Appl. Math. J. Chinese Univ.
2011, 26(1): 57-69

Projected gradient trust-region method for solving nonlinear systems with convex constraints

JIA Chun-xia[1] ZHU De-tong[2]

Abstract. In this paper, a projected gradient trust region algorithm for solving nonlinear equality systems with convex constraints is considered. The global convergence results are developed in a very general setting of computing trial directions by this method combining with the line search technique. Close to the solution set this method is locally Q-superlinearly convergent under an error bound assumption which is much weaker than the standard nonsingularity condition.

§1 Introduction

In this paper we consider the nonlinear systems subject to the convex constraints on variable

$$F(x) = 0, \qquad x \in \Omega, \tag{1}$$

where $F : \mathcal{X} \to \mathbb{R}^n$ is continuously differentiable and $\mathcal{X} \subseteq \mathbb{R}^n$ is an open set containing a nonempty closed convex set Ω.

There are quite a few articles using projected gradient methods to solve the convex constrained optimization problem [1, 2]. When the constrained conditions of the problem are geometrically simple or defined by a set of linear equalities and inequalities, projected gradient, conditional gradient and Newton-like methods are explored extensively in connection with two basic approaches, namely, the line search and the trust region. The classical Levenberg-Marquardt method for nonlinear systems (1) computes the trial step by

$$d_k = -(J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T F(x_k), \tag{2}$$

where $J(x_k) = F'(x_k)$ is the Jacobian and $\mu_k \geq 0$ is a parameter updated from one iteration to the next. Levenberg-Marquardt step (2) is a modification of Gauss-Newton's step

$$d_k^{\mathrm{GN}} = -(J(x_k)^T J(x_k))^{-1} J(x_k)^T F(x_k).$$

Received: 2008-01-09.
MR Subject Classification: 90C30, 49M39.
Keywords: Nonlinear equation, trust region method, projected gradient, local error bound.
Digital Object Identifier(DOI): 10.1007/s11766-011-1956-7.
Supported by the National Natural Science Foundation of China (10871130), the Research Fund for the Doctoral Program of Higher Education of China (20093127110005), and the Scientific Computing Key Laboratory of Shanghai Universities.

The parameter μ_k is used to prevent d_k from being too large when $J(x_k)^T J(x_k)$ is nearly singular. The Gauss-Newton's step is undefined when $J(x_k)^T J(x_k)$ is singular. A positive μ_k guarantees that (2) is well defined. The Levenberg-Marquardt method has a quadratic rate of convergence if the Jacobian at the solution x^* is nonsingular and if μ_k is chosen suitably at each step. However, the condition of the nonsingularity of $J(x^*)$ is too strong. Recently, Yamashita and Fukushima [8] showed that under a weaker condition that $\|F(x_k)\|$ provides a local error bound near the solution, the Levenberg-Marquardt method still has a quadratic convergence if $\mu_k = \|F(x_k)\|^2$. Trust region method [2, 7] is a well-accepted technique in the constrained optimization to assure global convergence. In the traditional trust region method, it is possible that the trust region subproblem needs to be resolved many times with a reduced trust region radius before obtaining an acceptable solution satisfying some criterions. Hence the total computation for completing one iteration might be expensive. In order to overcome this shortcoming, line search technique is often used [10].

Stimulated by the progress in these aspects, we present the projected gradient trust-region method with line search technique to solve (1). The new proposed algorithm is locally Q-superlinearly convergent under a weaker assumption which in particular allows the solution set to be (locally) nonunique. To this end, we replace the nonsingularity assumption by an error bound condition, which is much weaker than the standard nonsingularity condition.

The organization of this paper is as follows. In section 2, we describe our modified algorithm which combines the techniques of the trust region method, the projected gradient method and the line search technique. In Section 3, we prove the global convergence of this algorithm. We establish other convergence properties such as the order of local convergence of the method with a local error bound in Section 4.

§2 Algorithm

This section describes and investigates the projected gradient trust-region method for solving a convex constrained minimization reformulated by (1) under a weaker assumption.

Definition 2.1. Let the solution set X^* of problem (1) be nonempty. We say that $\|F(x)\|$ *provides a local error bound* on a neighborhood $\mathcal{B}_\delta$ if there exist constants $\delta > 0$ and $\tau > 0$ such that

$$\tau \ \mathrm{dist}(x, X^*) \leq \|F(x)\|$$

for $x \in \mathcal{B}_\delta(x^*) \cap \Omega$, where $\mathcal{B}_\delta(x) = \{y \in \Re^n \mid \|y - x\| \leq \delta\}$ is the closed ball centered at x with radius δ and $\mathrm{dist}(y, X^*) = \inf\{\|y - x\| \mid x \in X^*\}$ is the distance from a point y to the solution set X^*.

For solving (1) we usually consider the related optimization problem

$$\min f(x) = \frac{1}{2}\|F(x)\|^2 \qquad \text{s.t.} \ \ x \in \Omega, \tag{3}$$

where $f(x)$ is the natural merit function corresponding to the mapping F. Similar to (2), we can get d_k by solving a regularized problem

$$\min q_k(d) = \frac{1}{2}\|F(x_k) + H_k d\|^2 + \mu_k \|d\|^2 \qquad \text{s.t.} \ \ x_k + d \in \Omega, \tag{4}$$

where $H_k \in \Re^{n\times n}$ is an approximation to the Jacobian $F'(x_k)$ (not necessarily existing), $\mu_k = \min\{1, \mu\|F(x_k)\|\}$ and μ is a positive constant. A positive μ_k guarantees that the solution d_k of the subproblem (4) always uniquely exists.

We first introduce some standard notations for this paper. Given an inner product norm $\|\cdot\|$, the projection into a nonempty closed convex set Ω is the mapping $P:\mathbb{R}^n \to \Omega$ defined by

$$P_\Omega(x) = \mathrm{argmin}\{\ \|z-x\| \mid z\in\Omega\}.$$

The dependence of P on Ω is usually clear from the context, but if there is possibility of confusion we shall use $P_\Omega(x)$ to denote the projection of x into Ω.

