*\|)\right)\|w_{k+1}-w_k\|^2.\end{aligned}$$

Thus (a) implies (b).
(b) $\Longrightarrow$ (c): Since

(2.20)

$$\begin{aligned}r_0(w_{k+1}) &= (r_0(w_{k+1})-r_0(w_k)-\nabla r_0(w^*)(w_{k+1}-w_k))\\ &\quad -(J_k-\nabla r_0(w^*))(w_{k+1}-w_k)-J_k(\Lambda_k-I)J_k^{-1}(r_0(w_k)-r_0(w^*))\\ &\quad +(J_k(\Lambda_k-I)J_k^{-1}+I)p_k,\end{aligned}$$

we have

$$\begin{aligned}\|r_0(w_{k+1})\| &= O(\|w_{k+1}-w_k\|\|w_k-w^*\|)+\|(J_k-\nabla r_0(w^*))(w_{k+1}-w_k)\|\\ &\quad +O(\|w_k-w^*\|^{1+\tau_2})+O(\|w_k-w^*\|^{1+\tau_1}).\end{aligned} \tag{2.21}$$

Therefore, expressions (2.18) and (2.21) yield

$$\lim_{k\to\infty} \frac{\|r_0(w_{k+1})\|}{\|w_{k+1} - w_k\|} = 0.$$

(c) $\Longrightarrow$ (b): Since (2.20) is rewritten by

$$\begin{aligned}(J_k - \nabla r_0(w^*))(w_{k+1} - w_k) &= (r_0(w_{k+1}) - r_0(w_k) - \nabla r_0(w^*)(w_{k+1} - w_k)) \\ &\quad -J_k(\Lambda_k - I)J_k^{-1}(r_0(w_k) - r_0(w^*)) \\ &\quad +(J_k(\Lambda_k - I)J_k^{-1} + I)p_k - r_0(w_{k+1}),\end{aligned}$$

we see that

$$\begin{aligned}\|(J_k - \nabla r_0(w^*))(w_{k+1} - w_k)\| &= O(\|w_{k+1} - w_k\|\|w_k - w^*\|) + O(\|w_k - w^*\|^{1+\tau_2}) \\ &\quad +O(\|w_k - w^*\|^{1+\tau_1}) + o(\|w_{k+1} - w_k\|).\end{aligned}$$

From (2.18), this implies (b).
(c) $\Longleftrightarrow$ (d): The result follows directly from Broyden, Dennis and Moré [2].
Therefore, the theorem is proved. □

3. Convergence properties of primal-dual interior point methods

Now we consider the primal-dual interior point methods for solving the constrained optimization problem (1.1). The algorithm is given by Algorithm I in Section 1. Within the framework of this algorithm, Yamashita and Yabe [26] analyzed three kinds of step size rules given below:

Step size rule A (Single step size for x and z)

$$\alpha_{xk} = \alpha_{zk} = \alpha_k,$$

(3.1)

$$\alpha_k \equiv \min\left\{1, \gamma_k \min_i\left\{-\frac{(x_k)_i}{(\Delta x_k)_i}\middle|(\Delta x_k)_i < 0\right\}, \gamma_k \min_i\left\{-\frac{(z_k)_i}{(\Delta z_k)_i}\middle|(\Delta z_k)_i < 0\right\}\right\},$$

where $\gamma_k \in (0,1)$. The step size α_{yk} is determined by

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

Step size rule B (Different step sizes for x and z)

$$\alpha_{xk} = \min\left\{1, \gamma_k \min_i\left\{-\frac{(x_k)_i}{(\Delta x_k)_i}\middle|(\Delta x_k)_i < 0\right\}\right\}, \tag{3.2}$$

and

$$\alpha_{zk} = \min\left\{1, \gamma_k \min_i\left\{-\frac{(z_k)_i}{(\Delta z_k)_i}\middle|(\Delta z_k)_i < 0\right\}\right\}, \tag{3.3}$$

where $\gamma_k \in (0,1)$. The step size α_{yk} is determined by

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

Step size rule C (Globally convergent step size rule)

(3.4)
$$\alpha_{xk} = \min\left\{1, \gamma_k \min_i \left\{ -\frac{(x_k)_i}{(\Delta x_k)_i} \middle| (\Delta x_k)_i < 0 \right\}\right\},$$

where $\gamma_k \in (0,1)$. The step size α_{zk} is the largest step that satisfies

$$\alpha_{zk} \leq 1,$$

(3.5)

$$\min\left\{\frac{\mu_k}{M_{Lk}((x_k)_i + \alpha_{xk}(\Delta x_k)_i)}, (z_k)_i\right\} \leq (z_k)_i + \alpha_{zk}(\Delta z_k)_i \leq \max\left\{\frac{M_{Uk}\mu_k}{(x_k)_i + \alpha_{xk}(\Delta x_k)_i}, (z_k)_i\right\}, \quad i = 1, \cdots n,$$

where $\mu_k > 0$, and where M_{Lk} and M_{Uk} are positive numbers that satisfy

(3.6)

$$M_{Lk} > \max\left\{1, \frac{2\mu_k}{(1-\gamma_k)\min_i\{(x_k)_i(z_k)_i\}}\right\}, \quad M_{Uk} > \max\left\{3, \frac{3\max_i\{(x_k)_i(z_k)_i\}}{\mu_k}\right\}.$$

The step size α_{yk} is determined by

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

By taking these rules into account, Algorithm I stated in Section 1 yields various types of algorithms. Specifically, **Algorithms A, B and C** denote those such that, in Algorithm I, the step sizes are determined by Step size rule A, B and C, respectively. Note that the parameter μ_k is chosen to satisfy $\mu_k \geq 0$ for Algorithms A and B, and $\mu_k > 0$ for Algorithm C.

