Journal of Computational Mathematics
Vol.29, No.3, 2011, 261–286.
http://www.global-sci.org/jcm
doi:10.4208/jcm.1009-m3216

A SMOOTHING TRUST REGION METHOD FOR NCPS BASED ON THE SMOOTHING GENERALIZED FISCHER-BURMEISTER FUNCTION*

Xuebin Wang Changfeng Ma Meiyan Li
School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China
School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin 541004, China
Email: macf@fjnu.edu.cn

Abstract

Based on a reformulation of the complementarity problem as a system of nonsmooth equations by using the generalized Fischer-Burmeister function, a smoothing trust region algorithm with line search is proposed for solving general (not necessarily monotone) nonlinear complementarity problems. Global convergence and, under a nonsingularity assumption, local Q-superlinear/Q-quadratic convergence of the algorithm are established. In particular, it is proved that a unit step size is always accepted after a finite number of iterations. Numerical results also confirm the good theoretical properties of our approach.

Mathematics subject classification: 90C33, 90C30.
Key words: Nonlinear complementarity problem, Smoothing method, Trust region method, Global convergence, Local superlinear convergence.

1. Introduction

Let $F : R^n \to R^n$ be continuously differentiable. The nonlinear complementarity problem, denoted by NCP(F), is to find a vector $x \in R^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad \langle x, F(x) \rangle = 0. \tag{1.1}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. If F is an affine function of x, then NCP(F) reduces to a linear complementarity problem (LCP). The NCP is theoretically and practically useful, and has been used to study and formulate various equilibrium problems in economics and engineering, such as Nash equilibrium problems, traffic equilibrium problems, contact mechanics problems and so on [1, 2].

There have been many methods proposed for the solution of NCP(F). Among them, one of the most popular and powerful approaches that has been studied intensively recently is to reformulate NCP(F) as a system of nonlinear equations [10], as an unconstrained minimization problem using suitable merit functions [7], or as a parametric problem. Here we concentrate on the equation-based method, where the NCP(F) can be written equivalently as

$$\Phi(x) = 0 \tag{1.2}$$

for a suitable equation operator $\Phi : R^n \to R^n$. Recently, for solving complementarity problems, various equivalent equation-based reformulations have been proposed and seem attractive. For

* Received September 27, 2009 / Revised version received March 5, 2010 / Accepted March 22, 2010 / Published online February 28, 2011 /

more details of these reformulations, please see [6,10,12,14,15,19]. Generally the operator Φ is locally Lipschitz continuous but not differentiable, so that we cannot apply the classical Newton method directly to solve Eq. (1.2). Nevertheless, recent research shows that one can still design globally and locally fast convergent algorithms for the solution of Eq. (1.2). Most of these methods are classified by Kanzow and Pieper [20] into three categories: namely, nonsmooth Newton methods, Jacobian smoothing methods and smoothing (Newton) methods. We refer the interested reader to [20] and references therein.

In this paper, we aim to develop a trust region-type method for solving nonsmooth equations (1.2). Trust region methods for solving nonsmooth equations have been studied in [3, 9, 11]. The algorithm proposed in [3] was devoted to solving a semismooth equation reformulation for generalized complementarity problems by adopting the squared natural residual of the semismooth equations as a merit function. Actually, it used a trust region strategy for solving the following unconstrained minimization problem

$$\min_{x\in R^n} \ \widetilde{\Phi}(x),$$

where its merit function $\widetilde{\Phi}(x) := \|\Phi(x)\|_2^2/2$. In their method, the trust region subproblem was the following minimization problem

$$\begin{cases} \min & \nabla\widetilde{\Phi}(x^k)^T d + \dfrac{1}{2} d^T V_k^T V_k d \\ \text{s.t.} & \|d\| \le \Delta_k, \end{cases}$$

where Δ_k was the trust region radius and V_k was an arbitrary element in the Clarke's [13] generalized Jacobian of Φ at x^k. Global convergence and, under a nonsingularity assumption, local Q-superlinear (or Q-quadratic) convergence of this trust region method were established.

Inspired by Jiang *et al.*'s work [3], we develop a smoothing trust region method for solving the NCPs. The method is based on the recently presented smoothing technique which is to construct a smoothing approximation function $\Phi_\mu : R^n \to R^n$ of Φ such that for any $\mu > 0$, Φ_μ is continuously differentiable and

$$\|\Phi(x) - \Phi_\mu(x)\| \to 0, \qquad \text{as } \mu \downarrow 0 \text{ for all } x \in R^n.$$

The parameter μ is called smoothing parameter. Based on the smoothing function Φ_μ, we define a merit function

$$\theta_\mu(x) := \frac{1}{2}\|\Phi_\mu(x)\|^2 \tag{1.3}$$

and propose a trust region method where, at each iterate point x^k, the trial step d^k is obtained by solving the following trust region subproblem,

$$\begin{cases} \min & \Theta(d) := \dfrac{1}{2}\|\Phi_{\mu_k}(x^k) + J\Phi_{\mu_k}(x^k)d\|^2 \\ \text{s.t.} & \|d\| \le \Delta_k. \end{cases} \tag{1.4}$$

And define the actual reduction and the predicted reduction as follows:

$$\begin{aligned} Ared_k &:= \theta_{\mu_k}(x^k) - \theta_{\mu_k}(x^k + d^k), \\ Pred_k &:= \Theta(0) - \Theta(d^k). \end{aligned}$$

The trust region radius Δ_k is updated according to the value of the ratio

$$r_k := \frac{Ared_k}{Pred_k}. \tag{1.5}$$

In the algorithm presented in [3], at each iteration step, if the trial step d^k is not successful, one rejects the trial step, reduces the trust region radius, and resolves the subproblem. This process is repeated until d^k is a sufficiently descent direction of its merit function at the point x^k. Thus, at each iteration step of the algorithm presented in [3], the subproblem needs to be solved many times. This strategy is quite adequate for small problems. However, if the number of variables is large, resolving the trust region problem can be costly.

In contrast to Jiang *et al.*'s work [3], in our approach, we incorporate backtracking line searches in a trust region method, so as to avoid resolving the subproblem when the trial step d^k is not accepted. Specifically, at each iteration step, we only solve the trust region subproblem (1.4) once. If d^k is failed, then we take advantage of a line search strategy to determine a steplength t_k such that $\tilde{d}^k := t_k d^k$ provides a descent direction of the merit function θ_{μ_k} at x^k. The next iterate is obtained by letting $x^{k+1} = x^k + \tilde{d}^k$.

Obviously, the algorithm proposed in this paper enjoys the following advantages:

- It shares the advantages of trust region methods. The trust region subproblem (1.4) always has a solution whether $J\Phi_{\mu_k}(x^k)$ is nonsingular or singular. So our algorithm is well defined for general nonlinear complementarity problem without requiring that F is a P_0 function. It is well known that $J\Phi_{\mu_k}(x^k)$ is nonsingular if F is a P_0 function [17,18].
- At each iteration step, the trust region subproblem (1.4) is solved once only but need not be resolved when the trial step d^k is not accepted. It reduces the computational cost very much.

The remainder of this paper is organized as follows: In the next section, we summarize some preliminary results and important properties of smoothing reformulation equation based on the generalized Fischer-Burmeister function. In section 3, a smoothing trust region method is proposed for solving NCP(F) in detail. We proceed in section 4 by proving the global convergence of this algorithm. In section 5, the locally Q-superlinear/Q-quadratic convergence of the algorithm will be established under a nonsingularity assumption. Extensive and very encouraging numerical results are reported in section 6. Finally, in the last section we make some conclusive remarks.

