This article was downloaded by: [Academy of Mathematics and System Sciences]
On: 06 February 2013, At: 03:47
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Optimization Methods and Software
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/goms20

A regularized Newton method for monotone nonlinear equations and its application

Jinyan Fan [a] & Yaxiang Yuan [b]

[a] Department of Mathematics, and MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China

[b] State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, The Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, PO Box 2719, Beijing, 100190, People's Republic of China

Accepted author version posted online: 07 Nov 2012.Version of record first published: 08 Jan 2013.

To cite this article: Jinyan Fan & Yaxiang Yuan (2013): A regularized Newton method for monotone nonlinear equations and its application, Optimization Methods and Software, DOI:10.1080/10556788.2012.746344

To link to this article: http://dx.doi.org/10.1080/10556788.2012.746344

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Optimization Methods & Software, 2012
http://dx.doi.org/10.1080/10556788.2012.746344

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

A regularized Newton method for monotone nonlinear equations and its application

Jinyan Fan[a]* and Yaxiang Yuan[b]

[a]*Department of Mathematics, and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China;* [b]*State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, The Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, PO Box 2719, Beijing 100190, People's Republic of China*

(*Received 9 June 2012; final version received 31 October 2012*)

In this paper, we propose a regularized Newton method for the system of monotone nonlinear equations. The regularization parameter is taken as the norm of the residual, and a correction step with little additional calculations is also computed to compensate for the shorter trial step due to the introduction of the regularization parameter. Under the local error bound condition which is weaker than nonsingularity, we show that the new regularized Newton method with correction has quadratic convergence. We also apply the new method to the unconstrained convex optimization problems which may have singular Hessian at the solutions and develop a globally convergent regularized Newton algorithm by using trust region technique. Numerical results show that the algorithm is very efficient and robust.

Keywords: monotone nonlinear equations; regularized Newton method; correction technique; local error bound; unconstrained convex optimization

AMS Subject Classification: 65K05; 90C30

1. Introduction

We consider the problem of solving a system of monotone nonlinear equations, represented by

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

where $F(x) : R^n \to R^n$ is continuously differentiable and monotone, that is,

$$< F(x) - F(y), x - y > \geq 0 \quad \forall x, y \in R^n, \tag{2}$$

where $< \cdot, \cdot >$ denotes the inner product in R^n. Monotone nonlinear equations have many real applications. For example, some monotone nonlinear complementarity problems and variational inequality problems can be transformed into monotone nonlinear equations [1,3,12,14,19–21], and the unconstrained convex nonlinear optimization problems with singular solutions can also be reduced to a special case of monotone nonlinear equations [8]. The purpose of this work is to propose a regularized Newton method for (1). Due to the nonlinearity of $F(x)$, (1) may have

*Corresponding author. Email: jyfan@sjtu.edu.cn

© 2012 Taylor & Francis

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

no solutions. Throughout this paper, we assume that the solution set of (1) denoted by X^* is nonempty, and in all cases, $\|\cdot\|$ refers to the 2-norm.

The Newton method is a classical method for nonlinear equations. At every iteration, it computes the trial step

$$d_k^N = -J_k^{-1}F_k, \tag{3}$$

where $F_k = F(x_k)$ and $J_k = F'(x_k)$ is the Jacobian. It is well known that Newton's method has quadratic convergence if the Jacobian is Lipschitz continuous and nonsingular at the solution.

However, the condition of nonsingularity is too strong. Since $F(x)$ is a monotone function, the Jacobian $J(x)$ is positive semidefinite which implies it may be singular. In this case, we may compute the Moore–Penrose step $d_k^{\mathrm{MP}} = -J_k^{+}F_k$, which is a minimizer of $\min_d \|F_k + J_k d\|^2$. However, the computation of the singular value decomposition to obtain J_k^{+} is sometimes prohibitive. Hence, computing a direction that is close to d_k^{MP} may be a good idea.

To overcome the difficulty caused by the possible singularity of J_k, the regularized Newton method solves the following linear equations

$$(J_k + \lambda_k I)d = -F_k \tag{4}$$

to obtain the trial step d_k, where the regularization parameter $\lambda_k > 0$ is updated from iteration to iteration. If the Jacobian is Lipschitz continuous and nonsingular at the solution and if the initial iterative point is chosen sufficiently close to the solution, then the trial steps satisfy

$$\|d_{k+1}\| \le \kappa(\|d_k\|^2 + \lambda_k\|d_k\|) \tag{5}$$

for some positive κ [6]. The above inequality implies that the convergence rate of the regularized Newton method is at most quadratic under the condition of nonsingularity.

Since J_k is positive semidefinite, the positive λ_k makes d_k away from d_k^{MP}. If we replace $-F_k$ by $-F_k + \lambda_k d_k^{\mathrm{MP}}$ on the right-hand side of (4), then in fact d_k^{MP} is a solution of (4), which is desirable. Since we do not want to compute d_k^{MP}, we instead replace it with the best approximation we have available, that is, d_k. Thus, it is very likely that the solution s_k of the following linear equations

$$(J_k + \lambda_k I)d = -F_k + \lambda_k d_k \tag{6}$$

is closer to d_k^{MP} than d_k. We call

$$\tilde{d}_k = \lambda_k (J_k + \lambda_k I)^{-1} d_k \tag{7}$$

a correction step. Note that the coefficient matrix of linear equations (6) is the same as that of (4); therefore, we can make use of the available factorization of $J_k + \lambda_k I$ in (4), and only need a small amount of additional calculations to obtain s_k, which is favourable and easily implemented in real applications.

Now the interesting issue comes how to choose the regularization parameter λ_k, which will play an important role not only in theory but also in the efficient implementation of the method. In view of (5), the smaller the λ_k, the faster may be the convergence. However, if λ_k is chosen too small, it will lose its role. On the other hand, if λ_k is chosen too large, the trial step will be short and far away from the Moore–Penrose step, which may also be unsatisfactory. Anyway, the role of λ_k is the same as that of the Levenberg–Marquardt parameter [7,9,10]. Yamashita and Fukushima [15] chose $\lambda_k = \|F_k\|^2$ and showed that the Levenberg–Marquardt method has quadratic convergence under the local error bound condition which is weaker than nonsingularity. Fan and Yuan took $\lambda_k = \|F_k\|^\delta$ with $\delta \in [1, 2]$ and proved that the Levenberg–Marquardt method preserves the

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

quadratic convergence under the same conditions [2]. Based on the better performance of the Levenberg–Marquardt method with $\lambda_k = \|F_k\|$, we will consider the choice of

$$\lambda_k = \|F_k\| \tag{8}$$

in this paper.

We will show that our regularized Newton method with correction for monotone nonlinear equations could also achieve the quadratic convergence under the local error bound condition. Furthermore, we will apply the regularized Newton method with correction to solve the unconstrained convex nonlinear optimization problems with singular solutions. Some of the results given in [8] will be improved, in which the regularized Newton method is designed for the unconstrained convex nonlinear optimization with singular solutions.

The paper is organized as follows. In Section 2, we analyse the regularized Newton method with correction for monotone nonlinear equations without line search and obtain the quadratic convergence under the local error bound condition. In Section 3, we present a globally convergent regularized Newton algorithm with correction for monotone nonlinear equations by using line search. We resort to a Levenberg–Marquardt step in the case that the trial step fails to decrease the merit function. In Section 4, we discuss the application of regularization and correction approaches to the unconstrained convex nonlinear optimization problems with singular solutions and develop a globally convergent regularized Newton algorithm with correction by using trust region technique. The quadratic convergence of the algorithm is also proven. In Section 5, we test the regularized Newton algorithm with correction on the problem given in [8]. We conclude the paper in Section 6.