Let $w^* = (x^*, y^*, z^*)$ be a K-K-T point of problem (1.1). Let D be an open convex set that contains w^*. We assume the following standard conditions on the functions which appear in problem (1.1) and on the point w^*.

(A1) The second derivatives of the functions f and g are Lipschitz continuous in D.

(A2) The point x^* satisfies the regularity condition, i.e. the vectors $\nabla g_i(x^*)$, $i = 1, ..., m$ and e_i, $i \in \{ i \mid (x^*)_i = 0\}$ are linearly independent, where e_i is the i-th column of the identity matrix.

(A3) The strict complementarity of the solution w^* is satisfied, i.e. $(z^*)_i > 0$ for $i \in \{ i \mid (x^*)_i = 0\}$.

(A4) The second order sufficiency condition for the optimality is satisfied at the point w^*.

We should note that the second order sufficiency condition for the optimality is that, for all $v \neq 0$ satisfying $\nabla g_i(x^*)^T v = 0$, $i = 1, ..., m$, $e_i^T v = 0$ for $i \in \{ i \mid (x^*)_i = 0, \ (z^*)_i > 0\}$ and $e_i^T v \geq 0$ for $i \in \{ i \mid (x^*)_i = 0, \ (z^*)_i = 0\}$, there holds $v^T \nabla_x^2 L(w^*) v > 0$. Since assumption (A3) implies $\{ i \mid (x^*)_i = 0, \ (z^*)_i = 0\} = \emptyset$, assumption (A4) means that for all $v \neq 0$ satisfying $\nabla g_i(x^*)^T v = 0$, $i = 1, ..., m$ and $e_i^T v = 0$, $i \in \{ i \mid (x^*)_i = 0\}$, there holds $v^T \nabla_x^2 L(w^*) v > 0$.

Under the assumptions stated above, we have the following lemmas.

Lemma 2 (see [10]) *Under assumptions (A1), (A2), (A3) and (A4), the matrix $\nabla r(w^*)$ is nonsingular.* □

Lemma 3 *Suppose that assumptions (A1), (A2), (A3) and (A4) hold and that the sequence $\{w_k\}$ is generated by Algorithm A, B or C. Then there exist an $\varepsilon > 0$ and a $\delta > 0$ such that, if*

$$\|w_k - w^*\| \le \varepsilon, \quad \left\|G_k - \nabla_x^2 L(w^*)\right\|_M \le \delta, \tag{3.7}$$

then there exists a positive constant ζ such that

$$\|J_k - \nabla r_0(w^*)\| \le \eta^2 \sqrt{\delta^2 + \xi^2 \varepsilon^2} \quad \text{and} \quad \left\|J_k^{-1}\right\| \le \zeta, \tag{3.8}$$

and furthermore

$$\|\Lambda_k - I\| \le (1 - \gamma_k) + O(\|r_0(w_k)\|) + O(\mu_k). \tag{3.9}$$

Proof. By setting $u = x, v = (y, z)^t$ and $r_1 = \nabla_x L$, the result follows directly from Lemma 1 of the present paper and Lemma 7 in [26]. □

Now we pay our attention to the local and quadratic/superlinear convergence property of primal-dual interior point Newton/quasi-Newton methods. By using convergence theorems given in Section 2, we obtain the following corollaries, and these are the main convergence results of our previous paper [26]. We also note that Corollary 1 corresponds to the result by El-Bakry et al. [8].

Corollary 1 *Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Let $G_k = \nabla_x^2 L(w_k)$ and choose the parameters such that*

$$1 - \gamma_k = O(\|r_0(w_k)\|) \quad \text{and} \quad \mu_k = O(\|r_0(w_k)\|^2) \tag{3.10}$$

in Algorithms A, B and C. Let $\{w_k\}$ be generated by Algorithm A, B or C. Then there exists a positive constant ε such that if

$$\|w_0 - w^*\| < \varepsilon, \qquad w_0 \in D,$$

then the sequence $\{w_k\}$ is well defined and converges to w^*. Furthermore

$$\|w_{k+1} - w^*\| \le \nu \|w_k - w^*\|^2$$

for each $k \ge 0$, where ν is a positive constant.