The analysis of the algorithm requires some basic properties of the projection operator.

Lemma 2.1. *Let $\Omega \in \mathbb{R}^n$ be a nonempty closed convex set and let P be the projection into Ω.*

(1) *If $z\in\Omega$, then $\langle P(x)-x, z-P(x)\rangle \ge 0$ for $x\in\mathbb{R}^n$.*

(2) *$\langle P(y)-P(x), y-x\rangle \ge 0$ for $x,y\in\mathbb{R}^n$. If $P(y)\ne P(x)$, then strict inequality holds.*

(3) *$\|P(y)-P(x)\| \le \|y-x\|$ for $x,y\in\mathbb{R}^n$.*

See [9] for the proof of Lemma 2.1 and additional information on projections. If we take $z = x_k\in\Omega$ and $x = x_k - \xi\nabla f(x_k)\in\Re^n$, then we can conclude from Lemma 2.1(1) that

$$\langle \nabla f(x_k),\ x_k - P(x_k-\xi\nabla f(x_k))\rangle \ge \frac{\|P(x_k-\xi\nabla f(x_k)) - x_k\|^2}{\xi}, \tag{5}$$

for $\xi > 0$. The convergence analysis of the projected gradient trust-region method presented in the following section also requires the following properties of the projection operator.

Lemma 2.2. *If P is the projection into Ω, then the function ϕ_1 defined by*

$$\phi_1(\alpha) = \|P(x+\alpha d) - x\|$$

for $\alpha > 0$ is isotone (nondecreasing) for $x\in\mathbb{R}^n$ and $d\in\mathbb{R}^n$, and the function ϕ_2 defined by

$$\phi_2(\alpha) = \frac{\|P(x+\alpha d)-x\|}{\alpha}$$

for $\alpha>0$ is antitone (nonincreasing) for $x\in\mathbb{R}^n$ and $d\in\mathbb{R}^n$.

The isonicity of ϕ_1 is established by Toint [7], and the result on ϕ_2 is due to Gafni and Bertsekas [4]. Moré and Calamai [5] provided an alternate proof of this last result.

The definition of the projected gradient requires some notions from convex analysis. A direction v is feasible at $x\in\Omega$ if $x+\tau v$ belongs to Ω for $\tau>0$ sufficient small. The *tangent cone* $T(x)$ is defined as the closure of the cone of all feasible directions. The *projection gradient* $\nabla_\Omega f$ of f is defined by

$$\nabla_\Omega f(x) = \arg\min\{\|v+\nabla f(x)\| : v\in T(x)\}. \tag{6}$$

The proof of the following Lemma can be found in [5].

Lemma 2.3. *Let $\nabla_\Omega f(x)$ be the projected gradient of f at $x\in\Omega$.*

(1) $-\langle\nabla f(x), \nabla_\Omega f(x)\rangle = \|\nabla_\Omega f(x)\|^2$.

(2) $\min\{\langle \nabla f(x), v\rangle : \ v \in T(x), \|v\| \le 1\} = -\|\nabla_\Omega f(x_k)\|$.

(3) *The point* $x \in \Omega$ *is a stationary point of problem (3) if and only if* $\nabla_\Omega f(x) = 0$.

We are now going to describe more precisely the strategy we propose in order to choose, at the k-th iteration, a candidate for the $(k+1)$-th iteration. Our model of the objective function has the form of problem (4). To meet the need of the trust region strategy, we adopt the suggestion of Toint [7] or a linearly approximation of $d_k(\xi)$ (see (7) below) and choose a step d_k which gives as much reduction in the model $q_k(d)$ as one step of the gradient projection method applied to the subproblem

$$\min\{q_k(d) \mid x_k + d \in \Omega, \|d\| \le \mu_2 \Delta_k\}.$$

The step path of the gradient projection algorithm is given by

$$d_k(\xi_k) = P(x_k - \xi_k \nabla f(x_k)) - x_k. \tag{7}$$

Now we turn from the above subproblem to solve a quadratic subproblem

$$(S_k) \quad \begin{array}{ll} \min & q_k(d_k(\xi)) \\ \text{s.t.} & \|d_k(\xi)\| \le \mu_2 \Delta_k \end{array}$$

with μ_2 a positive constant. Let ξ_k be the optimal solution of the subproblem (S_k). Following Toint [7] (also see [2]) instead of solving the subproblem (S_k), the prepared step generated by the gradient projection algorithm is of the form $d_k(\xi_k)$, where $d_k(\xi)$ is given in (7) and ξ_k satisfies the following two requirements. Given constants $\mu_0, \ \mu_1, \ \mu_2$ with

$$0 < \mu_0 < \frac{1}{2} \quad \text{and} \quad 0 < \mu_1 < \mu_2,$$

the first requirement is that

$$q_k(d_k(\xi_k)) - q_k(0) \le \mu_0 \langle \nabla f(x_k), d_k(\xi_k)\rangle \quad \text{and} \quad \|d_k(\xi_k)\| \le \mu_2 \Delta_k, \tag{8}$$

while the second requirement is that there are positive constants γ_1 and γ_2 such that

$$\xi_k \ge \gamma_1 \quad \text{or} \quad \xi_k \ge \gamma_2 \bar{\xi}_k, \tag{9}$$

where $\bar{\xi}_k > 0$ satisfies

$$q_k(d_k(\bar{\xi}_k)) - q_k(0) \ge (1-\mu_0)\langle \nabla f(x_k), d_k(\bar{\xi}_k)\rangle \quad \text{or} \quad \|d_k(\bar{\xi}_k)\| \ge \mu_1 \Delta_k.$$

Similar to [2], we choose a step $d_k = d_k(\hat{\xi}_k)$, where $\gamma_3 > 0$ is a constant and $\hat{\xi}_k = \min\{\xi_k, \gamma_3\}$. If we take the *Cauchy point* x_k^{C} as

$$x_k^{\mathrm{C}} = P(x_k - \hat{\xi}_k \nabla f(x_k)) = x_k + d_k(\hat{\xi}_k),$$

then (5) implies that

$$\langle \nabla f(x_k), \ x_k - x_k^{\mathrm{C}}\rangle \ge \frac{\|x_k^{\mathrm{C}} - x_k\|^2}{\hat{\xi}_k}$$

for $\hat{\xi}_k > 0$. Hence

$$g_k^T d_k \le -\frac{\|x_k^{\mathrm{C}} - x_k\|^2}{\hat{\xi}_k} = -\frac{\|d_k\|^2}{\hat{\xi}_k}, \tag{10}$$

where $g_k = \nabla f(x_k) = F(x_k)^T F'(x_k)$ and $d_k = d_k(\hat{\xi}_k)$.