2. Local convergence analysis of the regularized Newton method with correction

In this section, we will study the convergence rate of the regularized Newton method with correction for (1) without line search. That is,

$$x_{k+1} := x_k + s_k, \tag{9}$$

where s_k is computed by (6) with d_k being obtained by (4). We assume that x^* is a solution of (1).

The local convergence theory requires the following assumptions.

Assumption 2.1

(a) *$F(x)$ is continuously differentiable and monotone.*
(b) *The Jacobian $J(x)$ is Lipschitz continuous on $N(x^*, b_1)$ with $b_1 < 1$, that is, there exists a positive constant L_1 such that*

$$\|J(y) - J(x)\| \le L_1\|y - x\| \quad \forall x, y \in N(x^*, b_1). \tag{10}$$

By Assumption 2.1(a), we know that the Jacobian $J(x)$ is positive semidefinite, that is, $J(x) \succeq 0$, and therefore

$$d^{\mathrm{T}}J(x)d \ge 0 \quad \forall d, x \in R^n. \tag{11}$$

Note that $J(x)$ is not necessarily symmetric or nonsingular. For example, the matrix $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is semidefinite, but is neither symmetric nor nonsingular.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Furthermore, it follows from Assumption 2.1(b) that

$$\|F(y) - F(x) - J(x)(y - x)\| \leq L_1 \|y - x\|^2 \quad \forall x, y \in N(x^*, b_1), \tag{12}$$

and there exists a constant $L_2 > 0$ such that

$$\|F(y) - F(x)\| \leq L_2 \|y - x\| \quad \forall x, y \in N(x^*, b_1). \tag{13}$$

ASSUMPTION 2.2 $\|F(x)\|$ *provides a local error bound on some* $N(x^*, b_1)$ *for (1), that is, there exists a positive constant* $c_1 > 0$ *such that*

$$\|F(x)\| \geq c_1 \operatorname{dist}(x, X^*) \quad \forall x \in N(x^*, b_1) = \{x \mid \|x - x^*\| \leq b_1\}, \tag{14}$$

where $\operatorname{dist}(x, X^*)$ *is defined as the distance from x to the solution set* X^*.

Note that if $J(x^*)$ is nonsingular, then x^* is an isolated solution; hence, $\|F(x)\|$ provides a local error bound on some neighbourhood of x^*. However, the converse is not necessarily true (e.g. see [13,15]). So, the local error bound condition is weaker than nonsingularity.

In the following, we denote $\bar{x}_k$ the vector in the solution set X^* that satisfies

$$\|x_k - \bar{x}_k\| = \operatorname{dist}(x_k, X^*).$$

To analyse the convergence rate of $\{x_k\}$, we need to investigate the properties of the positive semidefinite matrix J_k. Though J_k is not necessarily symmetric, it still has some desirable properties similar to symmetric positive semidefinite matrix. The first lemma given below shows the relationship between the positive semidefinite matrix and symmetric positive semidefinite matrix; the second one gives the bounds of a positive definite matrix and its inverse.

LEMMA 2.3 *A real-valued matrix A is positive semidefinite if and only if* $(A + A^{\mathrm{T}})/2$ *is positive semidefinite.*

LEMMA 2.4 *Suppose A is positive semidefinite. Then,* $\|A + \alpha I\| \geq \alpha$ *and* $\|(A + \alpha I)^{-1}\| \leq \alpha^{-1}$ *hold for any* $\alpha > 0$.

Proof It follows from Lemma 2.3 and the definition of the 2-norm that

$$\begin{aligned}
\|A + \alpha I\| &= \sqrt{\lambda_{\max}((A + \alpha I)^{\mathrm{T}}(A + \alpha I))} \\
&= \sqrt{\lambda_{\max}(A^{\mathrm{T}}A + \alpha(A^{\mathrm{T}} + A) + \alpha^2 I)} \\
&\geq \sqrt{\lambda_{\max}(\alpha^2 I)} \\
&= \alpha,
\end{aligned}$$

where $\lambda_{\max}((A + \alpha I)^{\mathrm{T}}(A + \alpha I))$ means the largest eigenvalue of $(A + \alpha I)^{\mathrm{T}}(A + \alpha I)$. Similarly, we have

$$\begin{aligned}
\|(A + \alpha I)^{-1}\| &= \sqrt{\lambda_{\max}((A + \alpha I)^{-T}(A + \alpha I)^{-1})} \\
&= \sqrt{\lambda_{\max}(((A + \alpha I)(A + \alpha I)^{\mathrm{T}})^{-1})} \\
&= \sqrt{\frac{1}{\lambda_{\min}(AA^{\mathrm{T}} + \alpha(A + A^{\mathrm{T}}) + \alpha^2 I)}} \\
&\leq \alpha^{-1}.
\end{aligned}$$

The proof is completed. ■

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

LEMMA 2.5 *Suppose Assumptions* 2.1 *and* 2.2 *hold. If* $x_k \in N(x^*, b_1/2)$, *then there exists a constant* $c_2 > 0$ *such that*

$$\|s_k\| \le c_2 \operatorname{dist}(x_k, X^*). \tag{15}$$

Proof Since $x_k \in N(x^*, b_1/2)$, we have

$$\|\bar{x}_k - x^*\| \le \|\bar{x}_k - x_k\| + \|x_k - x^*\| \le 2\|x_k - x^*\| \le b_1,$$

which means $\bar{x}_k \in N(x^*, b_1)$. Then, it follows from (8) and Assumption 2.2 that the regularization parameter satisfies

$$\lambda_k = \|F_k\| \ge c_1 \|\bar{x}_k - x_k\|. \tag{16}$$

By the definition of $\tilde{d}_k$ and Lemma 2.4, we have

$$\|\tilde{d}_k\| = \| - \lambda_k (J_k + \lambda_k I)^{-1} d_k\| \le \lambda_k \|(J_k + \lambda_k I)^{-1}\| \|d_k\| \le \|d_k\|. \tag{17}$$

Moreover, we deduce from (12), (16), Lemma 2.4 and $F(\bar{x}_k) = 0$ that

$$\begin{aligned}
\|d_k - (\bar{x}_k - x_k)\| &= \| - (J_k + \lambda_k I)^{-1} F_k - \bar{x}_k + x_k\| \\
&= \|(J_k + \lambda_k I)^{-1} (F_k + (J_k + \lambda_k I)(\bar{x}_k - x_k))\| \\
&\le \|(J_k + \lambda_k I)^{-1}\| (\|F_k + J_k(\bar{x}_k - x_k)\| + \lambda_k \|\bar{x}_k - x_k\|) \\
&\le \lambda_k^{-1} L_1 \|\bar{x}_k - x_k\|^2 + \|\bar{x}_k - x_k\| \\
&\le (c_1^{-1} L_1 + 1) \|\bar{x}_k - x_k\|,
\end{aligned}$$

which yields

$$\|d_k\| \le (c_1^{-1} L_1 + 2) \|\bar{x}_k - x_k\|. \tag{18}$$

Combining (17) and (18), we obtain

$$\|s_k\| = \|d_k + \tilde{d}_k\| \le c_2 \|\bar{x}_k - x_k\|, \tag{19}$$

where $c_2 = 2(c_1^{-1} L_1 + 2)$ is a positive constant. ∎

THEOREM 2.6 *Suppose Assumptions* 2.1 *and* 2.2 *hold. If* x_0 *is chosen sufficiently close to the solution set* X^*, *then* $\{x_k\}$ *converges to some solution of (1) quadratically.*