Proof. By setting $u = x$, $v = (y, z)^t$ and $\tau_1 = \tau_2 = 1$, Theorem 1 and Lemma 3 yield the result. □

Corollary 2 *Let $\{w_k\}$ be generated by Algorithm A, B or C. Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Choose the parameters μ_k and γ_k such that*

$$\mu_k = O(\|r_0(w_k)\|^{1+\tau_1}) \quad \text{and} \quad 1 - \gamma_k = O(\|r_0(w_k)\|^{\tau_2}) \tag{3.11}$$

for positive constants τ_1 and τ_2 in each algorithm. Assume that the sequence of matrices $\{G_k\}$ satisfies the bounded deterioration property

$$\|G_{k+1} - \nabla_x^2 L(w^*)\|_M \le (1 + \beta_1 \sigma_k) \|G_k - \nabla_x^2 L(w^*)\|_M + \beta_2 \sigma_k, \tag{3.12}$$

where β_1 and β_2 are positive constants, and

$$\sigma_k = \max(\|w_{k+1} - w^*\|, \|w_k - w^*\|). \tag{3.13}$$

Then for each of Algorithm A, B and C, and each $\nu \in (0,1)$, there exist positive constants $\varepsilon = \varepsilon(\nu)$ and $\delta = \delta(\nu)$ such that if

$$\|w_0 - w^*\| < \varepsilon, \qquad w_0 \in D$$

and

$$\|G_0 - \nabla_x^2 L(w^*)\|_M < \frac{\delta}{2},$$

the sequence $\{w_k\}$ is well defined and converges to w^*. Furthermore,

$$\|w_{k+1} - w^*\| \le \nu \|w_k - w^*\| \tag{3.14}$$

for each $k \ge 0$.

Proof. By setting $u = x$ and $v = (y, z)^t$, Theorem 2 and Lemma 3 yield the result. □

Corollary 3 *Let $\{w_k\}$ be generated by Algorithm A, B or C. Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Choose the parameters μ_k and γ_k satisfying (3.11). Assume that the sequence $\{w_k\}$ converges linearly to w^*. Then the following four conditions are equivalent.*

(a) The sequence $\{G_k\}$ satisfies

$$\lim_{k\to\infty} \frac{\|(G_k - \nabla_x^2 L(w^*))(x_{k+1} - x_k)\|}{\|w_{k+1} - w_k\|} = 0. \tag{3.15}$$

(b) The sequence $\{J_k\}$ satisfies

$$\lim_{k\to\infty} \frac{\|(J_k - \nabla r_0(w^*))(w_{k+1} - w_k)\|}{\|w_{k+1} - w_k\|} = 0. \tag{3.16}$$

(c) The sequence $\{r_0(w_k)\}$ satisfies

$$\lim_{k\to\infty} \frac{\|r_0(w_{k+1})\|}{\|w_{k+1} - w_k\|} = 0. \tag{3.17}$$

(d) The sequence $\{w_k\}$ converges superlinearly to w^*, i.e.

$$\lim_{k\to\infty} \frac{\|w_{k+1} - w^*\|}{\|w_k - w^*\|} = 0. \tag{3.18}$$

Proof. By setting $u = x$ and $v = (y, z)^t$, Theorem 3 yields the result. □

Before closing this section, we show examples of primal-dual interior point quasi-Newton updates that satisfy the bounded deterioration property (3.12) and condition (3.15) given above. Since the matrix G_k approximates the Hessian matrix $\nabla_x^2 L(w_k)$, we impose the following secant condition on G_{k+1}:

$$G_{k+1} s_k = q_k, \tag{3.19}$$

where

$$s_k = x_{k+1} - x_k \quad \text{and} \quad q_k = \nabla_x L(x_{k+1}, y_{k+1}, z_{k+1}) - \nabla_x L(x_k, y_{k+1}, z_{k+1}). \tag{3.20}$$

For simplicity, we omit the subscript k and denote the subscript $(k+1)$ by $+$.

First we consider the PSB update:

$$G_+ = G + \frac{(q - Gs)s^t + s(q - Gs)^t}{s^t s} - \frac{s^t(q - Gs)}{(s^t s)^2} ss^t. \tag{3.21}$$

The following lemma indicates that the PSB update satisfies the conditions stated in Corollaries 2 and 3. The proof can be found in Appendix.

Lemma 4 *(1) The PSB update (3.21) has the bounded deterioration property (3.12) at each iteration.*
(2) Suppose that the sequence $\{w_k\}$ converges linearly to w^*. Then the PSB update satisfies (3.15).□

Combining Corollaries 2, 3 and Lemma 4 , we obtain the following result.

Theorem 4 *Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Choose the parameters μ_k and γ_k satisfying (3.11). Let $\{w_k\}$ be generated by Algorithm A, B or C, and let $\{G_k\}$ be generated by the PSB update (3.21). Then the sequence $\{w_k\}$ converges locally and superlinearly to w^*.* □

Next we consider the Broyden family:

$$G_+ = G - \frac{Gss^tG}{s^tGs} + \frac{qq^t}{q^ts} + \phi(s^tGs)hh^t, \tag{3.22}$$

where

$$h = \frac{q}{q^ts} - \frac{Gs}{s^tGs}$$

and ϕ is a parameter. This family contains important members. For example, setting $\phi = 0$ yields the BFGS update:

$$G_+ = G - \frac{Gss^tG}{s^tGs} + \frac{qq^t}{q^ts}, \tag{3.23}$$

and setting $\phi = 1$ yields the DFP update:

$$G_+ = G - \frac{Gsq^t + qs^tG}{q^ts} + \left(1 + \frac{s^tGs}{q^ts}\right)\frac{qq^t}{q^ts}. \tag{3.24}$$

The following theorem indicates that the convex class of the Broyden family, i.e. $0 \le \phi \le 1$, satisfies the conditions stated in Corollaries 2 and 3.