We close this section by giving a list of the **NOTATION** employed.

All vectors are column vectors, the subscript ^{T} denotes transpose, R^n (respectively, R) denotes the space of n-dimensional real column vectors (respectively, real numbers), R^n_+ and R^n_{++} denote the nonnegative and positive orthants of R^n, respectively. R_+ (respectively, R_{++}) denotes the nonnegative (respectively, positive) orthant in R. We define $N := \{1, 2, \cdots, n\}$ and $\mathfrak{N} := \{0, 1, 2, \cdots\}$.

For any vector $y \in R^n$, we denote by diag$\{y_i : i \in N\}$ the diagonal matrix whose ith diagonal element is y_i and by vec$\{y_i : i \in N\}$ the column vector y. The matrix I represents the identity matrix of arbitrary dimension. e^i indicates the i-th column of the n-dimensional identity matrix.

If $x \in R^n$, we denote by $\|x\|_p$ the p-norm of x and by $\|x\|$ the Euclidean norm of x. For a given $M \in R^{n\times n}$ and a nonempty set of matrices $\mathscr{M} \in R^{n\times n}$, $\|M\|$ denotes the spectral norm of M and $dist(M, \mathscr{M}) := \inf_{B\in\mathscr{M}} \|M - B\|$ denotes the distance of M to $\mathscr{M}$. For any continuously differentiable mapping $G : R^n \to R^m$, the Jacobian of G at x is denoted by $JG(x) \in R^{m\times n}$. If $m = 1$, the symbol ∇G is used for the gradient of G. $\mathrm{sgn}(\cdot)$ is the sign function. Landau symbols $o(\cdot)$ and $O(\cdot)$ are defined in usual way.

2. Reformulation and Preliminaries

Throughout this paper, we assume that F is continuously differentiable on R^n.

A function $\bar{\phi} : R^2 \to R$ is called an NCP-function if it satisfies

$$\bar{\phi}(a, b) = 0 \iff a \geq 0, \quad b \geq 0, \quad ab = 0. \tag{2.1}$$

We concentrate on one particular reformulation of the NCP(F) and fully exploit the (additional) properties of this special reformulation. It is based on the generalized Fischer-Burmeister function [4, 5] , i.e., $\phi : R^2 \to R$ given by

$$\phi(a, b) := \|(a, b)\|_p - (a + b), \tag{2.2}$$

where p is any fixed real number in the interval $(1, +\infty)$ and $\|(a, b)\|_p$ denotes the p-norm of (a, b), i.e., $\|(a, b)\|_p = \sqrt[p]{|a|^p + |b|^p}$. Notice that in the function ϕ, the 2-norm of (a, b) in the original Fischer-Burmeister function is replaced by more generally a p-norm of (a, b) with $p > 1$. This function ϕ is still an NCP-function as was noted in Tseng's paper [8]. Then it is well known and easy to see that the NCP(F) is equivalent to a system of nonsmooth equations

$$\Phi(x) := \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0. \tag{2.3}$$

Its natural merit function $\theta : R^n \to R$ is given by

$$\theta(x) := \frac{1}{2}\|\Phi(x)\|^2. \tag{2.4}$$

Let $\mu > 0$ be the smoothing parameter, then the corresponding smooth operator $\Phi_\mu : R^n \to R^n$ is defined similarly by

$$\Phi_\mu(x) := \begin{pmatrix} \phi_\mu(x_1, F_1(x)) \\ \vdots \\ \phi_\mu(x_n, F_n(x)) \end{pmatrix},$$

where $\phi_\mu : R^2 \to R$ denotes the smooth approximation

$$\phi_\mu(a, b) := \|(a, b, \mu)\|_p - (a + b), \quad \mu > 0, \tag{2.5}$$

of the generalized Fischer-Burmeister function.

Note that Φ is locally Lipschitz on R^n and Fréchet differentiable on the set Ω, where

$$\Omega := \{x \in R^n | (x_i, F_i(x)) \neq 0, i \in N\}.$$

However, Φ is not necessarily differentiable at $x \notin \Omega$. Let $\partial\Phi(x)$ denote the Clarke's generalized Jacobian of Φ at $x \in R^n$ [13], which can be defined as follows:

$$\partial\Phi(x) := \mathrm{conv}\Big\{V \in R^{n\times n} | V = \lim_{x^k \to x} J\Phi(x^k), \Phi \text{ is differentiable at } x^k \text{ for all } k\Big\}.$$

Usually, $\partial\Phi(x)$ is difficult to compute. For the purpose of this paper, the C-subdifferential $\partial_C\Phi$ is considerably more important than the more familiar generalized Jacobian; here

$$\partial_C\Phi(x)^T := \partial\Phi_1(x) \times \cdots \times \partial\Phi_n(x),$$

which was discussed in [21].

Now we summarize some properties of the functions Φ, ϕ_μ and Φ_μ, which will be important for our subsequent analysis. The following result follows directly from Property 2.1(d) in [4] and the definition of the C-subdifferential.

Lemma 2.1. *For any $x \in R^n$, we have*

$$\partial_C\Phi(x) = \widetilde{D}_1(x) + \widetilde{D}_2(x)JF(x),$$

where $\widetilde{D}_1(x) = \mathrm{diag}\{a_i(x) : i \in N\}$, $\widetilde{D}_2(x) = \mathrm{diag}\{b_i(x) : i \in N\}$ with

$$a_i(x) = \frac{\mathrm{sgn}(x_i)|x_i|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1, \quad b_i(x) = \frac{\mathrm{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1,$$

if $(x_i, F_i(x)) \neq 0$, and

$$a_i(x) = \tilde{\xi}_i - 1, \quad b_i(x) = \tilde{\zeta}_i - 1,$$

for every $(\tilde{\xi}_i, \tilde{\zeta}_i) \in R^2$, such that $|\tilde{\xi}_i|^{p/(p-1)} + |\tilde{\zeta}_i|^{p/(p-1)} \le 1$, if $(x_i, F_i(x)) = 0$.

Lemma 2.2. *For all $(a,b) \in R^2$ and all $\mu_1, \mu_2 \ge 0$, the function ϕ_μ defined by (2.5) satisfies*

(i) *when $\mu = 0$, we have ϕ_0 is an NCP-function, i.e.,*

$$\phi_0(a,b) = 0 \iff a \ge 0, \quad b \ge 0, \quad ab = 0;$$

(ii) *if $\mu > 0$, then $\phi_\mu(a,b)$ is continuously differentiable on R^2; and*

$$\nabla\phi_\mu(a,b) = \left(\frac{\mathrm{sgn}(a)|a|^{p-1}}{\|(a,b,\mu)\|_p^{p-1}} - 1, \frac{\mathrm{sgn}(b)|b|^{p-1}}{\|(a,b,\mu)\|_p^{p-1}} - 1\right)^T;$$

(iii) *$|\phi_{\mu_1}(a,b) - \phi_{\mu_2}(a,b)| \le |\mu_1 - \mu_2|$. In particular, we have $|\phi_\mu(a,b) - \phi(a,b)| \le \mu$ for all $\mu > 0$.*

Proof. The conclusions (i) and (ii) can be easily obtained by simple calculation. From the definition of ϕ_μ and the property of p-norm, we have

$$\begin{aligned}|\phi_{\mu_1}(a,b) - \phi_{\mu_2}(a,b)| =& \big|\|(a,b,\mu_1)\|_p - (a,b,\mu_2)\|_p\big| \\ \le& \|(a,b,\mu_1) - (a,b,\mu_2)\|_p \\ =& \|(0,0,\mu_1-\mu_2)\|_p = |\mu_1 - \mu_2|,\end{aligned}$$

which gives (iii). □

As an immediate consequence of Lemma 2.2, we get the following corollary.