Proof First we assume both $x_{k+1} \in N(x^*, b_1/2)$ and $x_k \in N(x^*, b_1/2)$. It then follows from (6), (12)–(14) and (17)–(19) that

$$\begin{aligned}
\|\bar{x}_{k+1} - x_{k+1}\| &\le c_1^{-1} \|F(x_{k+1})\| \\
&\le c_1^{-1} \|F_k + J_k s_k\| + c_1^{-1} L_1 \|s_k\|^2 \\
&= c_1^{-1} \|\lambda_k (d_k - s_k)\| + c_1^{-1} L_1 \|s_k\|^2 \\
&\le c_3 \|\bar{x}_k - x_k\|^2,
\end{aligned} \tag{20}$$

where $c_3 = c_1^{-1}(c_1^{-1} L_1 + 2) L_2 + c_1^{-1} c_2^2 L_1$ is a positive constant.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

We now show by induction that if x_0 is chosen sufficiently close to X^*, then $x_k \in N(x^*, b_1/2)$ for all k. Let

$$r = \min\left\{\frac{b_1}{2(1+3c_2)}, \frac{1}{2c_3}\right\} \tag{21}$$

and $x_0 \in N(x^*, r)$. It follows from Lemma 2.5 that

$$\|x_1 - x^*\| = \|x_0 + s_0 - x^*\| \le \|x_0 - x^*\| + c_2\|x_0 - \bar{x}_0\| \le (1+c_2)r \le \frac{b_1}{2},$$

which means $x_1 \in N(x^*, b_1/2)$. Suppose $x_i \in N(x^*, b_1/2)$ for $i = 2, \dots, k$. By Lemma 2.5, (20) and (21), we obtain that

$$\|x_i - \bar{x}_i\| \le c_3\|x_{i-1} - \bar{x}_{i-1}\|^2 \le \cdots \le c_3^{2^i-1}\|x_0 - \bar{x}_0\|^{2^i} \le 2\left(\frac{1}{2}\right)^{2^i} r.$$

Furthermore, we have

$$\begin{aligned}
\|x_{k+1} - x^*\| &\le \|x_1 - x^*\| + \sum_{i=1}^{k}\|s_k\| \\
&\le (1+c_2)r + c_2\sum_{i=1}^{k}\|x_i - \bar{x}_i\| \\
&\le (1+c_2)r + 2c_2 r\sum_{i=1}^{k}\left(\frac{1}{2}\right)^{2^i} \\
&\le (1+c_2)r + 2c_2 r\sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^{i} \\
&\le (1+3c_2)r \\
&\le \frac{b_1}{2},
\end{aligned}$$

which implies $x_{k+1} \in N(x^*, b_1/2)$. So, if x_0 is chosen sufficiently close to X^*, then all x_k are in $N(x^*, b_1/2)$. Therefore, $\{x_k\}$ converges to the solution set X^* quadratically due to (20).

Note that since

$$\|\bar{x}_k - x_k\| \le \|\bar{x}_{k+1} - x_k\| \le \|\bar{x}_{k+1} - x_{k+1}\| + \|s_k\|, \tag{22}$$

we may deduce from (20) that

$$\|\bar{x}_k - x_k\| \le 2\|s_k\| \tag{23}$$

for all sufficiently large k. Combining this inequality with (19) and (20), we obtain that

$$\|s_{k+1}\| \le O(\|s_k\|^2), \tag{24}$$

which indicates that $\{x_k\}$ converges to some solution of (1) quadratically. The proof is completed. ∎

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

3. A globally convergent regularized Newton algorithm with correction for monotone nonlinear equations

We follow the classical approach of minimizing a merit function involving some norm of the residual [16]. For simplicity, we choose to consider

$$\phi(x) = \tfrac{1}{2}\|F(x)\|^2. \tag{25}$$

We call a point x such that $\phi'(x) = 0$ is a stationary point of $\phi(x)$.

Since J_k is positive semidefinite, we have

$$d_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k = -d_k^{\mathrm{T}} J_k^{\mathrm{T}} J_k d_k - \lambda_k d_k^{\mathrm{T}} J_k^{\mathrm{T}} d_k \le 0, \tag{26}$$

$$s_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k = -d_k^{\mathrm{T}} J_k^{\mathrm{T}} J_k d_k - 2\lambda_k d_k^{\mathrm{T}} J_k^{\mathrm{T}} d_k \le 0. \tag{27}$$

But these two inequalities cannot guarantee d_k and s_k to be descent directions of $\phi(x)$ at x_k as J_k may not be positive definite. So, when s_k fails to decrease the merit function, we resort to the Levenberg–Marquardt step

$$\bar{s}_k = -(J_k^{\mathrm{T}} J_k + \lambda_k I)^{-1} J_k^{\mathrm{T}} F_k, \tag{28}$$

which is always a decent direction of $\phi(x)$ at x_k because of

$$\bar{s}_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k = -\bar{s}_k^{\mathrm{T}} (J_k^{\mathrm{T}} J_k + \lambda_k I)\bar{s}_k < 0. \tag{29}$$

If the Jacobian is Lipschitz continuous, then the step size α_k obtained by Wolfe or Armijo line search along $\bar{s}_k$ satisfies

$$\|F(x_k + \alpha_k \bar{s}_k)\|^2 \le \|F(x_k)\|^2 - \beta_1\beta_3 \frac{(\bar{s}_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k)^2}{\|\bar{s}_k\|^2} \tag{30}$$

for some positive constants β_1 and β_3 [17,18].

The regularized Newton algorithm with correction for monotone nonlinear equations is presented as follows.

ALGORITHM 1

Step 1. Given $x_0 \in R^n, \eta \in (0,1), \varepsilon \ge 0$. *Set* $k = 0$.

Step 2. If $\|J_k^{\mathrm{T}} F_k\| \le \varepsilon$, *stop.*

Step 3. Compute $\lambda_k = \|F_k\|$. *Solve*

$$(J_k + \lambda_k I)d = -F_k \tag{31}$$

to obtain d_k. *Solve*

$$(J_k + \lambda_k I)d = -F_k + \lambda_k d_k \tag{32}$$

to obtain s_k.

Step 4. If s_k *satisfies*

$$\|F(x_k + s_k)\| \le \eta \|F(x_k)\|, \tag{33}$$

set $x_{k+1} = x_k + s_k$; *otherwise, solve*

$$(J_k^{\mathrm{T}} J_k + \lambda_k I)d = -J_k^{\mathrm{T}} F_k \tag{34}$$

to obtain $\bar{s}_k$, *compute the step size* α_k *that satisfies (30), and set* $x_{k+1} = x_k + \alpha_k \bar{s}_k$. *Set* $k := k + 1$ *and go to Step* 2.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

To study the global convergence of the above algorithm, we make the following assumption.