Theorem 5 *Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Choose the parameters μ_k and γ_k satisfying (3.11). Suppose that the Hessian matrix $\nabla_x^2 L(w^*)$ is positive definite. Let $\{w_k\}$ be generated by Algorithm A, B or C, and let $\{G_k\}$ be generated by the convex class of the Broyden family (3.22). Then the sequence $\{w_k\}$ converges locally and superlinearly to w^*.* □

The proof of this theorem can be similarly shown by combining the proofs of Lemma 4 and Theorem 4 and the proofs of the standard Broyden family (for example, see [9]). Thus we omit it.

4. Superlinear convergence in (x, z)

In the previous section, we have presented the necessary and sufficient conditions for the sequence $\{(x_k, y_k, z_k)\}$, which is generated by the primal-dual interior point quasi-Newton methods, to converge superlinearly to the K-K-T point (x^*, y^*, z^*). In this section, we will investigate a local behavior of the sequence $\{(x_k, z_k)\}$. In general, a Q-rate in (x, y, z) implies no more than the corresponding R-rate in (x, z). Therefore, it is interesting to consider necessary and sufficient conditions for the Q-rate in (x, z). This section gives such necessary and sufficient conditions for the part $\{(x_k, z_k)\}$ of the sequence $\{w_k\}$ to converge superlinearly to (x^*, z^*). The results mentioned below correspond to those by Boggs, Tolle and Wang [1] and Coleman [3] for the SQP method.

First we show a relationship between the sequences $\{(x_k, z_k)\}$ and $\{y_k\}$. The following theorem shows how the convergence of $\{(x_k, z_k)\}$ affects the convergence of $\{y_k\}$.

Theorem 6 *Let $\{w_k\}$ be generated by Algorithm A, B or C. Suppose that assumptions (A1), (A2), (A3) and (A4) hold, and that the matrices $\{G_k\}$ and vectors $\{y_k\}$ are bounded. Then, for some positive constants c_1, c_2, c_3 and c_4,*

$$\|y_{k+1} - y^*\| \le c_1 \left\| \begin{pmatrix} x_{k+1} - x^* \\ z_{k+1} - z^* \end{pmatrix} \right\| + c_2 \left\| \begin{pmatrix} x_k - x^* \\ z_k - z^* \end{pmatrix} \right\| + c_3 |1 - \alpha_{yk}| \|y_k - y^*\|. \tag{4.1}$$

If $\alpha_{yk} = 1$, then

$$\|y_{k+1} - y^*\| \le c_1 \left\| \begin{pmatrix} x_{k+1} - x^* \\ z_{k+1} - z^* \end{pmatrix} \right\| + c_2 \left\| \begin{pmatrix} x_k - x^* \\ z_k - z^* \end{pmatrix} \right\|. \tag{4.2}$$

If the sequence $\{(x_k, z_k)\}$ converges linearly to (x^*, z^*), then

$$\|y_{k+1} - y^*\| \le c_4 \left\| \begin{pmatrix} x_k - x^* \\ z_k - z^* \end{pmatrix} \right\| + c_3 |1 - \alpha_{yk}| \|y_k - y^*\|. \tag{4.3}$$

Proof. Since $y_{k+1} = y_k + \alpha_{yk} \Delta y_k$ and assumption (A2) implies that the matrix $A(x^*)A(x^*)^t$ is nonsingular, we have

$$\begin{aligned} y_{k+1} - y^* &= (A(x^*)A(x^*)^t)^{-1}(A(x^*)A(x^*)^t)(y_{k+1} - y^*) \\ &= (A(x^*)A(x^*)^t)^{-1}A(x^*)\{(A(x^*)^t - A(x_k)^t)y_{k+1} + (A(x_k)^t(y_k + \Delta y_k) - A(x^*)^t y^*) \\ &\qquad + (\alpha_{yk} - 1)A(x_k)^t \Delta y_k\}. \end{aligned}$$

Then, for a positive constant c_5,

$$\|y_{k+1} - y^*\| \le c_5(\|x_k - x^*\| + \|A(x_k)^t(y_k + \Delta y_k) - A(x^*)^t y^*\| + |1 - \alpha_{yk}| \, \|A(x_k)^t \Delta y_k\|). \tag{4.4}$$

Since equation (1.9) and the K-K-T conditions of problem (1.1) yield

$$A(x_k)^t(y_k + \Delta y_k) = G_k \Delta x_k + \nabla f(x_k) - (z_k + \Delta z_k) \tag{4.5}$$

and

$$A(x^*)^t y^* = \nabla f(x^*) - z^*, \tag{4.6}$$

we have, for positive constants c_6 and c_7,

$$\begin{aligned} &\|A(x_k)^t(y_k+\Delta y_k)-A(x^*)^t y^*\| \\ &\le \|G_k\|\|\Delta x_k\|+\|\nabla f(x_k)-\nabla f(x^*)\|+\|z_k-z^*\|+\|\Delta z_k\| \\ &\le c_6\left(\left\|\begin{pmatrix}\Delta x_k\\ \Delta z_k\end{pmatrix}\right\|+\left\|\begin{pmatrix}x_k-x^*\\ z_k-z^*\end{pmatrix}\right\|\right) \\ &\le c_7\left(\left\|\begin{pmatrix}x_{k+1}-x^*\\ z_{k+1}-z^*\end{pmatrix}\right\|+\left\|\begin{pmatrix}x_k-x^*\\ z_k-z^*\end{pmatrix}\right\|\right). \end{aligned} \tag{4.7}$$