Now, we describe the projected gradient trust-region (PGTRA) algorithm with line search technique.

Initialization step

Choose a parameter $\varepsilon > 0$. Choose $\beta \in (0, \frac{1}{2})$, $\Delta_{\max} \geq 0$, $x_0 \in \Omega$, $\Delta_0 < \Delta_{\max}$, and a nonsingular matrix $H_0 \in \mathbb{R}^{m\times n}$. Set $k = 0$ and go to the main step.

Main step

1. Compute $\nabla_\Omega f(x_k)$.

2. If $\|\nabla_\Omega f(x_k)\| \leq \varepsilon$, stop with the approximate solution x_k.

3. Approximately solve the subproblem and obtain the step $d_k(\xi_k)$ which is required to satisfy the two requirements (8) and (9). In particular, we choose $d_k = d_k(\hat{\xi}_k)$ in this paper.

4. Choose $\alpha_k = 1, \omega, \omega^2, \ldots$, until the following inequality is satisfied
$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta g_k^T d_k. \tag{11}$$
Set $x_{k+1} = x_k + \alpha_k d_k$.

5. Calculate $\text{pred}_k = q_k(0) - q_k(\alpha_k d_k)$, $\text{ared}_k = f(x_k) - f(x_k + \alpha_k d_k)$, and $\rho_k = \frac{\text{ared}_k}{\text{pred}_k}$. Updating trust region size Δ_{k+1} from Δ_k
$$\Delta_{k+1} \in \begin{cases} [\gamma_1\Delta_k, \gamma_2\Delta_k], & \text{if } \rho_k \leq \eta_1, \\ (\gamma_2\Delta_k, \Delta_k], & \text{if } \eta_1 < \rho_k < \eta_2, \\ (\Delta_k, \min\{\gamma_3\Delta_k, \Delta_{\max}\}], & \text{if } \rho_k \geq \eta_2. \end{cases}$$

6. Update H_k to obtain H_{k+1}, then set $k \leftarrow k+1$ and go to step 1.

Remark. It is easy to see that the nonmonotonic line search technique
$$f(x_k + \alpha_k d_k) \leq f(x_l(k)) + \alpha_k \beta g_k^T d_k$$
can be used in (11), where $f(x_{l(k)}) = \max\limits_{0\leq j\leq m(k)} f(x_{k-j}), m(k) = \min\{m(k-1), M\}$, $m(0) = 0$, and M is a positive integer. We can also prove the global convergence and local convergent rate of the proposed algorithm with nonmonotonic technique.

§3 Global convergence analysis

Throughout this section we assume that $F : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. Given $x_0 \in \text{int}(\Omega) \subset \mathbb{R}^n$, the algorithm generates a sequence $\{x_k\} \subset \Omega \subseteq \mathbb{R}^n$. In our analysis, we denote the level set of f by
$$\mathcal{L}(x_0) = \{\ x \in \Omega \mid f(x) \leq f(x_0),\ x_0 \in \Omega\ \}.$$
The following assumptions are commonly used in convergence analysis of most methods for the convex constrained systems.

Assumption 1. The sequence $\{x_k\}$ generated by the algorithm is contained in a compact set $\mathcal{L}(x_0)$ on $\mathbb{R}^n$.

By Assumption 1 and the continuity of $F'(x)$, there exists a positive constant $\chi_g > 0$ such that $\|F'(x_k)\| \leq \chi_g$ for all k.

Assumption 2. There exists a constant ϵ_0 with $0 < \epsilon_0 \leq \frac{1}{\chi_g}$ such that the updating matrix $H_k \in \mathbb{R}^{n\times n}$ is an approximation to the Jacobian $F'(x_k)$, that is,

$$\|F'(x_k) - H_k\| \leq \epsilon_0 \min\{\|d_k\|, \|d_k(\bar{\xi}_k)\|\}.$$

Furthermore, we have $\|H_k\| \leq \|F'(x_k)\| + \|F'(x_k) - H_k\| \leq \chi_g + \epsilon_0\|d_k\| \leq \chi_g + \epsilon_0\mu_2\Delta_{\max} \triangleq \chi_H$.

Define

$$w_k(d) = \frac{q_k(d) - q_k(0) - \langle\nabla q_k(0), d\rangle}{\|d\|^2}$$

and

$$\beta_k = 1 + \max\{\ |\omega_k(d_k(\bar{\xi}_k))|,\ |\omega_k(d_k)|\ \}.$$

We can deduce the following lemma from Assumption 2 and (5).

Lemma 3.1. *If $\bar{\xi}_k$ satisfies*

$$q_k(d_k(\xi)) - q_k(0) \geq \mu_0\langle\nabla f(x_k), d_k(\xi)\rangle,$$

then $\omega_k(d_k(\bar{\xi}_k)) + \epsilon_0\chi_g$ is positive and

$$\bar{\xi}_k \geq \frac{1-\mu_0}{\beta_k}.$$

The following lemma shows the relation between the gradient $\nabla f(x_k)$ of the function and $d_k(\xi_k)$ generated in the subproblem (S_k). We can see that the direction of the trial step generated in the trust region subproblem is a sufficiently descent direction. By Lemma 3.1, the following lemma can be similarly proved as in [6].