ASSUMPTION 3.1 *$F(x)$ is continuously differentiable and the Jacobian $J(x)$ is Lipschitz continuous, that is, there exists a positive constant L_1 such that*

$$||J(y) - J(x)|| \le L_1||y - x|| \quad \forall x, y. \tag{35}$$

THEOREM 3.2 *Suppose Assumption 3.1 holds. Then, any accumulation point of the sequence generated by Algorithm 1 is a stationary point of $\phi(x)$.*

Proof It is easy to see that $\|F_k\|$ is monotonically decreasing and bounded below. If $\|F_k\|$ converges to zero, then any accumulation point of $\{x_k\}$ is a solution of (1). Otherwise, suppose $\|F(x_k)\| \to \gamma$, where γ is a positive constant. Then, inequality (33) only holds finitely many times and (30) will be satisfied for all sufficiently large k. So, we have

$$\sum_{k=1}^{\infty} \frac{(\bar{s}_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k)^2}{\|\bar{s}_k\|^2} < +\infty. \tag{36}$$

By the definition of $\bar{s}_k$ and $\lambda_k = \|F(x_k)\| \ge \gamma$, we know that

$$(\bar{s}_k^{\mathrm{T}} J_k^{\mathrm{T}} F_k)^2 = (\bar{s}_k^{\mathrm{T}} (J_k^{\mathrm{T}} J_k + \lambda_k I)\bar{s}_k)^2 \ge \gamma^2 \|\bar{s}_k\|^4. \tag{37}$$

The above two inequalities imply that

$$\lim_{k\to\infty} \|\bar{s}_k\| = 0. \tag{38}$$

It then follows from (34) that

$$\lim_{k\to\infty} J_k^{\mathrm{T}} F_k = 0, \tag{39}$$

which means any accumulation point of $\{x_k\}$ is a stationary point of $\phi(x)$. ■

Fukushima [3] showed that the asymmetric variational inequality problem could be equivalent to an optimization problem. Moreover, any stationary point of such optimization problem is a solution of the variational inequality problem under the condition that $J(x)$ is positive definite. Actually, the result also holds true here. Suppose $\tilde{x}$ is a stationary point of the merit function, that is, $J(\tilde{x})^{\mathrm{T}} F(\tilde{x}) = 0$. Then, $F(\tilde{x})^{\mathrm{T}} J(\tilde{x})^{\mathrm{T}} F(\tilde{x}) = 0$. Hence, if $J(x)$ is positive definite, then $F(\tilde{x}) = 0$. So, we have the following result.

THEOREM 3.3 *Suppose $F(x)$ is continuously differentiable and the Jacobian $J(x)$ is Lipschitz continuous and positive definite. Then, any accumulation point of the sequence generated by Algorithm 1 is a solution of (1). Moreover, if $F(x)$ satisfies the local error bound condition, then the convergence rate of Algorithm 1 is quadratic.*

Proof We now proceed to prove the second part of the theorem. Without loss of generality, we suppose x^* is an accumulation point of $\{x_k\}$. Then, there exists a large $\tilde{k}$ such that $\|F_{\tilde{k}}\| \le \eta c_1^2/c_3 L_2$ and $x_k \in N(x^*, r)$ for all $k \ge \tilde{k}$, where c_1, c_3, L_2 and r are defined in Section 2, and $\eta \in (0, 1)$ is a parameter in Algorithm 1. It now follows from (13), (14) and (20) that

$$\frac{\|F(x_{\tilde{k}+1})\|}{\|F(x_{\tilde{k}})\|} \le \frac{L_2\|x_{\tilde{k}+1} - \bar{x}_{\tilde{k}+1}\|}{c_1\|x_{\tilde{k}} - \bar{x}_{\tilde{k}}\|} \le \frac{L_2 c_3}{c_1}\|x_{\tilde{k}} - \bar{x}_{\tilde{k}}\| \le \frac{L_2 c_3 \|F(x_{\tilde{k}})\|}{c_1^2} \le \eta.$$

Furthermore, we have $\|F(x_{k+1})\| \le \eta\|F(x_k)\|$ for all $k > \tilde{k}$, which implies that $x_{k+1} = x_k + s_k$ holds for all sufficiently large k. By Theorem 2.6, we know that the convergence rate is quadratic. ■

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

In Algorithm 1, when s_k fails to satisfy (33), we use an alternative step $\bar{s}_k$, because $\bar{s}_k$ is a sufficiently descent direction of the merit function at x_k. As a matter of fact, any other sufficient descent directions could be used instead of $\bar{s}_k$, and the global convergence will be preserved.

4. Application to unconstrained convex optimization problems with singular Hessian at solutions

In this section, we will consider the unconstrained minimization problem

$$\min_{x \in R^n} f(x), \tag{40}$$

where $f : R^n \to R$ is a convex LC^2 function, that is, f is twice continuously differentiable and the Hessian $\nabla^2 f$ is Lipschitz continuous. Suppose the minimizer set S^* of $f(x)$ is nonempty, and $g(x) = \nabla f(x)$ and $H(x) = \nabla^2 f(x)$. It is well known that $f(x)$ is convex if and only if $H(x)$ is symmetric positive semidefinite for all $x \in R^n$. Moreover, $x^* \in S^*$ if and only if x^* is a solution of the following nonlinear equations:

$$g(x) = 0. \tag{41}$$

Hence, we could get the minimizer of $f(x)$ by solving (41).

Hanger and Zhang presented the proximal point methods for (40) and studied the convergence properties of the methods under the local error bound condition in [4,5]. Since $H(x)$ is positive semidefinite, it is natural to apply the regularized Newton method to solve (40). Li *et al.* [8] presented the globally convergent inexact regularized Newton algorithms for (40) by line search and discussed the convergence rate under the local error bound condition. In this section, we will propose a globally convergent regularized Newton algorithm with correction for (40) by using trust region technique and show that the algorithm has quadratic convergence.

4.1 *Algorithm and global convergence*

Due to the symmetry of $H(x)$, some attractive properties of the trial step could be obtained. Suppose d_k is the solution of the linear equations

$$(H_k + \lambda_k I)d = -g_k, \tag{42}$$

where $\lambda_k = \|g_k\|$ and s_k is the solution of

$$(H_k + \lambda_k I)d = -g_k + \lambda_k d_k. \tag{43}$$

Then, we have

$$s_k^{\mathrm{T}} g_k = -d_k^{\mathrm{T}} H_k d_k - 2\lambda_k d_k^{\mathrm{T}} d_k < 0, \tag{44}$$

which implies that s_k is always a descent direction for $f(x)$ at x_k. So, when s_k fails to satisfy (33) in Algorithm 1, we can perform line search directly along the available s_k instead of the Levenberg–Marquardt step $\bar{s}_k$, which will be preferable in the implementation.

In the following, we will develop a globally convergent regularized Newton algorithm with correction for (41) using another kind of globalization approach: the trust region technique.

We begin with the definitions of the predicted reduction

$$\mathrm{Pred}_k = -g_k^{\mathrm{T}} s_k - \tfrac{1}{2} s_k^{\mathrm{T}} H_k s_k,$$

and the actual reduction

$$\mathrm{Ared}_k = f_k - f(x_k + s_k).$$

The ratio of the actual reduction to the predicted reduction

$$r_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$$

plays an important role in deciding whether to accept the trial step s_k and how to adjust the regularization parameter λ_k.