It also follows from (4.5) and (4.6) that, for a positive constant c_8,

$$\begin{aligned} \|A(x_k)^t\Delta y_k\| &\le \|G_k\|\|\Delta x_k\|+\|\nabla f(x_k)-\nabla f(x^*)\|+\|z_k-z^*\|+\|\Delta z_k\| \\ &\quad +\|A(x_k)^t y_k-A(x^*)^t y^*\| \\ &\le c_8\left(\left\|\begin{pmatrix}x_{k+1}-x^*\\ z_{k+1}-z^*\end{pmatrix}\right\|+\left\|\begin{pmatrix}x_k-x^*\\ z_k-z^*\end{pmatrix}\right\|+\|y_k-y^*\|\right). \end{aligned} \tag{4.8}$$

Therefore, by using equations (4.4), (4.7) and (4.8), we obtain

$$\|y_{k+1}-y^*\|\le c_1\left\|\begin{pmatrix}x_{k+1}-x^*\\ z_{k+1}-z^*\end{pmatrix}\right\|+c_2\left\|\begin{pmatrix}x_k-x^*\\ z_k-z^*\end{pmatrix}\right\|+c_3|1-\alpha_{yk}|\|y_k-y^*\|,$$

and the proof is complete. ◻

Expression (4.2) suggests that for $\alpha_{yk}=1$, the convergence of the sequence $\{(x_k,z_k)\}$ implies the convergence of the sequence $\{y_k\}$. Note that the boundedness of $\{y_k\}$ and the convergence of $\{(x_k,y_k)\}$ are discussed below.

Now we will give necessary and sufficient conditions for the Q-rate in (x,z). In what follows, we suppose that assumptions (A1), (A2), (A3) and (A4) given in Section 3 hold. Let $A^-(x)\in R^{n\times m}$ be a generalized inverse of $A(x)$ that satisfies

$$A(x)A^-(x)A(x)=A(x).$$

Since the matrix $A(x)$ is of full row rank, we have

$$A(x)A^-(x)=I.$$

Let $B(x)\in R^{(n-m)\times n}$ be a full rank matrix that satisfies

$$B(x)A(x)^t=O$$

and be differentiable with respect to x. The existence of such a matrix $B(x)$ near the solution w^* was proved by Goodman [13]. Then it is easily shown that a matrix

$$\begin{pmatrix}A^-(x)^t\\ B(x)\end{pmatrix}\in R^{n\times n}$$

is nonsingular.

First we consider the quasi-Newton method for solving the nonlinear system of equations (1.3). Then equation (1.9) in Algorithm I is equivalent to the following system

$$\begin{pmatrix}\begin{pmatrix}(A^-(x_k))^t\\ B(x_k)\end{pmatrix} & & O\\ & I & \\ O & & I\end{pmatrix}J_k\Delta w_k=-\begin{pmatrix}\begin{pmatrix}(A^-(x_k))^t\\ B(x_k)\end{pmatrix} & & O\\ & I & \\ O & & I\end{pmatrix}r(w_k).$$

This yields

$$\Delta y_k = (A^-(x_k))^t(G_k\Delta x_k + \nabla f(x_k) - (z_k + \Delta z_k)) - y_k, \tag{4.9}$$

$$\begin{pmatrix} B(x_k)G_k & -B(x_k) \\ A(x_k) & O \\ Z_k & X_k \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta z_k \end{pmatrix} = -\begin{pmatrix} B(x_k)(\nabla f(x_k) - z_k) \\ g(x_k) \\ X_k Z_k e - \mu_k e \end{pmatrix}. \tag{4.10}$$

Note that (4.10) does not correspond to the Newton equation with respect to (x, z). This fact makes it difficult to get convergence property directly from the preceding result. So we must consider the quasi-Newton method after rearranging (1.3).

We note that (1.3) is equivalent to the equation

$$\begin{aligned} \begin{pmatrix} (A^-(x))^t \\ B(x) \end{pmatrix} r_L(w) &= 0, \\ r_E(x) &= 0, \\ r_C(x, z) &= 0. \end{aligned}$$

Then we have

$$y = (A^-(x))^t(\nabla f(x) - z), \tag{4.11}$$

$$B(x)(\nabla f(x) - z) = 0, \tag{4.12}$$

$$g(x) = 0, \tag{4.13}$$

$$XZe - \mu e = 0. \tag{4.14}$$

Here expression (4.11) shows that (x, z) gives y. On the other hand, equations (4.12), (4.13) and (4.14) only depend on (x, z).