Lemma 3.2. *If ξ_k satisfies inequalities (8) and (9), then there is a constant $\mu_3 > 0$ such that*

$$-\langle\nabla f(x_k), d_k(\xi_k)\rangle \geq \mu_3[\frac{\|d_k(\xi_k)\|}{\xi_k}]\min\{\Delta_k, \frac{1}{\beta_k}[\frac{\|d_k(\xi_k)\|}{\xi_k}]\}. \tag{12}$$

Note that if $\hat{\xi}_k = \xi_k$, then (12) holds for $\xi_k = \hat{\xi}_k$; and if $\hat{\xi}_k = \gamma_3$, then we have

$$-\langle\nabla f(x_k), d_k(\hat{\xi}_k)\rangle \geq [\frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k}]^2\gamma_3. \tag{13}$$

Since $\|H_k\| \leq \chi_H$, we have the estimate

$$\omega_k(d_k(\xi)) = \frac{\frac{1}{2}\|H_k d_k\|^2 + \mu_k\|d_k\|^2}{\|d_k\|^2} \leq \frac{1}{2}\chi_H^2 + 1,$$

which will be used in our following convergence analysis. We have $1 \leq \beta_k \leq \chi_H^2 + 2 \triangleq b$.

Lemma 3.3. *Assume that Assumptions 1 and 2 hold and $\nabla f(\cdot)$ is Lipschitz continuous with Lipschitz constant L. If there exists $\epsilon > 0$ such that*

$$\frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k} \geq \epsilon$$

for all k, then there is $\hat{\tau} > 0$ such that

$$\Delta_k \geq \hat{\tau} \tag{14}$$

for all k and

$$\hat{\tau} = \min\{\frac{\epsilon\gamma_3}{\mu_2},\ \frac{\epsilon\gamma_1}{b},\ \frac{\mu_3\epsilon(1-\beta)\gamma_1}{L\mu_2^2},\ \frac{(1-\eta_2)\mu_0\mu_3\epsilon\gamma_1}{L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b}}\}.$$

Proof. If $\xi_k \geq \gamma_3$, then $\hat{\xi}_k = \gamma_3$. Since ϕ_1 is increasing, we have

$$\|P(x_k - \gamma_3\nabla f(x_k)) - x_k\| \leq \|P(x_k - \xi_k\nabla f(x_k)) - x_k\| \leq \mu_2\Delta_k.$$

Therefore,
$$\Delta_k \geq \frac{1}{\mu_2} \frac{\|P(x_k - \gamma_3 \nabla f(x_k)) - x_k\|}{\gamma_3} \gamma_3 \geq \frac{\epsilon\gamma_3}{\mu_2}$$
and hence (14) holds.

If $\xi_k < \gamma_3$, then $\hat{\xi}_k = \xi_k$. We first prove that if
$$\Delta_k \leq \min\{\frac{\epsilon}{b}, \frac{\mu_3\epsilon(1-\beta)}{L\mu_2^2}\}, \tag{15}$$
then $\alpha_k = 1$ must satisfy the inequality (11) in step 4, i.e.,
$$f(x_k + d_k) \leq f(x_k) + \beta g_k^T d_k.$$
If the above inequality is not true, then
$$f(x_k + d_k) > f(x_k) + \beta g_k^T d_k.$$
Since
$$f(x_k + d_k) = f(x_k) + \nabla f(x_k + \eta_k d_k)^T d_k,$$
where $\eta_k \in (0, 1)$ and $\nabla f(\cdot)$ is Lipschitz continuous with Lipschitz constant L, we have
$$L\|d_k\|^2 + (1-\beta)\nabla f(x_k)^T d_k \geq 0.$$
By (8) and (12), we obtain $L\mu_2^2\Delta_k^2 - \mu_3\epsilon(1-\beta)\min\{\Delta_k, \frac{\epsilon}{b}\} > 0$. Since $\Delta_k < \frac{\epsilon}{b}$, we have $L\mu_2^2\Delta_k^2 - \mu_3\epsilon(1-\beta)\Delta_k > 0$. Then $\Delta_k > \frac{\mu_3\epsilon(1-\beta)}{L\mu_2^2}$, which contradicts (15).

From the above we see that if (15) holds, then the step size $\alpha_k = 1$ and hence $x_{k+1} = x_k + d_k(\xi_k)$. We now assume that there is a number k such that
$$\Delta_k < \min\{\frac{\epsilon\gamma_1}{b}, \frac{\mu_3\epsilon(1-\beta)\gamma_1}{L\mu_2^2}, \frac{(1-\eta_2)\mu_0\mu_3\epsilon\gamma}{L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b}}\}, \tag{16}$$
from which we will derive a contradiction.