The regularized Newton algorithm with correction for unconstrained convex optimization problems is stated as follows.

Algorithm 2

Step 1. Given $x_0 \in R^n, \varepsilon \geq 0, \mu_0 > m > 0, 0 < p_0 \leq p_1 \leq p_2 < 1, 0 < p_4 < 1 < p_3$. *Set* $k = 0$.

Step 2. If $\|g_k\| \leq \varepsilon$, *stop.*

Step 3. Compute $\lambda_k = \mu_k \|g_k\|$. *Solve*

$$(H_k + \lambda_k I)d = -g_k \tag{45}$$

to obtain d_k. *Solve*

$$(H_k + \lambda_k I)d = -g_k + \lambda_k d_k \tag{46}$$

to obtain s_k.

Step 4. Compute $r_k = \mathrm{Ared}_k/\mathrm{Pred}_k$. *Set*

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } r_k \geq p_0, \\ x_k, & \text{otherwise}. \end{cases} \tag{47}$$

Step 5. Choose μ_{k+1} *as*

$$\mu_{k+1} = \begin{cases} p_3\mu_k, & \text{if } r_k < p_1, \\ \mu_k, & \text{if } r_k \in [p_1, p_2], \\ \max\{p_4\mu_k, m\}, & \text{if } r_k > p_2. \end{cases} \tag{48}$$

Set $k := k + 1$ *and go to Step 2.*

It is easy to see from (45), (46) and Lemma 2.4 that the correction step

$$\tilde{d}_k = s_k - d_k = (H_k + \lambda_k I)^{-1} \lambda_k d_k \tag{49}$$

satisfies

$$\|\tilde{d}_k\| \leq \|d_k\|, \tag{50}$$

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

which gives

$$\|s_k\| = \|d_k + \tilde{d}_k\| \leq 2\|d_k\|. \tag{51}$$

Note that the regularization step d_k is the minimizer of the following unconstrained convex minimization problem:

$$\min_{d \in R^n} g_k^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} (H_k + \lambda_k I) d. \tag{52}$$

If we let

$$\Delta_k = \| - (H_k + \lambda_k I)^{-1} g_k\| = \|d_k\|,$$

then d_k is also a solution of the trust region problem

$$\begin{aligned} \min_{d \in R^n} \quad & \varphi(d) = g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} H_k d, \\ \text{s.t.} \quad & \|d\| \leq \Delta_k. \end{aligned} \tag{53}$$

Hence, by the famous result given by Powell [11], we obtain that

$$\varphi(0) - \varphi(d_k) \geq \frac{1}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|H_k\|}\right\}. \tag{54}$$

By some simple calculations, we deduce from (49) that

$$\begin{aligned} \varphi(d_k) - \varphi(s_k) &= g_k^{\mathrm{T}} d_k + \tfrac{1}{2} d_k^{\mathrm{T}} H_k d_k - g_k^{\mathrm{T}} s_k - \tfrac{1}{2} s_k^{\mathrm{T}} H_k s_k \\ &= -g_k^{\mathrm{T}} \tilde{d}_k - \tfrac{1}{2} \tilde{d}_k^{\mathrm{T}} H_k \tilde{d}_k - \tilde{d}_k^{\mathrm{T}} H_k d_k \\ &= \lambda_k \tilde{d}_k^{\mathrm{T}} d_k - \tfrac{1}{2} \tilde{d}_k^{\mathrm{T}} H_k \tilde{d}_k \\ &= \tfrac{1}{2} \tilde{d}_k^{\mathrm{T}} H_k \tilde{d}_k + \lambda_k \tilde{d}_k^{\mathrm{T}} \tilde{d}_k \\ &\geq 0, \end{aligned} \tag{55}$$

so the predicted reduction satisfies

$$\mathrm{Pred}_k = \varphi(0) - \varphi(s_k) \geq \varphi(0) - \varphi(d_k). \tag{56}$$

Combining the above inequality with (51) and (54), we have the following lemma.

LEMMA 4.1 *Let s_k be computed by (45) and (46). Then, the predicted reduction satisfies*

$$\mathrm{Pred}_k \geq \frac{1}{2}\|g_k\| \min\left\{\frac{\|s_k\|}{2}, \frac{\|g_k\|}{\|H_k\|}\right\}. \tag{57}$$

Inequality (57) plays an essential role in the global convergence of Algorithm 2. The following result shows that Algorithm 2 has global convergence under some suitable conditions.

THEOREM 4.2 *Suppose $f(x)$ is convex, twice continuously differentiable and bounded below, and $H(x)$ is bounded above. Then, the sequence generated by Algorithm 2 satisfies*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{58}$$

Proof Now assume, for the purpose of a contradiction, that there exists a positive K such that

$$\|g_k\| \geq \tau \quad \forall k \geq K \tag{59}$$

for some $\tau > 0$. Let T be the set of all indices at which $r_k \geq p_1$. Since $\{f_k\}$ is monotonically decreasing, it follows from (51), (56) and (59) that

$$\begin{aligned} f_1 &\geq \sum_{k\in T}(f_k - f_{k+1}) \\ &\geq \sum_{k\in T} p_1 \mathrm{Pred}_k \\ &\geq \sum_{k\in T} \frac{p_1}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|H_k\|}\right\} \\ &\geq \sum_{k\in T} \frac{p_1\tau}{2} \min\left\{\frac{\|s_k\|}{2}, \frac{\tau}{\|H_k\|}\right\}. \end{aligned} \tag{60}$$

Then, by the boundedness of $f(x)$ and $H(x)$, we have that

$$\sum_{k\in T}\|d_k\| < +\infty \quad \text{and} \quad \sum_{k\in T}\|s_k\| < +\infty. \tag{61}$$

If T is infinite, (61) implies that $d_k \to 0, k \in T$. It then follows from (45), (59) and the boundedness of H_k that

$$1 \leq \frac{\|H_k\|}{\|g_k\|}\|d_k\| + \mu_k\|d_k\| \leq \frac{\|H_k\|}{\tau}\|d_k\| + \mu_k\|d_k\|. \tag{62}$$

Hence $\mu_k \to +\infty, k \in T$. Note that since $\mu_{k+1} = p_3\mu_k$ for all $k \notin T$ with $p_3 > 1$, we have $\mu_k \to +\infty$. On the other case that T is finite, we have $\mu_{k+1} = p_3\mu_k$ for all sufficiently large k, which also yields $\mu_k \to +\infty$. Therefore, (45) gives $d_k \to 0$, and hence

$$\lim_{k\to\infty} s_k = 0. \tag{63}$$

It now follows from (59), Lemma 4.1 and the boundedness of H_k that

$$\begin{aligned} |r_k - 1| &= \left|\frac{\mathrm{Ared}_k - \mathrm{Pred}_k}{\mathrm{Pred}_k}\right| \\ &\leq \left|\frac{f(x_k + s_k) - f_k - g_k^{\mathrm{T}} s_k - \frac{1}{2}s_k^{\mathrm{T}} H_k s_k}{\frac{1}{2}\|g_k\| \min\{\|s_k\|/2, \|g_k\|/\|H_k\|\}}\right| \\ &\leq \frac{o(\|s_k\|^2)}{\|s_k\|} \longrightarrow 0, \end{aligned} \tag{64}$$

which means

$$r_k \longrightarrow 1. \tag{65}$$

So, we deduce from (48) that there exists a positive constant $\bar{m} > m$ such that $\mu_k < \bar{m}$ holds for all sufficiently large k, which yields the desirable contradiction to $\mu_k \to +\infty$. Therefore, the assumption (59) cannot be true, which implies (58) holds true. The proof is completed. ∎

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Due to the introduction of the correction step $\tilde{d}_k$, we can see from (56) that the predicted reduction $\varphi(0) - \varphi(s_k)$ of our new Algorithm 2 is always larger than the one $\varphi(0) - \varphi(d_k)$ which is usually defined for nonlinear optimization in trust region methods. So, in view of (47), we can expect that the actual reduction $f_k - f(x_k + s_k)$ may also be larger than the usual one $f_k - f(x_k + d_k)$, which may accelerate convergence. Hence, it is reasonable to believe that s_k should generally be better than d_k.