We apply the Newton method to (4.12), (4.13) and (4.14), and we have the Newton equation

$$\tilde{J}(x_k, z_k) \begin{pmatrix} \Delta x_k \\ \Delta z_k \end{pmatrix} = -\tilde{r}(x_k, z_k), \tag{4.15}$$

where

$$\tilde{r}(x, z) = \begin{pmatrix} B(x)(\nabla f(x) - z) \\ g(x) \\ XZe - \mu e \end{pmatrix}, \tag{4.16}$$

$$\tilde{J}(x, z) = \begin{pmatrix} \dot{B}(x) r_L(w) + B(x)\nabla_x^2 L(w) & -B(x) \\ A(x) & O \\ Z & X \end{pmatrix} \tag{4.17}$$

and $\dot{B}(x)$ denotes the derivative of $B(x)$ with respect to x. At (x^*, z^*), the matrix $\tilde{J}$ is given by

$$\tilde{J}(x^*, z^*) = \begin{pmatrix} B(x^*)\nabla_x^2 L(w^*) & -B(x^*) \\ A(x^*) & O \\ Z^* & X^* \end{pmatrix} \tag{4.18}$$

and becomes nonsingular. We consider a quasi-Newton method in which the matrix

$$\dot{B}(x_k) r_L(w_k) + B(x_k)\nabla_x^2 L(w_k)$$

is approximated by $B(x_k)G_k$, where G_k is an approximation to the Hessian matrix $\nabla_x^2 L(w_k)$. Then a quasi-Newton equation is identical with (4.10), so we can apply the general theory in Section 2 to (4.10). Specifically speaking, by setting $u = x$ and $v = z$ in Theorem 3, we obtain the following theorem.

Theorem 7 *Suppose that assumptions (A1), (A2), (A3) and (A4) hold. Choose the parameters μ_k and γ_k satisfying (3.11). Let $\{w_k\}$ be generated by the quasi-Newton method based on Algorithm A , B or C . Assume that $\{y_k\}$ converges to y^* and that the sequence $\{(x_k, z_k)\}$ converges linearly to (x^*, z^*). Let $\{G_k\}$ be a sequence such that the matrix*

$$\begin{pmatrix} B(x_k)G_k & -B(x_k) \\ A(x_k) & O \\ Z_k & X_k \end{pmatrix}$$

is nonsingular for each $k \geq 0$.

Then the following three conditions are equivalent.

(a) *The sequence $\{(x_k, z_k)\}$ converges superlinearly to (x^*, z^*).*

(b) *The sequence $\{G_k\}$ satisfies*

$$\lim_{k\to\infty} \frac{\|(B(x_k)G_k - B(x^*)\nabla_x^2 L(w^*))(x_{k+1} - x_k)\|}{\left\| \begin{pmatrix} x_{k+1} - x_k \\ z_{k+1} - z_k \end{pmatrix} \right\|} = 0. \tag{4.19}$$

(c) *The sequence $\{G_k\}$ satisfies*

$$\lim_{k\to\infty} \frac{\|B(x_k)(G_k - \nabla_x^2 L(w^*))(x_{k+1} - x_k)\|}{\left\| \begin{pmatrix} x_{k+1} - x_k \\ z_{k+1} - z_k \end{pmatrix} \right\|} = 0. \tag{4.20}$$

Proof. The parts (a) and (b) are straightforward results from Theorem 3. The continuity of $B(x)$ near x^* yields the part (c). Therefore, the theorem is proved. □

We note that a special choice of $B(x)$ yields a specific result. For example, let $B(x)$ be an orthonormal basis such that

$$B(x)B(x)^t = I. \tag{4.21}$$

Then a matrix

$$P(x) \equiv B(x)^t B(x) = I - A(x)^t (A(x)A(x)^t)^{-1} A(x) \tag{4.22}$$

becomes an orthogonal projection matrix onto the orthogonal complement of the range space of $A(x)^t$. This fact gives us the following result.

Corollary 4 *Suppose that all the assumptions of Theorem 7 hold. Assume that the matrix $B(x)$ satisfies (4.21). Then the following conditions are equivalent.*

(a) *The sequence $\{(x_k, z_k)\}$ converges superlinearly to (x^*, z^*).*

(b) *The sequence $\{G_k\}$ satisfies*

$$\lim_{k\to\infty} \frac{\|P(x_k)(G_k - \nabla_x^2 L(w^*))(x_{k+1} - x_k)\|}{\left\| \begin{pmatrix} x_{k+1} - x_k \\ z_{k+1} - z_k \end{pmatrix} \right\|} = 0. \quad \square \tag{4.23}$$

The preceding corollary corresponds to the main result (Theorem 5.1) by Martinez, Parada and Tapia [18].

Appendix: Proof of Lemma 4

The proof is similar to that of convergence theorem in [2].