Let t be the first iteration number such that the inequality (16) holds. When
$$\Delta_{t-1} \leq \frac{\Delta_t}{\gamma_1} \leq \min\{\frac{\epsilon}{b}, \frac{\mu_3\epsilon(1-\beta)}{L\mu_2^2}\},$$
(15) holds for $k = t-1$. Hence $\alpha_{t-1} = 1$. From (8) and (12), we have
$$\mathrm{Pred}_{t-1} \geq -\mu_0\langle\nabla f(x_{t-1}), d_{t-1}(\xi_{t-1})\rangle \geq \mu_0\mu_3\epsilon\min\{\Delta_{t-1}, \frac{\epsilon}{b}\} = \mu_0\mu_3\epsilon\Delta_{t-1}. \tag{17}$$
From Assumptions 1 and 2 and that $\nabla f(\cdot)$ being Lipschitz continuous, we have
$$|f(x_k + d_k) - f(x_k) + q_k(0) - q_k(d_k))| \leq (L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b})\|d_k\|^2 \tag{18}$$
with $\eta_k \in (0, 1)$. Combining (17) and (18), we have
$$\begin{aligned} |\rho_{t-1} - 1| &= \frac{|f(x_{t-1} + d_{t-1}) - f(x_{t-1}) + \mathrm{Pred}_{t-1}|}{|\mathrm{Pred}_{t-1}|} \leq \frac{(L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b})\mu_2^2\Delta_{t-1}^2}{\mu_0\mu_3\epsilon\Delta_{t-1}} \\ &\leq \frac{(L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b})}{\mu_0\mu_3\epsilon}\frac{\Delta_t}{\gamma_1} \leq 1 - \eta_2. \end{aligned}$$
This implies that $\rho_{t-1} \geq \eta_2$. By the updating rule for the trust region radius Δ_k in step 5, we have
$$\Delta_{t-1} \leq \Delta_t < \min\{\frac{\epsilon\gamma_1}{b}, \frac{\mu_3\epsilon(1-\beta)\gamma_1}{L\mu_2^2}, \frac{(1-\eta_2)\mu_0\mu_3\epsilon\gamma}{L + \sqrt{2f(x_0)}\epsilon_0 + \hat{b}}\}.$$
This contradicts the assumption that t is the first index with (16) holding. □

We are ready to state one of our main results.

Theorem 3.1. *Let Assumptions 1 and 2 hold and $\nabla f(\cdot)$ be Lipschitz continuous with Lipschitz constant L. Let $\{x_k\} \subset \Omega$ be a sequence generated by the algorithm. Then*

$$\liminf_{k\to+\infty} \frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k} = 0.$$

Proof. If the conclusion is not true, then there exists $\epsilon > 0$ such that

$$\frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k} \geq \epsilon.$$

According to the acceptance rule in step 4, we have

$$f(x_{k+1}) \leq f(x_k) + \alpha_k \beta g_k^T d_k \leq f(x_k) - \alpha_k \beta \mu_3 \epsilon \min\{\Delta_k, \frac{\epsilon}{b}\}.$$

Since $\{f(x_k)\}$ is convergent, we have $\lim\limits_{k\to\infty} \alpha_k \Delta_k = 0$. By Lemma 3.3, $\Delta_k \geq \hat{\tau}$, so we get

$$\lim_{k\to\infty} \alpha_k = 0.$$

The acceptance rule of α_k means that

$$\frac{\alpha_k}{\omega}\langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle + o(\frac{\alpha_k}{\omega}\|d_k(\hat{\xi}_k)\|) = f(x_k + \frac{\alpha_k}{\omega} d_k(\hat{\xi}_k)) - f(x_k) > \beta \frac{\alpha_k}{\omega} g_k^T d_k(\hat{\xi}_k),$$

which implies

$$(1-\beta)\frac{\alpha_k}{\omega}\langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle + o(\frac{\alpha_k}{\omega}\|d_k(\hat{\xi}_k)\|) \geq 0. \tag{19}$$

Dividing (19) by $\frac{\alpha_k}{\omega}\|d_k(\hat{\xi}_k)\|$ and noting that $1-\beta > 0$ and $\langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle \leq 0$, we obtain

$$\lim_{k\to\infty} \frac{\langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle}{\|d_k(\hat{\xi}_k)\|} = 0.$$

From $\|d_k(\hat{\xi}_k)\| \leq \max\{\mu_2 \Delta_{\max}, d_k(\gamma_3)\}$ and $\|d_k(\gamma_3)\| = \|P(x_k - \gamma_3 \nabla f(x_k)) - P(x_k)\| \leq \gamma_3 \sqrt{2f(x_0)}\chi_g$, we have that

$$\lim_{k\to\infty} \langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle = 0.$$

Combing the above formula with $\langle \nabla f(x_k), d_k(\hat{\xi}_k)\rangle \leq -\mu_3 \epsilon \min\{\Delta_k, \frac{\epsilon}{b}\} \leq 0$, we can get

$$\lim_{k\to\infty} \Delta_k = 0,$$

which contradicts Lemma 3.3. □

Theorem 3.2. *Let Assumptions 1 and 2 hold and $\nabla f(\cdot)$ be Lipschitz continuous with Lipschitz constant L. Let $\{x_k\} \subseteq \Omega$ be a sequence generated by the algorithm. Then*

$$\lim_{k\to+\infty} \frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k} = 0.$$

Proof. Assume that there is an $\epsilon_1 \in (0,1)$. Let $\frac{\|d_{m_i}(\hat{\xi}_{m_i})\|}{\hat{\xi}_{m_i}}$ be a subsequence such that

$$\frac{\|d_{m_i}(\hat{\xi}_{m_i})\|}{\hat{\xi}_{m_i}} \geq \epsilon_1$$

for all m_i $(i = 1, 2, \dots)$. Theorem 3.1 guarantees the existence of another subsequence $\frac{\|d_{l_i}(\hat{\xi}_{l_i})\|}{\hat{\xi}_{l_i}}$ such that

$$\frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k} \geq \epsilon_2$$

for $m_i \le k < l_i$ and

$$\frac{\|d_{l_i}(\hat{\xi}_{l_i})\|}{\hat{\xi}_{l_i}} \le \epsilon_2$$

for some $\epsilon_2 \in (0, \epsilon_1)$. By the acceptance rule in step 4 and Lemma 3.2, we have

$$f(x_{k+1}) - f(x_k) \le \alpha_k \beta g_k^T d_k \le -\alpha_k \beta \mu_3 \epsilon_2 \min\{\Delta_k, \frac{\epsilon_2}{b}\}$$

for $m_i \le k < l_i$. Since $\{f(x_k)\}$ is convergent, we have that

$$\lim_{k\to\infty,\ m_i \le k < l_i} \alpha_k \Delta_k = 0.$$

Similar to the proof of Theorem 3.1, we have the result. □

The following result is an immediate consequence of Theorem 3.2 with the condition $0 \le \|d_k(\hat{\xi}_k)\| = \frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k}\hat{\xi}_k \le \frac{\|d_k(\hat{\xi}_k)\|}{\hat{\xi}_k}\gamma_3$.