Corollary 2.1. *The following properties hold for the smoothing operator Φ_μ.*

(i) *For all $\mu > 0$, $\Phi_\mu(x)$ is continuously differentiable on R^n with its Jacobian*

$$J\Phi_\mu(x) = D_1(\mu, x) + D_2(\mu, x) JF(x)$$

where

$$D_1(\mu, x) = \mathrm{diag}\Big\{ \frac{\mathrm{sgn}(x_i)|x_i|^{p-1}}{\|(x_i, F_i(x), \mu)\|_p^{p-1}} - 1 : i \in N \Big\},$$
$$D_2(\mu, x) = \mathrm{diag}\Big\{ \frac{\mathrm{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i, F_i(x), \mu)\|_p^{p-1}} - 1 : i \in N \Big\}.$$

(ii)
$$\|\Phi_{\mu_1}(x) - \Phi_{\mu_2}(x)\| \le \kappa|\mu_1 - \mu_2|, \tag{2.6}$$

for all $x \in R^n$ and $\mu_1, \mu_2 \ge 0$, where $\kappa := \sqrt{n}$. In particular, we have

$$\|\Phi_\mu(x) - \Phi(x)\| \le \kappa\mu, \tag{2.7}$$

for all $x \in R^n$ and $\mu \ge 0$.

Lemma 2.3. *Let $\mu > 0$ and $p > 1$ be arbitrary but fixed. Then the following statements hold:*

(i)

$$\varpi_1(t) := t^{\frac{(1-p)}{p}} - (t + \mu^p)^{\frac{(1-p)}{p}},$$
$$\varpi_2(t) := (t + \mu^p)^{\frac{(p-1)}{p}} - t^{\frac{(p-1)}{p}}$$

are strictly decreasing in $t > 0$ and in $t \ge 0$, respectively.

(ii) *Furthermore, $\varpi_2(t) \le \mu^{p-1}$ for all $t \ge 0$.*

Proof. (i) The function ϖ_1 is continuously differentiable and

$$\varpi_1'(t) = \frac{1-p}{p}\Big(t^{\frac{(1-2p)}{p}} - (t + \mu^p)^{\frac{(1-2p)}{p}} \Big).$$

Obviously, from $\mu > 0$ and $p > 1$, it follows that $\varpi_1'(t) < 0$ for all $t > 0$. This implies that $\varpi_1(t)$ is strictly decreasing in $t > 0$.

By a similar argument, it is not difficult to see that $\varpi_2(t)$ is strictly decreasing in $t \ge 0$.

(ii) From the statement (i), it is easy to see that $\varpi_2(t) \le \varpi_2(0) = \mu^{p-1}$ for all $t \ge 0$. The proof is complete. ☐

In order to guarantee local fast convergence of our algorithm, we have to control the distance between $J\Phi_\mu(x^k)$ and the set $\partial_C\Phi(x^k)$ at each iteration step x^k as $\mu \downarrow 0$. Note that the set $\partial_C\Phi(x)$ is nonempty and compact for any $x \in R^n$.

Proposition 2.1. *Let $x \in R^n$ be arbitrary but fixed.*

(i) *Then we have*

$$\lim_{\mu\downarrow 0} \mathrm{dist}(J\Phi_\mu(x), \partial_C\Phi(x)) = 0. \tag{2.8}$$

(ii) *Assume that x is not a solution of NCP(F). Let us define the constants*

$$\gamma(x) := \max_{i \notin \beta(x)} \{\|\mathrm{sgn}(x_i)|x_i|^{p-1}e^i + \mathrm{sgn}(F_i(x))|F_i(x)|^{p-1}\nabla F_i(x)\|\} \ge 0,$$
$$\alpha(x) := \min_{i \notin \beta(x)} \{|x_i|^p + |F_i(x)|^p\} > 0,$$

where $\beta(x) := \{i | x_i = F_i(x) = 0\}$*. Let* $\delta > 0$ *be given, and define*

$$\hat{\xi}(x,\delta) := \begin{cases} 1, & \text{if } \left(\frac{\sqrt{n}\gamma(x)}{\delta}\right)^{\frac{p}{(p-1)}} - \alpha(x) \le 0, \\ \alpha(x)^{\frac{2}{p}} \left(\left(\frac{\sqrt{n}\gamma(x)}{\delta}\right)^{\frac{p}{(p-1)}} - \alpha(x) \right)^{-\frac{1}{p}}, & \text{otherwise.} \end{cases} \tag{2.9}$$

Then

$$dist(J\Phi_\mu(x), \partial_C \Phi(x)) \le \delta$$

for all μ *such that* $0 < \mu \le \hat{\xi}(x,\delta)$.

Proof. (i) Define

$$\beta(x) := \{i | x_i = F_i(x) = 0\}.$$

If we denote the i-th component function of Φ_μ by $\Phi_{\mu,i}$, from Corollary 2.1(i), we obtain

$$\lim_{\mu \downarrow 0} \nabla \Phi_{\mu,i}(x) = \begin{cases} \left(\frac{\operatorname{sgn}(x_i)|x_i|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1 \right) e^i + \left(\frac{\operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1 \right) \nabla F_i(x), & \text{for } i \notin \beta(x) \\ -e^i - \nabla F_i(x), & \text{for } i \in \beta(x). \end{cases}$$

Hence the assertion (i) follows from Lemma 2.1 with $(\tilde{\xi}_i, \tilde{\zeta}_i) = 0$ for $i \in \beta(x)$.

(ii) The assertion (i) implies that for any $\delta > 0$, there exists a parameter $\hat{\xi}(x,\delta) > 0$ such that

$$\operatorname{dist}(J\Phi_\mu(x), \partial_C \Phi(x)) \le \delta$$

for all $0 < \mu \le \hat{\xi}(x,\delta)$.

Now we prove (ii), that is, choose such $\hat{\xi}(x,\delta)$ as follows. Note that $N \setminus \beta(x) \neq \emptyset$ since x is not a solution of NCP(F) by assumption. Hence $\alpha(x) > 0$. From Corollary 2.1(i) and Lemma 2.1, we obtain

$$\nabla \Phi_{\mu,i}(x) = \left(\frac{\operatorname{sgn}(x_i)|x_i|^{p-1}}{\|(x_i, F_i(x), \mu)\|_p^{p-1}} - 1 \right) e^i + \left(\frac{\operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i, F_i(x), \mu)\|_p^{p-1}} - 1 \right) \nabla F_i(x)$$

and

$$\partial \Phi_i(x) = \begin{cases} \left(\frac{\operatorname{sgn}(x_i)|x_i|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1 \right) e^i + \left(\frac{\operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i, F_i(x))\|_p^{p-1}} - 1 \right) \nabla F_i(x), & \text{for } i \notin \beta(x) \\ (\tilde{\xi}_i - 1)e^i + (\tilde{\zeta}_i - 1)\nabla F_i(x), & \text{for } i \in \beta(x) \end{cases}$$

respectively, where $(\tilde{\xi}_i, \tilde{\zeta}_i) \in R^2$ denotes any vector such that $|\tilde{\xi}_i|^{p/(p-1)} + |\tilde{\zeta}_i|^{p/(p-1)} \le 1$. We consider the following two cases:

Case 1, $i \in \beta(x)$. Then $(x_i, F_i(x)) = 0$ and therefore

$$\nabla \Phi_{\mu,i}(x) = -e^i - \nabla F_i(x).$$

Hence, taking $(\tilde{\xi}_i, \tilde{\zeta}_i) = 0$, we see that $\nabla \Phi_{\mu,i}(x) \in \partial \Phi_i(x)$ so that

$$dist(\nabla \Phi_{\mu,i}(x), \partial \Phi_i(x)) = 0, \tag{2.10}$$

for all $i \in \beta(x)$.