4.2 *Quadratic convergence of Algorithm 2*

In the following, we will prove the quadratic convergence of Algorithm 2 under the local error bound condition. We assume that x^* is a minimizer of $f(x)$. The required assumptions are similar to Assumption 2.1.

ASSUMPTION 4.3

(a) *$f(x)$ is convex and continuously differentiable.*
(b) *$\|g(x)\|$ provides a local error bound on some $N(x^*, \tilde{b}_1)$ with $\tilde{b}_1 < 1$, that is, there exists a positive constant $\tilde{c}_1 > 0$ such that*

$$\|g(x)\| \geq \tilde{c}_1 \operatorname{dist}(x, S^*) \quad \forall x \in N(x^*, \tilde{b}_1). \tag{66}$$

(c) *The Hessian $H(x)$ is Lipschitz continuous on $N(x^*, \tilde{b}_1)$, that is, there exists a positive constant $\tilde{L}_1$ such that*

$$\|H(y) - H(x)\| \leq \tilde{L}_1 \|y - x\| \quad \forall x, y \in N(x^*, \tilde{b}_1). \tag{67}$$

By Assumption 4.1(c), we know

$$\|g(y) - g(x) - H(x)(y - x)\| \leq \tilde{L}_1 \|y - x\|^2 \quad \forall x, y \in N(x^*, \tilde{b}_1), \tag{68}$$

and there exists a constant $\tilde{L}_2 > 0$ such that

$$\|g(y) - g(x)\| \leq \tilde{L}_2 \|y - x\| \quad \forall x, y \in N(x^*, \tilde{b}_1). \tag{69}$$

Note that $\lambda_k = \mu_k \|g_k\|$ with $\mu_k \geq m$ for all k; taking the same process of deduction as the proof of Lemma 2.5, we obtain that there exist positive constants $\bar{c}_2$ and $\tilde{c}_2$ such that

$$\|d_k\| \leq \bar{c}_2 \operatorname{dist}(x_k, S^*) \quad \text{and} \quad \|s_k\| \leq \tilde{c}_2 \operatorname{dist}(x_k, S^*) \tag{70}$$

hold for all sufficiently large k.

Since inequality (69) implies the boundedness of $H(x)$, combining (66), (68), (70) and Lemma 4.1, we have

$$\begin{aligned}
|r_k - 1| &= \left| \frac{\text{Ared}_k - \text{Pred}_k}{\text{Pred}_k} \right| \\
&\leq \left| \frac{f(x_k + s_k) - f_k - g_k^{\mathrm{T}} s_k - \frac{1}{2} s_k^{\mathrm{T}} H_k s_k}{(\frac{1}{2}) \|g_k\| \min\{\|s_k\|/2, \|g_k\|/\|H_k\|\}} \right| \\
&\leq \frac{o(\|s_k\|^2)}{\|\bar{x}_k - x_k\| \min\{\|s_k\|, \|\bar{x}_k - x_k\|\}} \longrightarrow 0,
\end{aligned}$$

which implies $r_k \to 1$. Hence, there exists a positive constant $\tilde{m} > m$ such that

$$\mu_k < \tilde{m} \tag{71}$$

holds for all sufficiently large k. It then follows from (66) and (68)–(71) that

$$\begin{aligned} \|\bar{x}_{k+1} - x_{k+1}\| &\le \tilde{c}_1^{-1}\|g(x_{k+1})\| \\ &\le \tilde{c}_1^{-1}\|g_k + H_k s_k\| + \tilde{c}_1^{-1}\tilde{L}_1\|s_k\|^2 \\ &= \tilde{c}_1^{-1}\|\lambda_k(d_k - s_k)\| + \tilde{c}_1^{-1}\tilde{L}_1\|s_k\|^2 \\ &\le \tilde{c}_3\|\bar{x}_k - x_k\|^2 \end{aligned} \tag{72}$$

for some $\tilde{c}_3 > 0$. Using the same arguments as (22) and (23), we derived

$$\|s_{k+1}\| \le O(\|s_k\|^2). \tag{73}$$

Therefore, $\{x_k\}$ converges quadratically to some solution of (41).

As the solution of (41) is also a minimizer of $f(x)$, we summarize our results in the next theorem.

THEOREM 4.4 *Suppose Assumption* 4.3 *holds. Then, the convergence rate of Algorithm* 2 *is quadratic.*

5. Numerical experiments

We test Algorithm 2 on the unconstrained nonlinear optimization problem given in [8]. The function to be minimized is

$$f(x) = \frac{1}{2}\sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + \frac{1}{12}\sum_{i=1}^{n-1}\alpha_i(x_i - x_{i+1})^4, \tag{74}$$

where $\alpha_i \ge 0$ $(i = 1, \ldots, n-1)$ are constants. Obviously, $f(x)$ is convex and the minimizer set of $f(x)$ is

$$S = \{x \in R^n | x_1 = x_2 = \cdots = x_n\}.$$

It can also be verified that $\|g(x)\|$ provides a local error bound near the minimizer of $f(x)$. The Hessian $\nabla^2 f(x)$ is given as follows:

$$\nabla^2 f(x) = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} + \begin{pmatrix} a_1 & -a_1 & & & \\ -a_1 & a_1 + a_2 & -a_2 & & \\ & \ddots & \ddots & \ddots & \\ & & -a_{n-2} & a_{n-2} + a_{n-1} & -a_{n-1} \\ & & & -a_{n-1} & a_{n-1} \end{pmatrix},$$

where $a_i = \alpha_i(x_i - x_{i+1})^2$ $(i = 1, \ldots, n-1)$ is symmetric semidefinite, but singular as the sum of every column is zero.

The aims of the experiments are as follows: to check whether Algorithm 2 converges quadratically as stated in Section 4 and to see how well the technique of correction works. We set $p_0 = 0.0001, p_1 = 0.25, p_2 = 0.75, p_3 = 4, p_4 = 0.25, \mu_0 = 10^{-2}$ and $m = \varepsilon = 10^{-5}$ for Algorithm 2.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Table 1. Results of Algorithm 2 to test quadratic convergence.

k	0	1	2	3	4
$\|\nabla f(x_k)\|$	1.8856	0.4921	0.0320	1.1e–05	2.5e–15
$\|s_k\|$	6.0092	2.8629	0.2109	7.6e–05	

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

Table 1 reports the norms of $\nabla f(x_k)$ and s_k at every iteration when $n = 10$, $\alpha_i = 1$ $(i = 1, \ldots, n-1)$ and $x_0 = (1, 2, \ldots, n)^{\mathrm{T}}$. Algorithm 2 only takes four iterations to obtain the minimizer of $f(x)$; both $\|\nabla f(x_k)\|$ and $\|s_k\|$ decrease very quickly.