Theorem 3.3. *Let Assumptions 1 and 2 hold, $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on Ω, $\nabla f(\cdot)$ be Lipschitz continuous with Lipschitz constant L, and $\{x_k\}$ be the sequence generated by the algorithm. Then*

$$\lim_{k\to\infty} \|d_k\| = 0.$$

Theorem 3.4. *Let Assumptions 1 and 2 hold, $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on Ω, $\nabla f(\cdot)$ be Lipschitz continuous with Lipschitz constant L, and $\{x_k\}$ be the sequence generated by the algorithm. If $\lim_{k\in K_1} x_k = x^*$, then $\lim_{k\in K_1} x_k^C = x^*$ and $\lim_{k\in K_1} \|\nabla_\Omega f(x_k^{\mathrm{C}})\| = 0$.*

The first part of Theorem 3.4 can be proved by $\|x_k^{\mathrm{C}} - x^*\| \le \|x_k^{\mathrm{C}} - x_k\| + \|x_k - x^*\|$ and Theorem 3.3; the second part can be found in [5].

§4 Properties of the local convergence

In this section, we discuss the convergence rate of the proposed algorithm. To establish the (local) convergence results, we change Assumption 2 to (20) and add another assumption. We also assume that d_k is the optimal solution of (4).

Assumption 2. There exist constants $\delta > 0$ and $\tau > 0$ such that

$$\|[H_k - F'(x_k)](x - x_k)\| \le \tau \|x - x_k\|^2 \tag{20}$$

for $x \in B_\delta(x^*) \cap \Omega$.

By (20), H_k is bounded and we still use $\|H_k\| \le \chi_H$ to express the boundedness of H_k.

Assumption 3. The solution set X^* of problem (1) is nonempty. For some solution $x^* \in X^*$, there exist constants $\delta > 0,\ \tau_1 > 0$ and $\tau_2 > 0$ such that

$$\tau_1 \mathrm{dist}(x, X^*) \le \|F(x)\|$$

and

$$\|F(x) - F(x_k) - F_k'(x - x_k)\| \le \tau_2 \|x - x_k\|^2$$

for $x, x_k \in B_\delta(x^*) \cap \Omega$.

Using the mean value theorem and the boundedness of $F'(x_k)$ on $B_\delta(x^*) \cap \Omega$, there is a constant $L_1 > 0$ such that

$$\|F(x) - F(y)\| \le L_1 \|x - y\| \tag{21}$$

for $x, y \in B_\delta(x^*) \cap \Omega$.

The constants $\tau,\ \delta,\ \tau_1,\ \tau_2$ and L_1 that appear in the subsequence analysis are always the constants from Assumptions 2, 3 and (21).

Lemma 4.1. *Assume that Assumptions 1 to 3 hold and $\mu_k = \min\{1, \mu\|F(x_k)\|\}$. There exist constants $\tau_3 > 0$ and $\tau_4 > 0$ such that the following inequalities hold for $x_k \in B_{\frac{\delta}{2}}(x^*) \cap \Omega$:*

(1) $\|d_k\| \leq \tau_3\ dist(x_k, X^*)$;

(2) $\|F(x_k) + H_k d_k\| \leq \tau_4\ dist(x_k, X^*)^{\frac{3}{2}}$.

Proof. The proof is similar to Lemma 2.3 in [3]. □

Lemma 4.2. *Assume that Assumptions 1 to 3 hold and $x_k \in B_{\frac{\delta}{2}}(x^*)$. Then the full stepsize*

$$\alpha_k = 1$$

is always accepted for k sufficiently large. Moreover, if $f(x)$ is twice continuously differentiable, then there exists $\hat{\Delta} > 0$ such that $\Delta_k \geq \hat{\Delta}$.

Proof. Lex $\bar{x}_k \in X^*, \|x_k - \bar{x}_k\| = \mathrm{dist}(x_k, X^*)$, and $d_k = \mathrm{argmin}\ q_k(d)$. Then

$$\begin{aligned} g_k^T d_k &= \frac{1}{2}\|F(x_k) + F'(x_k)d_k\|^2 - \frac{1}{2}\|F(x_k)\|^2 - \frac{1}{2}d_k^T F'(x_k)^T F'(x_k) d_k \\ &\leq \frac{1}{2}\|F(x_k) + F'(x_k)d_k\|^2 + \mu_k\|d_k\|^2 - \frac{1}{2}\|F(x_k)\|^2 \\ &\leq \frac{1}{2}\|F(x_k) - F(\bar{x}_k) + F'(x_k)(\bar{x}_k - x_k)\|^2 + \mu_k\|F(x_k)\| \cdot \|\bar{x}_k - x_k\|^2 - \frac{1}{2}\|F(x_k)\|^2. \end{aligned}$$

By $x_k \in B_{\frac{\delta}{2}}(x^*)$, we know $x_k, \bar{x}_k \in B_\delta(x^*)$. By Assumption 3 and inequalities $\|x_k - \bar{x}_k\| \leq \|x_k - x^*\| \leq \delta$ and $\delta \leq \min\{1, \frac{\tau_1^2}{2(\tau_2^2 + 2\mu L_1)}\}$, we have

$$\begin{aligned} g_k^T d_k &\leq \frac{1}{2}(\tau_2^2\|\bar{x}_k - x_k\|^4 + 2\mu L_1\|\bar{x}_k - x_k\|^3 - \tau_1^2\|\bar{x}_k - x_k\|^2) \\ &\leq \frac{1}{2}(\tau_2^2 + 2\mu L_1\delta - \tau_1^2)\|x_k - \bar{x}_k\|^2 \leq -\frac{\tau_1^2}{4}\mathrm{dist}(x_k, X^*)^2. \end{aligned}$$