Case 2, $i \notin \beta(x)$. Then $\partial\Phi_i(x) = \{\nabla\Phi_i(x)\}$. So, by an easy calculation, we have

$$\begin{aligned}
&dist(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)) = \|\nabla\Phi_{\mu,i}(x) - \nabla\Phi_i(x)\| \\
=&\Big\|\Big(\frac{\operatorname{sgn}(x_i)|x_i|^{p-1}}{\|(x_i,F_i(x),\mu)\|_p^{p-1}} - 1\Big)e^i + \Big(\frac{\operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i,F_i(x),\mu)\|_p^{p-1}} - 1\Big)\nabla F_i(x) \\
&- \Big(\frac{\operatorname{sgn}(x_i)|x_i|^{p-1}}{\|(x_i,F_i(x))\|_p^{p-1}} - 1\Big)e^i - \Big(\frac{\operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}}{\|(x_i,F_i(x))\|_p^{p-1}} - 1\Big)\nabla F_i(x)\Big\| \\
=&\Big\|\Big(\frac{1}{\|(x_i,F_i(x),\mu)\|_p^{p-1}} - \frac{1}{\|(x_i,F_i(x))\|_p^{p-1}}\Big)\big(\operatorname{sgn}(x_i)|x_i|^{p-1}e^i + \operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}\nabla F_i(x)\big)\Big\| \\
=&\Big(\frac{1}{\|(x_i,F_i(x))\|_p^{p-1}} - \frac{1}{\|(x_i,F_i(x),\mu)\|_p^{p-1}}\Big)\big\|\operatorname{sgn}(x_i)|x_i|^{p-1}e^i + \operatorname{sgn}(F_i(x))|F_i(x)|^{p-1}\nabla F_i(x)\big\|.
\end{aligned}$$

From the definitions of $\gamma(x)$ and $\alpha(x)$, by using Lemma 2.3 (i), we therefore obtain

$$\begin{aligned}
&\operatorname{dist}\big(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)\big) \le \Big(\alpha(x)^{\frac{(1-p)}{p}} - (\alpha(x)+\mu^p)^{\frac{(1-p)}{p}}\Big)\cdot\gamma(x) \\
=&\frac{\big(\alpha(x)+\mu^p\big)^{\frac{(p-1)}{p}} - \alpha(x)^{\frac{(p-1)}{p}}}{\big(\alpha(x)(\alpha(x)+\mu^p)\big)^{\frac{(p-1)}{p}}}\cdot\gamma(x) \le \frac{\mu^{p-1}}{\big(\alpha(x)(\alpha(x)+\mu^p)\big)^{\frac{(p-1)}{p}}}\cdot\gamma(x),
\end{aligned}$$

where the latter inequality follows from Lemma 2.3 (ii). We now want to show that

$$\frac{\mu^{p-1}}{\big(\alpha(x)(\alpha(x)+\mu^p)\big)^{\frac{(p-1)}{p}}}\cdot\gamma(x) \le \frac{\delta}{\sqrt{n}}, \tag{2.11}$$

for all $0 < \mu \le \hat{\xi}(x,\delta)$, which then implies

$$\operatorname{dist}(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)) \le \frac{\delta}{\sqrt{n}}. \tag{2.12}$$

Obviously, if $\gamma(x) = 0$, then (2.11) holds for $\forall\ \mu > 0$. Thereby we assume that $\gamma(x) > 0$. Then a simple calculation shows that (2.11) is equivalent to

$$\alpha(x)^2 \ge \mu^p\bigg(\Big(\frac{\sqrt{n}\gamma(x)}{\delta}\Big)^{\frac{p}{(p-1)}} - \alpha(x)\bigg). \tag{2.13}$$

Hence, if $(\sqrt{n}\gamma(x)/\delta)^{p/(p-1)} - \alpha(x) \le 0$, inequality (2.11) holds for any $\mu > 0$, in particular for all $\mu \in (0,1]$. Otherwise, from (2.13), we have

$$\mu \le \alpha(x)^{\frac{2}{p}}\bigg(\Big(\frac{\sqrt{n}\gamma(x)}{\delta}\Big)^{\frac{p}{(p-1)}} - \alpha(x)\bigg)^{-\frac{1}{p}} =: \hat{\xi}(x,\delta).$$

Putting together (2.10) and (2.12), we therefore obtain

$$\operatorname{dist}(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)) \le \frac{\delta}{\sqrt{n}}, \qquad \text{for all}\ \ i \in N.$$

Thus

$$\operatorname{dist}(J\Phi_\mu(x), \partial_C\Phi(x)) \le \delta$$

for all $0 < \mu \le \hat{\xi}(x,\delta)$.

We complete the proof of this proposition. □

3. Algorithm

In this section, we give a detailed description of our smoothing trust region method and state some of its elementary properties.

For simplicity, in the remainder of this paper we denote

$$\Phi_k(x) := \Phi_{\mu_k}(x), \quad \theta_k(x) := \frac{1}{2}\|\Phi_k(x)\|^2.$$

Algorithm 3.1. (A Smoothing Trust Region Method)

Step 0 *Initialization: Choose constants* $\lambda, \eta, \alpha \in (0,1)$, $\sigma \in (0, \frac{1}{2})$, $\Delta_{\min} \geq 0$, $\tau > 0, \nu > 0$, $0 < \eta_1 < \eta_2 < 1$, $0 < \alpha_1 < 1 < \alpha_2$, $\Delta_0 > \Delta_{\min}$. *Let* $x^0 \in R^n$ *be an arbitrary point; Set* $\kappa := \sqrt{n}$, $\beta_0 := \|\Phi(x^0)\|$, $\mu_0 = \frac{\alpha}{2\kappa}\beta_0$ *and* $k := 0$.

Step 1 *If* $\|\nabla\theta(x^k)\| = 0$, *STOP.*

Step 2 *Solve the trust region subproblem* (1.4) *to obtain a trial step* d^k, *(approximately).*

Step 3 *Evaluate the reduction ratio* r_k *from* (1.5).

Step 4 *If* $r_k \geq \eta_1$, *then let* $x^{k+1} := x^k + d^k$, *and perform Step 5.*
Otherwise, let i_k *be the smallest nonnegative integer* i *such that*

$$\theta_k(x^k + \lambda^i d^k) \leq \theta_k(x^k) + \sigma\lambda^i \nabla\theta_k(x^k)^T d^k. \tag{3.1}$$

Set $t_k := \lambda^{i_k}$, $x^{k+1} := x^k + t_k d^k$, *and perform Step* 5.

Step 5 *(Updating* Δ_{k+1}*)*

$$\Delta_{k+1} := \begin{cases} \alpha_1\Delta_k, & \text{if} \quad r_k < \eta_1 \\ \max\{\Delta_{\min}, \Delta_k\}, & \text{if} \quad r_k \in [\eta_1, \eta_2) \\ \max\{\Delta_{\min}, \alpha_2\Delta_k\}, & \text{if} \quad r_k \geq \eta_2 \end{cases} \tag{3.2}$$

Step 6 *(Updating* μ_{k+1}*)*

6.1 *If*

$$\|\Phi(x^{k+1})\| \leq \max\left\{\eta\beta_k, \frac{1}{\alpha}\|\Phi(x^{k+1}) - \Phi_k(x^{k+1})\|\right\}, \tag{3.3}$$

then set $\beta_{k+1} := \|\Phi(x^{k+1})\|$ *and choose* μ_{k+1} *satisfying*

$$0 < \mu_{k+1} \leq \min\left\{\frac{\mu_k}{2}, \frac{\alpha}{2\kappa}\beta_{k+1}, \hat{\xi}(x^{k+1}, \nu\beta_{k+1}), \theta(x^{k+1})\right\}, \tag{3.4}$$

where $\hat{\xi}(\cdot,\cdot)$ *is given by* (2.9).