We also ran the regularized Newton algorithm without correction (Algorithm RNA), that is, we do not solve the linear equations (46) and just set the solution of (45) to be the trial step s_k. We first take the same values of α_i, n and x_0 as given in [8]. The results are reported in Table 2. 'niter' represents the number of iterations, '$\|\nabla f_k\|$' represents the final value of $\|\nabla f_k\|$, and x^* the final value of x_k. The result before the sign / is for Algorithm RNA and after for Algorithm 2. If they are the same, we just present one. We also generate x_0 randomly with $\|x_0\|$ being $1, 10^3, 10^6$ and 10^9, respectively. The results are given in Table 3, where 'Ng' represents the number of gradient calculations and 'NH' the number of Hessian calculations. If they are equal, we just present one.

Since the Hessian H is always singular, the Newton method cannot be used to solve nonlinear equations (41). But by using the regularization technique, both Algorithm RNA and Algorithm 2 work quite well. The sequence generated always converges to the minimizer of $f(x)$ in few

Table 2. Results of Algorithm RNA and Algorithm 2.

α_i	n	$(x_0)_i$	niter	$\|\nabla f_k\|$	x_i^*	n	$(x_0)_i$	niter	$\|\nabla f_k\|$	x_i^*
0	10	i	3/2	6.5e − 10/1.8e − 09	5.5	50	i	4/3	9.1e − 09/2.2e − 09	25.5
		$n-i$	3/2	6.5e − 10/1.8e − 09	4.5		$n-i$	4/3	9.1e − 09/2.2e − 09	24.5
		$1/i$	2	4.6e − 07/6.7e − 15	0.2929		$1/i$	2	5.0e − 06/3.7e − 10	0.0902/0.0899
	100	i	5/4	6.5e − 12/4.4e − 14	50.5	200	i	5/4	2.3e − 07/3.8e − 08	100.5
		$n-i$	5/4	6.5e − 12/4.4e − 14	49.5		$n-i$	5/4	2.3e − 07/3.8e − 08	99.5
		$1/i$	2	7.0e − 06/2.4e − 09	0.0529/0		$1/i$	2	8.0e − 06/1.0e − 08	0.0321/0.0294
	500	i	6/5	6.3e − 08/2.4e − 09	250.5	1000	i	7/6	8.9e − 10/1.5e − 12	500.5
		$n-i$	6/5	6.3e − 08/2.4e − 09	249.5		$n-i$	7/6	8.9e − 10/1.6e − 12	499.5
		$1/i$	2	8.4e − 06/4.7e − 08	0.0214/0		$1/i$	2	8.4e − 06/1.1e − 07	0.0202/0.0080
1	10	i	4	3.4e − 12/2.5e − 15	5.5	50	i	5/4	3.5e − 12/3.8e − 08	25.5
		$n-i$	4	3.4e − 12/2.5e − 15	4.5		$n-i$	5/4	3.5e − 12/3.8e − 08	24.5
		$1/i$	3	7.9e − 11/7.3e − 12	0.2929		$1/i$	3	4.6e − 09/7.6e − 12	0.0900
	100	i	5	7.2e − 08/2.6e − 13	50.5	200	i	6/5	9.7e − 11/6.0e − 08	100.5
		$n-i$	5	7.2e − 08/2.6e − 13	49.5		$n-i$	6/5	9.7e − 11/6.0e − 08	99.5
		$1/i$	3	1.9e − 08/7.8e − 12	0.0519		$1/i$	3	4.7e − 08/1.5e − 11	0.0294
	500	i	7/6	5.2e − 12/2.9e − 10	250.5	1000	i	7/6	2.7e − 07/6.3e − 06	500.5
		$n-i$	7/6	5.2e − 12/2.9e − 10	249.5		$n-i$	7/6	2.7e − 07/6.3e − 06	499.5
		$1/i$	3	7.6e − 08/1.9e − 10	0.0137		$1/i$	3	8.7e − 08/7.5e − 10	0.0078/0.0075
i	10	i	5	1.1e − 06/7.3e − 07	5.5	50	i	7	4.6e − 06/1.3e − 07	25.5
		$n-i$	5	1.1e − 06/7.3e − 07	4.5		$n-i$	7	4.6e − 06/1.3e − 07	24.5
		$1/i$	3	8.0e − 11/7.3e − 12	0.2929		$1/i$	3	4.5e − 09/7.6e − 12	0.0900
	100	i	9/8	1.2e − 10/7.7e − 07	50.5	200	i	10	6.7e − 08/1.2e − 11	100.5
		$n-i$	9/8	1.2e − 10/7.7e − 07	49.5		$n-i$	10	6.7e − 08/1.2e − 11	99.5
		$1/i$	3	1.9e − 08/8.0e − 12	0.0519		$1/i$	3	4.6e − 08/3.9e − 11	0.0294
	500	i	12/11	7.5e − 09/2.1e − 06	250.5	1000	i	13	1.1e − 06/9.3e − 10	500.5
		$n-i$	12/11	7.5e − 09/2.1e − 06	249.5		$n-i$	13	1.1e − 06/9.3e − 10	499.5
		$1/i$	3	7.6e − 08/2.0e − 10	0.0137		$1/i$	3	8.6e − 08/7.5e − 10	0.0078/0.0075

Table 3. Comparisons between RNA and Algorithm 2 with different x_0.

			RNA			Algorithm 2		
n	α_i	$\|x_0\|$	Ng/NH	$\|\nabla f_k\|$	x_i^*	Ng/NH	$\|\nabla f_k\|$	x_i^*
10	0	1	3	2.57e − 06	0.050754	3	4.59e − 13	0.050746
		10^3	7	1.05e − 07	50.746	6	1.90e − 11	50.746
		10^6	13	0	50,746	11	1.30e − 09	50,746
		10^9	33	0	50,746,000	24	0	50,746,000
	1	1	4	1.23e − 08	0.050746	4	6.20e − 09	0.050746
		10^3	21	1.96e − 12	50.746	20	4.91e − 08	50.746
		10^6	43	0	50,746	42	1.03e − 10	50,746
		10^9	78	0	50,746,000	68	3.60e − 06	50,746,000
	i	1	5	4.08e − 09	0.050746	5	2.96e − 09	0.050746
		10^3	23	2.01e − 14	50.746	22	1.30e − 09	50.746
		10^6	44	3.14e − 06	50,746	44	0	50,746
		10^9	79	0	50,746,000	70	1.05e − 08	50,746,000
100	0	1	3	8.81e − 06	−0.0068268	3	2.30e − 10	−0.0068912
		10^3	7	9.36e − 06	−6.8915	6	1.08e − 06	−6.8913
		10^6	13	1.61e − 10	−6891.2	12	0	−6891.2
		10^9	36	6.84e − 09	−6,891,200	26	0	−6,891,200
	1	1	4	3.12e − 10	−0.0068912	4	2.39e − 14	−0.0068912
		10^3	19	3.54e − 06	−6.8912	19	1.98e − 11	−6.8912
		10^6	41	4.22e − 07	−6891.2	41	0	−6891.2
		10^9	122/104	3.86e − 06	−6,891,200	118/98	7.92e − 06	−6,891,200
	i	1	6	7.52e − 08	−0.0068913	6	7.68e − 08	−0.0068912
		10^3	24	1.27e − 07	−6.8912	24	1.93e − 13	−6.8912
		10^6	46	1.71e − 08	−6891.2	45	1.27e − 06	−6891.2
		10^9	126/109	2.63e − 09	−6,891,200	140/115	0	−6,891,200
1000	0	1	4	1.37e − 10	0.0015583	3	1.28e − 08	0.0015437
		10^3	8	1.63e − 10	1.5583	7	1.11e − 13	1.5583
		10^6	13	4.01e − 07	1558.3	12	2.11e − 11	1558.3
		10^9	39	3.59e − 08	1.558300	28	0	1,558,300
	1	1	4	2.32e − 10	0.0015582	3	2.89e − 07	0.0015855
		10^3	17	2.25e − 11	1.5583	16	3.26e − 07	1.5583
		10^6	50/47	5.13e − 09	1558.3	40	9.03e − 06	1558.3
		10^9	802/543	6.46e − 08	1.558300	778/516	3.29e − 10	1,558,300
	i	1	7	3.07e − 10	0.0015583	7	2.92e − 10	0.0015583
		10^3	25	7.32e − 13	1.5583	24	2.80e − 08	1.5583
		10^6	308/211	3.88e − 07	1558.3	229/162	2.00e − 07	1558.3
		10^9	791/526	4.53e − 09	1.558300	777/506	5.79e − 06	1,558,300