(1) The update (3.21) yields

$$\begin{aligned} G_+ - \nabla_x^2 L(w^*) &= \left(I - \frac{ss^t}{s^t s}\right)(G - \nabla_x^2 L(w^*))\left(I - \frac{ss^t}{s^t s}\right) \\ &\quad + \frac{(q - \nabla_x^2 L(w^*)s)s^t + s(q - \nabla_x^2 L(w^*)s)^t}{s^t s} - \frac{s^t(q - \nabla_x^2 L(w^*)s)}{(s^t s)^2} ss^t. \end{aligned}$$

Setting

$$P = I - \frac{ss^t}{s^t s} \quad \text{and} \quad E = G - \nabla_x^2 L(w^*),$$

we have

$$\|G_+ - \nabla_x^2 L(w^*)\|_F \le \|P\|\|EP\|_F + 3\frac{\|q - \nabla_x^2 L(w^*)s\|}{\|s\|}.$$

By mean value theorem, we have

$$\begin{aligned} \|q - \nabla_x^2 L(w^*)s\| &\le \int_0^1 \|(\nabla_x^2 L(x+ts, y_+, z_+) - \nabla_x^2 L(x^*, y^*, z^*))s\| dt \\ &\le \int_0^1 \left\| (\nabla r_0(x+ts, y_+, z_+) - \nabla r_0(x^*, y^*, z^*)) \begin{pmatrix} s \\ 0 \\ 0 \end{pmatrix} \right\| dt \\ &\le \int_0^1 \|\nabla r_0(x+ts, y_+, z_+) - \nabla r_0(x^*, y^*, z^*)\| dt \|s\| \\ &\le \xi \|s\| \int_0^1 \left\| \begin{pmatrix} x+ts-x^* \\ y_+ - y^* \\ z_+ - z^* \end{pmatrix} \right\| dt \\ &\le \xi \|s\| \int_0^1 ((1-t)\|w - w^*\| + (1+t)\|w_+ - w^*\|) dt \\ &\le 2\xi \|s\| \sigma. \end{aligned}$$

If $E = 0$, then the bounded deterioration clearly holds.

Next we consider the case of $E \neq 0$. Since

$$\|P\| = 1$$

and

$$\begin{aligned} \left\| E\left(I - \frac{ss^t}{s^t s}\right) \right\|_F^2 &= \left(\|E\|_F - \frac{\|Es\|^2}{2\|E\|_F\|s\|^2}\right)^2 - \left(\frac{\|Es\|^2}{2\|E\|_F\|s\|^2}\right)^2 \\ &\le \left(\|E\|_F - \frac{\|Es\|^2}{2\|E\|_F\|s\|^2}\right)^2, \end{aligned}$$

we have

$$\|P\|\|EP\|_F \le \|E\|_F - \frac{\|Es\|^2}{2\|E\|_F\|s\|^2}.$$

Thus we obtain

$$\begin{aligned}\|G_+ - \nabla_x^2 L(w^*)\|_F &\le \|G - \nabla_x^2 L(w^*)\|_F - \frac{\|(G - \nabla_x^2 L(w^*))s\|^2}{2\|G - \nabla_x^2 L(w^*)\|_F\|s\|^2} + 6\xi\sigma \\ &\le \|G - \nabla_x^2 L(w^*)\|_F + 6\xi\sigma.\end{aligned}$$

This implies the bounded deterioration property.
(2) Define

$$\eta_k = \|G_k - \nabla_x^2 L(w^*)\|_F \qquad \text{and} \qquad \psi_k = \frac{\|(G_k - \nabla_x^2 L(w^*))s_k\|}{\|s_k\|}.$$

Then, for the case of $E_k \neq 0$, the result of (1) yields

$$\frac{\psi_k^2}{2\eta_k} + \eta_{k+1} \le \eta_k + 6\xi\sigma_k.$$

Since the bounded deterioration property guarantees that there exists a positive constant $\bar{\eta}$ such that $\eta_k \le \bar{\eta}$ for any $k \ge 0$, we have

$$\frac{\psi_k^2}{2\bar{\eta}} + \eta_{k+1} \le \eta_k + 6\xi\sigma_k.$$

On the other hand, for $E = 0$, the preceding result follows directly from the bounded deterioration property. Summing both sides from $k = 0$ to $k = N$, we have

$$\frac{1}{2\bar{\eta}}\sum_{k=0}^{N}\psi_k^2 \le \frac{1}{2\bar{\eta}}\sum_{k=0}^{N}\psi_k^2 + \eta_{N+1} \le \|G_0 - \nabla_x^2 L(w^*)\|_F + 6\xi\sum_{k=0}^{N}\sigma_k.$$

The linear convergence of the sequence $\{w_k\}$ to w^* with

$$\|w_{k+1} - w^*\| \le \nu\|w_k - w^*\|, \quad \text{where} \quad 0 < \nu < 1$$

yields

$$\sigma_k = \|w_k - w^*\| \le \nu^k\|w_0 - w^*\|.$$

Then we have

$$\begin{aligned}\frac{1}{2\bar{\eta}}\sum_{k=0}^{N}\psi_k^2 &\le \|G_0 - \nabla_x^2 L(w^*)\|_F + 6\xi\|w_0 - w^*\|\sum_{k=0}^{N}\nu^k \\ &\le \|G_0 - \nabla_x^2 L(w^*)\|_F + 6\xi\|w_0 - w^*\|\frac{1}{1-\nu}.\end{aligned}$$

Thus $\psi_k \to 0$ and we obtain

$$\begin{aligned}\lim_{k\to\infty}\frac{\|(G_k - \nabla_x^2 L(w^*))s_k\|}{\|w_{k+1} - w_k\|} &\le \lim_{k\to\infty}\frac{\|(G_k - \nabla_x^2 L(w^*))s_k\|}{\|s_k\|} \\ &= \lim_{k\to\infty}\psi_k = 0.\end{aligned}$$

Therefore, condition (3.15) is satisfied by the PSB update (3.21). □

References

[1] Boggs, P.T., Tolle, J.W. and Wang, P.: On the local convergence of quasi-Newton methods for constrained optimization, *SIAM J. on Control and Optimization*, Vol.20 (1982), pp.161-171.