6.2 *If (3.3) does not hold and*

$$\|\nabla\theta_k(x^{k+1})\| \leq \tau\mu_k \tag{3.5}$$

then set $\beta_{k+1} := \beta_k$ *and choose* μ_{k+1} *satisfying*

$$0 < \mu_{k+1} \leq \min\left\{\frac{\mu_k}{2}, \frac{\|\Phi_k(x^k)\| - \|\Phi_k(x^{k+1})\|}{\kappa}\right\}. \tag{3.6}$$

6.3 *If none of the above conditions is met, set*

$$\beta_{k+1} := \beta_k \quad \text{and} \quad \mu_{k+1} := \mu_k.$$

Step 7 *Let k be replaced by $k+1$, and return to Step* 1.

Without loss of generality, we assume that $\|\nabla\theta(x^k)\| \neq 0$ for all $k \in \mathfrak{N}$ in the following convergence analysis.

Remark 1. (i) The application of the function $\hat{\xi}(\cdot,\cdot)$ in the updating rules (3.4) is the crucial condition for superlinear/quadratic convergence of Algorithm 3.1.

(ii) In both updating rules, namely, in (3.4) and (3.6), we reduce μ_k by at least a factor of 1/2. This is reasonable since we want to force μ_k to go to 0. The sequence $\{\mu_k\}$ is monotonically decreasing and bounded from below. And the parameter β_k will be very useful in proving convergence results in the subsequent sections.

Define the index set

$$\mathcal{F} := \{0\} \cup \left\{ k \in \mathfrak{N} \middle| \|\Phi(x^k)\| \leq \max\left\{\eta\beta_{k-1}, \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{k-1}(x^k)\|\right\}\right\}. \tag{3.7}$$

From the construction of the above algorithm and the definition of $\hat{\xi}(\cdot,\cdot)$ in Proposition 2.1, we have

$$dist(J\Phi_k(x^k), \partial_C\Phi(x^k)) \leq \nu\beta_k = \nu\|\Phi(x^k)\| \tag{3.8}$$

for all $k \in \mathcal{F}$ with $k \geq 1$.

We conclude that the backtracking line search (3.1) should be performed provided the search direction is sufficiently downhill. In order to do this, we apply the algorithm proposed by Nocedal and Yuan, i.e., Algorithm 2.6 in [23], to approximately solve the subproblem (1.4) to obtain the trial step d^k. We show that the trial step d^k is always a direction of sufficient descent for the merit function θ_μ; see Lemma 3.1 below.

We present Algorithm 2.6 in [23] as follows.

Algorithm 3.2. (Algorithm for approximate solution of the subproblem (1.4)**)**
Input: x^k, μ_k, Δ_k. ***Output***: d^k.

$$g_k := J\Phi_{\mu_k}(x^k)^T\Phi_{\mu_k}(x^k), \quad B_k := J\Phi_{\mu_k}(x^k)^T J\Phi_{\mu_k}(x^k).$$

Step 1 *Given constants $\gamma > 1$ and $\epsilon > 0$, set $\hat{\lambda} := 0$.*
If B_k is positive definite, go to Step 2;
else find $\hat{\lambda} \in \left[0, \|B_k\| + (1+\epsilon)\|g_k\|/\Delta_k\right]$ such that $B_k + \hat{\lambda}I$ is positive definite.
Step 2 *Factorize $B_k + \hat{\lambda}I = R_k^T R_k$, where R_k is upper triangular, and solve $R_k^T R_k d = -g_k$ for d^k.*

Step 3 If $\|d^k\| \le \Delta_k$, stop; else solve $R_k^T q = d^k$ for q_k, and compute

$$\hat{\lambda} := \hat{\lambda} + \frac{\|d^k\|^2}{\|q_k\|^2} \frac{\gamma\|d^k\| - \Delta_k}{\Delta_k},$$

go to Step 2.

Note that, in the subproblem (1.4), the quadratic model

$$\begin{aligned}\Theta(d) =& \frac{1}{2}\|\Phi_k(x^k)\|^2 + \left(J\Phi_k(x^k)^T\Phi_k(x^k)\right)^T d + \frac{1}{2} d^T J\Phi_k(x^k)^T J\Phi_k(x^k) d \\ =& \theta_k(x^k) + \nabla\theta_k(x^k)^T d + \frac{1}{2} d^T J\Phi_k(x^k)^T J\Phi_k(x^k) d \\ =& \theta_k(x^k) + g_k^T d + \frac{1}{2} d^T B_k d.\end{aligned}$$

Therefore, we obtain the following result, and omit its proof here since it have been discussed in detail by Nocedal and Yuan in [23].

Lemma 3.1. *If the trial step d^k is computed at each iteration by Algorithm 3.2, then there exists a constant $\bar{\tau} > 0$ such that*

$$Pred_k := \Theta(0) - \Theta(d^k) \ge \bar{\tau}\|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}, \tag{3.9}$$

$$g_k^T d^k \le -\bar{\tau}\|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}, \tag{3.10}$$

where g_k and B_k are defined in Algorithm 3.2.

Remark 2. (i) Note that, inequality (3.10) particularly implies that the trial step d^k at each iteration provides a sufficiently descent direction of the function θ_k at x^k. Therefore for each k, there exists a finite nonnegative integer i_k such that (3.1) holds, which shows the well-definiteness of Algorithm 3.1.

(ii) The property (3.9) of the predicted reduction is very important for the proof of our main global convergence result, Theorem 4.1 below.

4. Global Convergence

The aim of this section is to show the global convergence of Algorithm 3.1, that is, any accumulation point of a sequence generated by Algorithm 3.1 is a stationary point of the merit function θ.

Applying Lemma 2.1, we first obtain the following result whose proof can be carried out in a similar way as the proof of Proposition 3.4 of [6].

Proposition 4.1. *The merit function θ defined by (2.4) is continuously differentiable everywhere and its gradient $\nabla\theta(x) = V^T\Phi(x)$ for an arbitrary $V \in \partial_C\Phi(x)$.*

The following technical result will be utilized in our global convergence analysis.

Lemma 4.1. *Let $\{x^k\} \subseteq R^n$ and $\{\mu_k\} \subseteq R_{++}$ be two sequences with $\{x^k\} \to x^*(\in R^n)$ and $\{\mu_k\} \downarrow 0$. Then*

$$\lim_{k\to\infty} \nabla\theta_k(x^k) = \nabla\theta(x^*).$$

Proof. For an arbitrary $V \in \partial_C \Phi(x^*)$, we obtain from Proposition 4.1

$$\nabla\theta(x^*) = V^T\Phi(x^*) = \sum_{i=1}^n \phi(x_i^*, F_i(x^*))V_i^T,$$

where V_i^T stands for the i-th column of the matrix V^T. On the other hand, obviously θ_μ is differentiable for all $\mu > 0$, so we have

$$\nabla\theta_k(x^k) = J\Phi_k(x^k)^T\Phi_k(x^k) = \sum_{i=1}^n \phi_{\mu_k}(x_i^k, F_i(x^k))\nabla\Phi_{k,i}(x^k),$$

where $\Phi_{k,i}$ denotes the i-th component function of Φ_k.