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

iterations, whatever the initial point is chosen, though different initial points may lead to different minimizers. However, for all the same n, x_0 and different α, Algorithm 2 converges to the same minimizer of $f(x)$. That happens to Algorithm RNA too. The reason may be that the first part of the Hessian $\nabla^2 f(x)$ is a constant, while the second part of $\nabla^2 f(x)$ at all minimizers is zero, which is independent of α.

Moreover, we can see that for the same α, n and x_0, the number of iterations of Algorithm 2 is always less than or equal to that of Algorithm RNA. Even if they are equal, the final norm of the gradient obtained in Algorithm 2 is always less than that in Algorithm RNA. Though it seems that the advantage of Algorithm 2 over Algorithm RNA is not significant when the initial point is very close to the minimizer of $f(x)$, Table 3 shows that the correction term does help to improve Algorithm 2 when the initial point is far away from the minimizer. These facts indicate that the introduction of correction is really useful and could accelerate the convergence of the regularized Newton method.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013

6. Concluding remarks

We propose a regularized Newton method with correction for monotone nonlinear equations. The technique of regularization is not new; it is used to overcome the possible singularity of the Jacobian. However, the idea of correction is novel; it has the advantages of compensating for the shorter trial step because of the regularization and needs just little additional calculations. Under the local error bound condition which is weaker than nonsingularity, we show that the method achieves the same quadratic convergence as the Newton method.

An unconstrained convex nonlinear optimization problem with singular Hessian at the solutions can be transformed to a special case of monotone nonlinear equations. Therefore, we applied the regularized Newton method with correction to solve these problems and developed a globally convergent algorithm using a trust region technique. The numerical results in Section 5 suggest that Algorithm 2 appears more efficient than Algorithm RNA, which does not perform the correction process.

Acknowledgements

The first author is supported by Chinese NSF grants 10871127 and 11171217, and the second author is supported by Chinese NSF grants 10831006, 11021101, and CAS grant kjcx-yw-s7.

References

[1] H. Dan, N. Yamashita, and M. Fukushima, *A superlinearly convergent algorithm for the monotone nonlinear complementarity problem without uniqueness and nondegeneracy conditions*, Math. Oper. Res. 27 (2002), pp. 743–754.

[2] J.Y. Fan and Y.X. Yuan, *On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption*, Computing 74 (2005), pp. 23–39.

[3] M. Fukushima, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Mathematical Programming Ser. A 53 (1992), pp. 99–110.

[4] W.W. Hager and H. Zhang, *Asymptotic convergence analysis of a new class of proximal point methods*, SIAM J. Control Optim. 46 (2007), pp. 1683–1704.

[5] W.W. Hager and H. Zhang, *Self-adaptive inexact proximal point methods*, Comput. Optim. Appl. 39 (2008), pp. 161–181.

[6] C.T. Kelley, *Solving Nonlinear Equations with Newton's Method*, Fundamentals of Algorithm, SIAM, Philadelphia, PA, 2003.

[7] K. Levenberg, *A method for the solution of certain nonlinear problems in least squares*, Quart. Appl. Math. 2 (1944), pp. 164–166.

[8] D.-H. Li, M. Fukushima, L. Qi, and N. Yamashita, *Regularized Newton methods for convex minimization problems with singular solutions*, Comput. Optim. Appl. 28 (2004), pp. 131–147.

[9] D.W. Marquardt, *An algorithm for least-squares estimation of nonlinear inequalities*, SIAM J. Appl. Math. 11 (1963), pp. 431–441.

[10] J.J. Moré, *The Levenberg–Marquardt algorithm: Implementation and theory*, in *Lecture Notes in Mathematics 630: Numerical Analysis*, G.A. Watson, ed., Springer-Verlag, Berlin, 1978, pp. 105–116.

[11] M.J.D. Powell, *Convergence properties of a class of minimization algorithms*, in *Nonlinear Programming 2*, O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, eds., Academic Press, New York, 1975, pp. 1–27.

[12] D. Sun, *A regularization Newton method for solving nonlinear complementarity problems*, Appl. Math. Optim. 40 (1999), pp. 315–339.

[13] P. Tseng, *Error bounds and superlinear convergence analysis of some Newton-type methods in optimization, in Nonlinear Optimization and Related Topics*, G.D. Pillo and F. Giannessi, eds., Kluwer Academic Publishers, Norwell, MA, 2000, pp. 445–462.

[14] N. Yamashita and M. Fukushima, *The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem*, SIAM J. Optim. 11 (2000), pp. 364–379.

[15] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg–Marquardt method*, Computing 15 (Suppl) (2001), pp. 237–249.

[16] Y.X. Yuan, *Trust region algorithms for nonlinear equations*, Information 1 (1998), pp. 7–20.

[17] Y.X. Yuan, *Problems on convergence of unconstrained optimization algorithms*, in *Numerical Linear Algebra and Optimization*, Y.X. Yuan, ed., Science Press, Beijing, 1999, pp. 95–107.

[18] Y.X. Yuan, *A review of trust region algorithms for optimization*, in *ICM99: Proceedings of the Fourth International Congress on Industrial and Applied Mathematics*, J.M. Ball and J.C.R. Hunt, eds., Oxford University Press, Edinburgh, 2000, pp. 271–282.

[19] Y.B. Zhao and D. Li, *Monotonicity of fixed point and normal mapping associated with variational inequality and its applications*, SIAM J. Optim. 11 (2001), pp. 962–973.

[20] G. Zhou and L. Qi, *On the convergence of an inexact Newton-type method*, Oper. Res. Lett. 34 (2006), pp. 647–652.

[21] G. Zhou and K.C. Toh, *Superlinear convergence of a Newton-type algorithm for monotone equations*, J. Optim. Theory Appl. 125 (2005), pp. 205–221.

Downloaded by [Academy of Mathematics and System Sciences] at 03:47 06 February 2013