From the acceptance rule of α_k and (10),

$$f(x_k) - f(x_{k+1}) \geq -\alpha_k \beta g_k^T d_k \geq \alpha_k \beta \frac{1}{\gamma_3}\|d_k\|^2 \geq 0. \tag{22}$$

This means that the sequence $\{f(x_k)\}$ is nonincreasing and hence convergent. Then from (22), we can get

$$\lim_{k\to\infty} \alpha_k \|d_k\|^2 = 0.$$

Therefore, we have either

$$\liminf_{k\to\infty} \alpha_k = 0 \tag{23}$$

or

$$\lim_{k\to\infty} \|d_k\|^2 = 0. \tag{24}$$

If (23) holds, then there exists a subset $K \subset \{k\}$ such that $\lim\limits_{k\in K} \alpha_k = 0$. From the acceptance rule of α_k, we have

$$f(x_k + \frac{\alpha_k}{\omega}d_k) > f(x_k) + \frac{\alpha_k}{\omega}\beta g_k^T d_k.$$

Hence,

$$\frac{\alpha_k}{\omega}(1-\beta)g_k^T d_k + o(\frac{\alpha_k}{\omega}\|d_k\|) \geq 0. \tag{25}$$

Dividing (25) by $\frac{\alpha_k}{\omega}\|d_k\|$ and noting that $g_k^T d_k \leq -\frac{1}{\gamma_3}\|d_k\|^2 < 0$, we can get $\lim_{k\in K}\frac{g_k^T d_k}{\|d_k\|} = 0$ and $\lim_{k\in K}\|d_k\| = 0$. Hence (24) holds. Since g_k is bounded and $|g_k^T d_k| = -g_k^T d_k \geq \frac{\tau_1^2}{4}\mathrm{dist}(x_k, X^*)^2 \geq 0$, we deduce that

$$\lim_{k\to\infty}\mathrm{dist}(x_k, X^*) = 0. \tag{26}$$

Since $d_k = \arg\min q_k(d) = \frac{1}{2}\|F(x_k) + H_k d\|^2 + \mu_k\|d\|^2$, we have $[F(x_k)^T H_k]^T + (H_k^T H_k + \mu_k I)d_k = 0$, which implies that

$$F(x_k)^T H_k d_k + d_k^T H_k^T H_k d_k = -\mu_k\|d_k\|^2. \tag{27}$$

By Assumption 3,

$$\mu\tau_1\mathrm{dist}(x_k, X^*) \leq \mu_k = \mu\|F(x_k)\| \leq \mu L_1\|x_k - \bar{x}_k\| = \mu L_1\mathrm{dist}(x_k, X^*). \tag{28}$$

Then by (26)-(28),

$$F(x_k)^T H_k d_k + d_k^T H_k^T H_k d_k = o(\|d_k\|^2). \tag{29}$$

By Assumption 2, we obtain

$$\begin{aligned} |\ F(x_k)^T[F'(x_k) - H_k]d_k\ | &= |\ \frac{1}{\alpha_k}F(x_k)^T[F'(x_k) - H_k](x_{k+1} - x_k)\ | \\ &\leq \frac{1}{\alpha_k}L_1\mathrm{dist}(x_k, X^*)\cdot\tau\|x_{k+1} - x_k\|^2 \\ &\leq L_1\tau\mathrm{dist}(x_k, X^*)\|d_k\|^2. \end{aligned}$$

Thus

$$g_k^T d_k = F(x_k)^T[F'(x_k) - H_k]d_k + F(x_k)^T H_k d_k = o(\|d_k\|^2) + F(x_k)^T H_k d_k \tag{30}$$

and

$$d_k^T F'(x_k)^T F'(x_k) d_k = o(\|d_k\|^2) + d_k^T H_k^T H_k d_k. \tag{31}$$

By(29)-(31), we have

$$g_k^T d_k + d_k^T F'(x_k)^T F'(x_k) d_k = o(\|d_k\|^2). \tag{32}$$

Combining (32) with Lemma 4.1 (1), we obtain

$$\begin{aligned} & f(x_k + d_k) - f(x_k) - \frac{1}{2}g_k^T d_k \\ = \ & \frac{1}{2}\|F(x_k) + F'(x_k)d_k + o(\|d_k\|)\|^2 - \frac{1}{2}\|F(x_k)\|^2 - \frac{1}{2}g_k^T d_k \\ = \ & \frac{1}{2}g_k^T d_k + \|F(x_k)\|\cdot o(\|d_k\|) + \frac{1}{2}d_k^T F'^T(x_k)F'(x_k)d_k + F'(x_k)d_k\cdot o(\|d_k\|) + o(\|d_k\|^2) \\ = \ & o(\|d_k\|^2) + \|F(x_k)\|\cdot o(\|d_k\|) = o(\|x_k - \bar{x}_k\|^2). \end{aligned}$$

Furthermore, for sufficiently large k we have

$$\begin{aligned} f(x_k + d_k) - f(x_k) - \beta g_k^T d_k &= (\frac{1}{2} - \beta)g_k^T d_k + f(x_k + d_k) - f(x_k) - \frac{1}{2}g_k^T d_k \\ &\leq -\frac{\tau_1^2}{4}(\frac{1}{2} - \beta)\mathrm{dist}(x_k, X^*)^2 + o(\|x_k - \bar{x}_k\|^2) < 0, \end{aligned}$$

which implies that $\alpha_k = 1$ holds.