[2] Broyden, C.G., Dennis, Jr., J.E. and Moré, J.J.: On the local and superlinear convergence of quasi-Newton methods, *Journal of Institute of Mathematics and its Applications*, Vol.12 (1973), pp.223-245.

[3] Coleman, T.F.: On characterizations of superlinear convergence for constrained optimization, in *Computational Solution of Nonlinear Systems of Equations*, E.L.Allgower and K.Georg, eds., Lectures in Applied Mathematics, 26, American Mathematical Society, Rhode Island, 1990, pp.113-133.

[4] Dembo, R.S. and Steihaug, T.: Truncated-Newton algorithms for large-scale unconstrained optimization, *Mathematical Programming*, Vol.26 (1983), pp.190-212.

[5] Dembo, R.S., Eisenstat, S.C. and Steihaug, T.: Inexact Newton methods, *SIAM J. on Numerical Analysis*, Vol.19 (1982), pp.400-408.

[6] Dennis, Jr., J.E. and Moré, J.J.: A characterization of superlinear convergence and its application to quasi-Newton methods, *Mathematics of Computation*, Vol.28 (1974), pp.549-560.

[7] Eisenstat, S.C. and Steihaug, T.: *Local analysis of inexact quasi-Newton methods*, TR82-7, Dept. of Mathematical Sciences, Rice University, Texas, USA, 1982.

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

[9] Engels, J.R. and Martínez, H.J.: Local and superlinear convergence for partially known quasi-Newton methods, *SIAM J. on Optimization*, Vol.1 (1991), pp.42-56.

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

[11] Fletcher, R.: *Practical Methods of Optimization*, Second Edition, John Wiley and Sons, New York, 1987.

[12] Fontecilla, R.: Inexact secant methods for nonlinear constrained optimization, *SIAM J. on Numerical Analysis*, Vol.27 (1990), pp.154-165.

[13] Goodman, J.: Newton's method for constrained optimization, *Mathematical Programming*, Vol.33 (1985), pp.162-171.

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

[15] Kojima, M., Mizuno, S. and Yoshise, A.: A primal-dual interior-point algorithm for linear programming, in *Progress in Mathematical Programming, Interior-Point and Related Methods*, N.Megiddo ed., Springer-Verlag, New York, 1989, pp.29-47.

[16] Lasdon, L.S., Plummer, J. and Yu, G.: *Primal-dual and primal interior point algorithms for general nonlinear programs*, Manuscript, MSIS Department, The University of Texas at Austin, Texas, USA, February 1993.

[17] Lustig, I.J., Marsten, R.E. and Shanno, D.F.: Computational experience with a primal-dual interior point method for linear programming, *Linear Algebra and its Applications*, Vol.152 (1991), pp.191-222.

[18] Martinez, H.J., Parada, Z. and Tapia, R.A.: 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.

[19] McCormick, G.P. and Falk, J.E.: *The superlinear convergence of a nonlinear primal-dual algorithm*, Report GWU/OR/Serial T-550/91, Dept. of Operations Research, The George Washington University, Washington DC, USA, July 1991.

[20] McShane, K.A., Monma, C.L. and Shanno, D.F.: An implementation of a primal-dual interior point method for linear programming, *ORSA Journal on Computing*, Vol.1 (1989), pp.70-83.

[21] Nocedal, J. and Overton, M.L.: Projected Hessian updating algorithms for nonlinearly constrained optimization, *SIAM J. on Numerical Analysis*, Vol.22 (1985), pp.821-850.

[22] Steihaug, T.: *Damped inexact quasi-Newton methods*, TR81-3, Dept. of Mathematical Sciences, Rice University, Texas, USA, 1981 (revised May 1984).

[23] Vial, J.-P.: Computational experience with a primal-dual interior-point method for smooth convex programming, *Optimization Methods and Software*, Vol.3 (1994), pp.285-310.

[24] Yabe, H. and Yamashita, H.: On the rate of convergence of primal-dual interior point methods based on quasi-Newton methods, in *The State of the Art of Scientific Computing and its Prospect*, Research Report (RIMS Kokyuroku) No.880, Research Institute for Mathematical Sciences, Kyoto University, Kyoto, Japan, *in Japanese*, July 1994, pp.211-219.

[25] Yamashita, H.: *A globally convergent primal-dual interior point method for constrained optimization*, Technical Report, Mathematical Systems, Inc., Tokyo, Japan, April 1992 (revised March 1994).

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

Hiroshi YABE
Department of Applied Mathematics,
Faculty of Science,
Science University of Tokyo,
1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162, Japan
e-mail: yabe@am.kagu.sut.ac.jp

Hiroshi YAMASHITA
Mathematical Systems, Inc.,
2-4-3, Shinjuku, Shinjuku-ku, Tokyo, 160, Japan
e-mail: hy@msi.co.jp