Let $\beta(x^*) := \{i | x_i^* = F_i(x^*) = 0\}$. We distinguish two cases:

Case 1, $i \notin \beta(x^*)$. Then the generalized Fischer-Burmeister function ϕ is continuously differentiable at $(x_i^*, F_i(x^*))$, and the i-th column of V^T is single valued and equal to $\nabla\Phi_i(x^*)$ from Lemma 2.1. It follows from the continuity of ϕ and JF that

$$\lim_{k\to\infty} \phi_{\mu_k}(x_i^k, F_i(x^k))\nabla\Phi_{k,i}(x^k) = \phi(x_i^*, F_i(x^*))\nabla\Phi_i(x^*) = \phi(x_i^*, F_i(x^*))V_i^T.$$

Case 2, $i \in \beta(x^*)$. From Lemma 2.2, we know

$$\frac{\partial\phi_\mu(a,b)}{\partial a} \in (-2, 0) \text{ and } \frac{\partial\phi_\mu(a,b)}{\partial b} \in (-2, 0)$$

for $\forall\ (a, b) \in R^2$ and $\mu > 0$. So the sequence $\{\nabla\Phi_{k,i}(x^k)\}$ is bounded as $k \to \infty$. Since

$$\lim_{k\to\infty} \phi_{\mu_k}(x_i^k, F_i(x^k)) = \phi(x_i^*, F_i(x^*)) = 0,$$

we get

$$\lim_{k\to\infty} \phi_{\mu_k}(x_i^k, F_i(x^k))\nabla\Phi_{k,i}(x^k) = 0.$$

Obviously, $\phi(x_i^*, F_i(x^*))V_i^T = 0$ for $\forall\ i \in \beta(x^*)$.

Hence, the statement follows from Case 1 and 2. The proof is completed. □

Next we will show that all the iterates x^k generated by our new algorithm remain in a certain level set. The following simple results are exploited.

Lemma 4.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1. Then we have*

(i)

$$\|\Phi_k(x^{k+1})\| < \|\Phi_k(x^k)\|, \qquad \textit{for all } k \in \mathfrak{N}.$$

(ii)

$$\|\Phi_k(x^k)\| + \kappa\mu_k < \|\Phi_{k-1}(x^{k-1})\| + \kappa\mu_{k-1}, \qquad \textit{for all } k \in \mathfrak{N} \setminus \mathcal{F}.$$

Proof. (i) We consider two cases.

Case 1, $r_k \geq \eta_1$. Then, from (3.9) in Lemma 3.1, it follows that

$$\theta_k(x^k) - \theta_k(x^{k+1}) \geq \eta_1 Pred_k > 0;$$

Case 2, $r_k < \eta_1$. From the line search rule (3.1) and Remark 2 (i), we get

$$\theta_k(x^{k+1}) < \theta_k(x^k).$$

Obviously, Case 1 and Case 2 imply $\|\Phi_k(x^{k+1})\| < \|\Phi_k(x^k)\|$ for all $k \in \mathfrak{N}$.

(ii) First assume that $k \in \mathfrak{N} \setminus \mathcal{F}$ and the updating rule (3.6) is not active, then we have $\mu_k = \mu_{k-1}$. This equality, together with statement (i), gives

$$\|\Phi_k(x^k)\| + \kappa\mu_k = \|\Phi_{k-1}(x^k)\| + \kappa\mu_{k-1} < \|\Phi_{k-1}(x^{k-1})\| + \kappa\mu_{k-1}.$$

On the other hand, if $k \in \mathfrak{N} \setminus \mathcal{F}$ and the updating rule (3.6) is active, then it holds that

$$\mu_k \le \frac{\|\Phi_{k-1}(x^{k-1})\| - \|\Phi_{k-1}(x^k)\|}{\kappa}.$$

Therefore, from Corollary 2.1 (ii), it follows that

$$\begin{aligned}
\|\Phi_k(x^k)\| + \kappa\mu_k &\le \|\Phi_k(x^k)\| - \|\Phi_{k-1}(x^k)\| + \|\Phi_{k-1}(x^{k-1})\| \\
&\le \|\Phi_k(x^k) - \Phi_{k-1}(x^k)\| + \|\Phi_{k-1}(x^{k-1})\| \\
&\le \kappa(\mu_{k-1} - \mu_k) + \|\Phi_{k-1}(x^{k-1})\| \\
&< \kappa\mu_{k-1} + \|\Phi_{k-1}(x^{k-1})\|,
\end{aligned}$$

where the last inequality follows from $\kappa\mu_k > 0$. This completes the proof. □

Lemma 4.3. *Assume that the index set $\mathcal{F}$ defined by* (3.7) *consists of $k_0 = 0 < k_1 < k_2 < \cdots$. Let $k \in \mathfrak{N}$ be an arbitrary but fixed index and k_j the largest number in $\mathcal{F}$ such that $k_j \le k$. Then the following three statements hold:*

(i) $\mu_k \le \mu_{k_j}$ *and* $\beta_k = \beta_{k_j}$.

(ii) *Let* $\bar\rho := \max\{\eta, \frac{1}{2}\}$, *then we have*

$$\beta_{k_j} \le \bar\rho^j \|\Phi(x^0)\| \quad \text{and} \quad \mu_{k_j} \le \left(\frac{1}{2}\right)^j \frac{\alpha}{2\kappa}\|\Phi(x^0)\|.$$

(iii) $\|\Phi(x^k)\| < \beta_{k_j} + 2\kappa\mu_{k_j}$.

Proof. (i) Statement (i) follows immediately from the updating rules for μ_k and β_k in Step 6 of Algorithm 3.1.

(ii) Let us denote

$$\begin{aligned}
\mathcal{F}_1 &:= \left\{k \in \mathcal{F} \middle| \eta\beta_{k-1} \ge \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{k-1}(x^k)\|\right\}, \\
\mathcal{F}_2 &:= \left\{k \in \mathcal{F} \middle| \eta\beta_{k-1} < \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{k-1}(x^k)\|\right\}.
\end{aligned}$$

Then $\mathcal{F} = \{0\} \cup \mathcal{F}_1 \cup \mathcal{F}_2$.

If $j = 0$, we get $k_0 = 0$ and thus

$$\beta_{k_0} = \beta_0 = \|\Phi(x^0)\| \quad \text{and} \quad \mu_{k_0} = \mu_0 = \frac{\alpha}{2\kappa}\beta_0 = \frac{\alpha}{2\kappa}\|\Phi(x^0)\|$$

by the definitions of β_0 and μ_0.

If $j \ge 1$, by Step 6 of Algorithm 3.1, if $k_j \in \mathcal{F}_1$, then

$$\beta_{k_j} \le \eta\beta_{k_j - 1} = \eta\beta_{k_{j-1}},$$

where the last equality follows from statement (i) (notice that $k_{j-1} \le k_j - 1$); and if $k_j \in \mathcal{F}_2$, using Corollary 2.1 (ii), then

$$\beta_{k_j} \le \frac{1}{\alpha}\|\Phi(x^{k_j}) - \Phi_{k_j-1}(x^{k_j})\| \le \frac{\kappa}{\alpha}\mu_{k_j-1} \le \frac{\kappa}{\alpha}\mu_{k_{j-1}} \le \frac{1}{2}\beta_{k_{j-1}},$$

where the third inequality follows from statement (i) (notice that $k_{j-1} \le k_j - 1$) and the last inequality follows from the updating rule (3.4). Then, letting $\bar{\rho} = \max\{\eta, \frac{1}{2}\}$, we have

$$\beta_{k_j} \le \bar{\rho}\beta_{k_{j-1}}.$$

Also, from the updating rule (3.4), we get

$$\mu_{k_j} \le \frac{1}{2}\mu_{k_j-1} \le \frac{1}{2}\mu_{k_{j-1}}.$$

Therefore, $\beta_{k_j} \le \bar{\rho}^j\beta_{k_0} = \bar{\rho}^j\|\Phi(x^0)\|$ and $\mu_{k_j} \le \left(\frac{1}{2}\right)^j\mu_{k_0} = \left(\frac{1}{2}\right)^j\frac{\alpha}{2\kappa}\|\Phi(x^0)\|$ from the definitions of β_0 and μ_0. This completes the proof of statement (ii).