By Assumption 2,

$$\begin{aligned}\|F'(x_k)d_k\|^2 - \|H_k d_k\|^2 &= \|(F'(x_k)-H_k)d_k + H_k d_k\|^2 - \|H_k d_k\|^2\\ &= \|(F'(x_k)-H_k)d_k\|^2 + d_k^T H_k^T (F'(x_k)-H_k)d_k\\ &= o(\|d_k\|^2).\end{aligned}$$

The above formula and (28) yield that

$$\begin{aligned}& f(x_k+d_k) - q_k(d_k)\\ =\ & f(x_k) + \nabla f(x_k)^T d_k + \frac{1}{2}d_k^T G(x_k) d_k + o(\|d_k\|^2) - \frac{1}{2}\|F(x_k)+H_k d_k\|^2 - \mu_k\|d_k\|^2\\ =\ & \frac{1}{2}\|F(x_k)\|^2 + g_k^T d_k + \frac{1}{2}d_k^T[F'(x_k)^T F'(x_k) + S(x)]d_k + o(\|d_k\|^2)\\ & -\frac{1}{2}\|F(x_k)+H_k d_k\|^2 - \mu_k\|d_k\|^2\\ =\ & F(x_k)^T(F'(x_k)-H_k)d_k + \frac{1}{2}[\|F'(x_k)d_k\|^2 - \|H_k d_k\|^2] + \frac{1}{2}d_k^T S(x_k)d_k + o(\|d_k\|^2)\\ & -\mu_k\|d_k\|^2\\ =\ & o(\|d_k\|^2).\end{aligned}$$

Since

$$\begin{aligned}|\mathrm{pred}_k| &= |q_k(d_k) - q_k(0)| \geq |q_k(d_k(\xi_k)) - q_k(0)|\\ &\geq |\mu_0 \nabla f(x_k)^T d_k(\xi_k)| \geq\ c\|d_k\|^2,\end{aligned}$$

we have

$$\lim_{k\to\infty}|\rho_k - 1| = \lim_{k\to\infty}\left|\frac{f(x_k) - f(x_k+d_k) - q_k(0) + q_k(d_k)}{\mathrm{pred}_k}\right| = 0,$$

which means that $\rho_k \to 1$. Therefore, $\rho_k \geq \eta_2$ and hence $\Delta_{k+1} \geq \Delta_k$. □

Assume that x_0 is close enough to x^* such that $\alpha_k = 1$ for all k. We get the following results with proofs similar to those in [11].

Lemma 4.3. *Assume that Assumptions 1 to 3 hold, $\mu_k = \min\{1, \mu\|F(x_k)\|\}$, and x_{k-1} and $x_k \in B_{\frac{\delta}{2}}(x^*)$. Then there is a constant $\tau_5 > 0$ such that*

$$dist(x_k, X^*) \leq \tau_5\, dist(x_{k-1}, X^*)^{\frac{3}{2}}.$$

The next result shows that the assumption of Lemma 4.2 is satisfied if the starting point x_0 in the proposed algorithm is chosen sufficiently close to the solution set X^*. Let

$$r = \min\{\frac{\delta}{2(1+6\tau_3)},\ \frac{2}{3\tau_5^2}\}. \tag{33}$$

Lemma 4.4. *Assume that Assumptions 1 to 3 hold. If the starting point $x_0 \in \Omega$ used in the proposed algorithm belongs to the ball $B_r(x^*)$, then all iterates x_k generated by the proposed algorithm belong to the ball $B_{\frac{\delta}{2}}(x^*)$.*

Lemma 4.5. *Assume that Assumptions 1 to 3 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm with starting point $x_0 \in B_r(x^*)$, where r is defined by (33). Then the sequence $\{x_k\}$ converges to a solution of (1) which belonging to the ball $B_{\frac{\delta}{2}}(x^*)$.*

We now obtain our main local convergence result of this section.

Theorem 4.1. *Assume that Assumptions 1 to 3 hold. Let* $\{x_k\}$ *be a sequence generated by the proposed algorithm with starting point* $x_0 \in B_r(x^*)$*, where* r *is defined by (33) and the limit point is* $\bar{x}$*. Then the sequence* $\{x_k\}$ *converges locally Q-superlinearly to* $\bar{x}$ *at 1.5-order rate of local convergence.*

Theorem 4.1 means that under Assumptions 1 to 3, the distance $\mathrm{dist}(x_k, X^*)$ converges to zero locally Q-superlinearly. We feel that the numerical test will be implemented in practice further.

References

[1] P L D Angelis, G Toraldo. *On the identification property of a projected gradient method*, SIAM J Numer Anal, 1993, 30(5): 1483-1497.

[2] J V Bucke, J J Moré, G Toraldo. *Convergence properties of trust region methods for linear and convex constraints*, Math Programming, 1990, 47: 305-336.

[3] K Christian, Y Nobuo, F Masao. *Levenberg-Marquardt methods with strong local convergence properities for solving nonlinear equations with convex constraints*, J Comput Appl Math, 2004, 172: 375-397.

[4] E M Gafni, D P Bertsekas. *Two-metric projection methods for constrained optimization*, SIAM J Control Optim, 1984, 22: 936-964.

[5] P H Calamai, J J Moré. *Projected gradient methods for linearly constrained problems*, Math Programming, 1987, 30: 93-116.

[6] J J Moré. *Trust regions and projected gradient*, In: *System Modelling and Optimization*, Springer, 1988, 1-13.

[7] P L Toint. *Global convergence of a class of trust region algorithms for nonconvex minimization in Hilbert space*, IMA J Numer Anal, 1988, 8: 231-252.

[8] N Yamashita, M Fukushima. *On the rate of convergence of the Levenberg-Marquardt mehod*, Computing, 2001, 15: 237-249.

[9] E H Zarantonello. *Projections on convex sets in Hilbert space and spectral theory*, In: *Contributions to Nonlinear Functional Analysis*, Academic Press, 1971, 237-424.

[10] D T Zhu. *Convergence properties of projected gradient methods with nonmonotonic back tracking technique for convex constrained optimization*, Systems Sci Math Sci, 2000, 13(4): 407-424.

[11] D T Zhu. *Affine scaling interior Levenberg-Marquardt method for bound-constrained semismooth equations under local error bound conditions*, J Comput Appl Math, 2008, 219: 198-215.

[1] Mathematics and Science College, Shanghai Normal University, Shanghai 200234, China
Email: jiachx@shnu.edu.cn

[2] Business College, Shanghai Normal University, Shanghai 200234, China