(iii) If $k_j = k$, statement (iii) obviously holds since $\beta_{k_j} = \|\Phi(x^{k_j})\| = \|\Phi(x^k)\|$ in this case. Thus we consider $k_j < k$ in the following. Denote the index set

$$\mathcal{L} := \{\ \bar{k} \mid k_j + 1 \le \bar{k} \le k\}.$$

Obviously, $\mathcal{L} \ne \emptyset$. From the definition of k_j (note that $k < k_{j+1}$), it follows that

$$\mathcal{L} \subseteq \mathfrak{N} \setminus \mathcal{F}$$

Thereby, from Lemma 4.2 (ii), we obtain

$$\|\Phi_{\bar{k}}(x^{\bar{k}})\| + \kappa\mu_{\bar{k}} < \|\Phi_{\bar{k}-1}(x^{\bar{k}-1})\| + \kappa\mu_{\bar{k}-1}$$

for all $\bar{k} \in \mathcal{L}$. It is equivalent to

$$\|\Phi_{\bar{k}+1}(x^{\bar{k}+1})\| + \kappa\mu_{\bar{k}+1} < \|\Phi_{\bar{k}}(x^{\bar{k}})\| + \kappa\mu_{\bar{k}} \tag{4.1}$$

for all $\bar{k}$ satisfying $k_j \le \bar{k} \le k - 1$. The above inequality, together with Corollary 2.1 (ii), gives

$$\begin{aligned}
\|\Phi(x^k)\| \le& \|\Phi(x^k) - \Phi_k(x^k)\| + \|\Phi_k(x^k)\| \\
\le& \kappa\mu_k + \|\Phi_k(x^k)\| < \kappa\mu_{k-1} + \|\Phi_{k-1}(x^{k-1})\| \\
<& \cdots \\
<& \kappa\mu_{k_j} + \|\Phi_{k_j}(x^{k_j})\| \\
\le& \kappa\mu_{k_j} + \|\Phi_{k_j}(x^{k_j}) - \Phi(x^{k_j})\| + \|\Phi(x^{k_j})\| \\
\le& \kappa\mu_{k_j} + \kappa\mu_{k_j} + \|\Phi(x^{k_j})\| = 2\kappa\mu_{k_j} + \beta_{k_j},
\end{aligned}$$

where the dots indicate the repeated use of (4.1). That is, statement (iii) holds. □

Proposition 4.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1). Then the following two assertions hold:*

(i) *Denote the level set*

$$\mathscr{L}_0 := \{x \in R^n | \theta(x) \le (1+\alpha)^2\theta(x^0)\},$$

we have

$$\{x^k\} \subseteq \mathscr{L}_0.$$

(ii) *If the index set $\mathcal{F}$ is infinite, then each accumulation point of the sequence $\{x^k\}$ must be a solution of NCP(F).*

Proof. (i) Let k_j and $\bar{\rho}$ be defined in Lemma 4.3. Then $\frac{1}{2} \leq \bar{\rho} < 1$. Therefore, from Lemma 4.3 (ii) and (iii), it follows that

$$\begin{aligned} \|\Phi(x^k)\| <& \beta_{k_j} + 2\kappa\mu_{k_j} \\ \leq& \bar{\rho}^j \|\Phi(x^0)\| + \left(\frac{1}{2}\right)^j \alpha\|\Phi(x^0)\| \\ \leq& \bar{\rho}^j (1+\alpha)\|\Phi(x^0)\| \\ \leq& (1+\alpha)\|\Phi(x^0)\|. \end{aligned} \tag{4.2}$$

The above inequality implies that $x^k \in \mathscr{L}_0$ for all $k \in \mathfrak{N}$. That is, assertion (i) holds.

(ii) Let x^* be a limit point of $\{x^k\}$ and $\{x^k\}_K \to x^*$ where $K \subseteq \mathfrak{N}$. The inequality (4.2) indicates that for all $k \in \mathfrak{N}$, we have

$$\|\Phi(x^k)\| \leq \bar{\rho}^j (1+\alpha)\|\Phi(x^0)\|.$$

Since the index set $\mathcal{F}$ is infinite, then from the definition of j, $k \to \infty$ obviously implies that $j \to \infty$. Therefore,

$$\|\Phi(x^*)\| = \lim_{k(\in K)\to\infty} \|\Phi(x^k)\| \leq \lim_{j\to\infty} \bar{\rho}^j (1+\alpha)\|\Phi(x^0)\| = 0,$$

where the last equality follows from $\frac{1}{2} \leq \bar{\rho} < 1$. Thus x^* is a solution of NCP(F). $\square$

We complete the proof.

Remark 3. If F is a uniform P-function or, more generally, an R_0-function, then the level set $\mathscr{L}_0$ is compact.

Note that Proposition 4.2 (ii) provides a sufficient condition for an accumulation point to be a solution of the complementarity problem NCP(F).

Now we can prove the main global convergence result for Algorithm 3.1.

Theorem 4.1. *Let $\{x^k\}$ be a sequence generated by Algorithm* 3.1. *Then each accumulation point of the sequence $\{x^k\}$ is a stationary point of $\theta(x)$.*

Proof. Let x^* be an accumulation point of the sequence $\{x^k\}$ and $\{x^k\}_K$ be a subsequence converging to x^*. Since the sequence $\{\mu_k\}$ is monotonically decreasing and bounded from below, it converges to some $\mu_* \geq 0$.

If the index set $\mathcal{F}$ is infinite, by Proposition 4.2 (ii), it follows that x^* is a solution of NCP(F), namely $\|\Phi(x^*)\| = 0$, which implies that $\nabla\theta(x^*) = 0$, i.e., x^* is a stationary point of $\theta(x)$. Hence we can suppose that $\mathcal{F}$ contains only finitely many indices. Without loss of generality, we assume that $\mathcal{F} \cap K = \emptyset$ in the remaining part of this proof.

Define the index set

$$K_1 := \left\{ k \in K \,\middle|\, \text{inequality (3.5) is satisfied for } k \right\}.$$

(a) We claim that the set $K_2 := K \setminus K_1$ contains only finitely many indices. Suppose that this assertion does not hold, namely K_2 is infinite. Then by Step 6 of Algorithm 3.1, there exists an integer $\hat{k} \in K_2$ such that for all $k(\in K_2) \geq \hat{k}$,

$$\mu_k = \mu_* > 0 \quad \text{and} \quad \|\nabla\theta_k(x^{k+1})\| > \tau\mu_k.$$

That is,

$$\|\nabla\theta_*(x^{k+1})\| > \tau\mu_* > 0,$$

for all $k(\in K_2) \geq \hat{k}$. Let g_k be defined as in Algorithm 3.2, then

$$\|g_k\| > \tau\mu_* > 0, \tag{4.3}$$

for all $k(\in K_2) \geq \hat{k}$.

Since the sequence $\{J\Phi_k(x^k)\}_K$ is obviously bounded on the convergent sequence $\{x^k\}_K$, then there must exist a constant $C_1 > 0$ such that

$$\|B_k\| \leq \|J\Phi_k(x^k)\|^2 \leq C_1, \tag{4.4}$$

for all $k \in K$, where B_k is defined in Algorithm 3.2.

From Lemma 4.2 (i) and $\theta_\mu(x) = \frac{1}{2}\|\Phi_\mu(x)\|^2$, it follows that $\{\theta_*(x^k)\}$ is monotonically decreasing for all $k(\in K_2) \geq \hat{k}$ and is bounded from below, and is therefore convergent. Thus, from Lemma 3.1, (4.3) and (4.4), we have

$$\begin{aligned}
+\infty >&\theta_{\hat{k}}(x^{\hat{k}}) = \theta_*(x^{\hat{k}}) \\
\geq&\sum_{k=\hat{k}}^{\infty}\Big(\theta_*(x^k) - \theta_*(x^{k+1})\Big) \geq \sum_{k\in\mathcal{I}}\Big(\theta_*(x^k) - \theta_*(x^{k+1})\Big) \\
\geq&\eta_1\sum_{k\in\mathcal{I}}\Big(\Theta(0) - \Theta(d^k)\Big) \geq \eta_1\sum_{k\in\mathcal{I}}\bar{\tau}\tau\mu_*\min\left\{\Delta_k, \frac{\tau\mu_*}{C_1}\right\},
\end{aligned}$$

where $\mathcal{I} := \{k \in K_2 | r_k \geq \eta_1, k \geq \hat{k}\}$. This, if $\mathcal{I}$ is infinite, implies

$$\lim_{k(\in\mathcal{I})\to\infty}\Delta_k = 0. \tag{4.5}$$

Denote $\mathcal{J} := \{k \in K_2 | r_k < \eta_1, k \geq \hat{k}\}$. Since $\Delta_{k+1} = \alpha_1\Delta_k$ with $\alpha_1 \in (0, 1)$ for all $k \in \mathcal{J}$ from Step 5 of Algorithm 3.1, then, if $\mathcal{J}$ is infinite, we also have

$$\lim_{k(\in\mathcal{J})\to\infty}\Delta_k = 0. \tag{4.6}$$

Hence, from (4.5) and (4.6), it follows that

$$\lim_{k(\in K_2)\to\infty}\Delta_k = 0. \tag{4.7}$$

Recalling $\|d^k\| \leq \Delta_k$, the above limit implies that

$$\lim_{k(\in K_2)\to\infty}\|d^k\| = 0. \tag{4.8}$$

On the other hand, for all $k(\in K_2) \geq \hat{k}$, from the definitions of g_k and B_k and (4.8), we can obtain

$$\begin{aligned}
|r_k - 1| =&\left|\frac{\theta_*(x^k) - \theta_*(x^{k+1})}{\Theta(0) - \Theta(d^k)} - 1\right| = \left|\frac{-g_k^T d^k + o(\|d^k\|)}{-g_k^T d^k - \frac{1}{2}{d^k}^T B_k d^k} - 1\right| \\
=&\frac{\left|\frac{1}{2}{d^k}^T B_k d^k + o(\|d^k\|)\right|}{|\Theta(0) - \Theta(d^k)|} \leq \frac{\frac{1}{2}C_1\|d^k\|^2 + o(\|d^k\|)}{\bar{\tau}\tau\mu_*\min\left\{\Delta_k, \frac{\tau\mu_*}{C_1}\right\}} \\
\leq&\frac{\frac{1}{2}C_1\|d^k\|^2 + o(\|d^k\|)}{\bar{\tau}\tau\mu_*\|d^k\|} \to 0,
\end{aligned} \tag{4.9}$$

where the first inequality follows from (4.3), (4.4) and Lemma 3.1(i), and the second inequality follows from (4.7) and $\|d^k\| \le \Delta_k$. So (4.9) means that $r_k \ge \eta_1$ for all large $k(\in K_2) \ge \hat{k}$, which implies that $\Delta_k \ge \Delta_{\min} > 0$ by Step 5 of Algorithm 3.1. However, this contradicts (4.7). Therefore, the set K_2, i.e., $K \setminus K_1$ contains only finitely many indices.

(b) Then, without loss of generality, we can assume that $K = K_1$ so that the updating rule (3.6) is active for all $k \in K$. This, in particular, implies that $\{\mu_k\} \to 0$. Then from the test (3.5), it follows that

$$\lim_{k(\in K)\to\infty} \|\nabla\theta_k(x^{k+1})\| \le \lim_{k(\in K)\to\infty} \tau\mu_k = 0.$$

Therefore, from Lemma 4.1, we obtain

$$\nabla\theta(x^*) = \lim_{k(\in K)\to\infty} \|\nabla\theta_k(x^{k+1})\| = 0,$$

which shows that x^* is a stationary point of $\theta(x)$. This completes the proof. □

5. Local Convergence

In this section, we want to show that Algorithm 3.1 is locally Q-superlinearly/Q-quadratically convergent under certain assumptions.

The first result in this section follows from [5] together with known results for (strongly) semismooth functions [22] and the theory of C-differentiable functions [21].

Lemma 5.1. *Let $\{x^k\} \subseteq R^n$ be any convergent sequence with limit point $x^* \in R^n$. Then the following statements hold:*

(i) *The function Φ defined by* (2.3) *is semismooth so that*

$$\lim_{k\to\infty} \frac{\|\Phi(x^k) - \Phi(x^*) - V_k(x^k - x^*)\|}{\|x^k - x^*\|} = 0,$$

for any $V_k \in \partial_C\Phi(x^k)$.

(ii) *If F continuously differentiable with a locally Lipschitzian Jacobian, then Φ defined by* (2.3) *is strongly semismooth so that*

$$\limsup_{k\to\infty} \frac{\|\Phi(x^k) - \Phi(x^*) - V_k(x^k - x^*)\|}{\|x^k - x^*\|^2} < \infty,$$

for any $V_k \in \partial_C\Phi(x^k)$.

Lemma 5.2. *Let $\{x^k\}$ be a sequence generated by Algorithm* 3.1*, and x^* be an accumulation point of $\{x^k\}$. If x^* is a solution of NCP(F), then the index set $\mathcal{F}$ is infinite and $\{\mu_k\} \to 0$.*

Proof. For the sake of contradiction, we assume that $\mathcal{F}$ is finite. Then from the updating rules for β_k in Step 6 of Algorithm 3.1, we deduce that there exists $\bar{\beta} > 0$, such that

$$\beta_k = \bar{\beta} > 0,$$

for sufficiently large k. Thus, it follows from the test (3.3) that

$$\|\Phi(x^{k+1})\| > \max\left\{\eta\beta_k, \frac{1}{\alpha}\|\Phi(x^{k+1}) - \Phi_k(x^{k+1})\|\right\} \ge \eta\beta_k = \eta\bar{\beta} > 0,$$

for sufficiently large k. This contradicts the fact that $\|\Phi(x^*)\| = 0$ by assumption. Therefore, $\mathcal{F}$ must be infinite.

So the updating rules for μ_k imply that $\{\mu_k\} \to 0$. This completes the proof. $\square$

Remark 4. We explicitly point out that the proof of Lemma 5.2 shows that, each accumulation point of the sequence $\{x^k : k \in \mathfrak{N} \setminus \mathcal{F}\}$ (if the set $\mathfrak{N} \setminus \mathcal{F}$ is infinite) is impossibly a solution of NCP(F).

Lemma 5.3. *Let $\{x^k\}$ be a sequence generated by Algorithm* 3.1*, and x^* be an accumulation point of $\{x^k\}$. Assume that all elements of $\partial_C\Phi(x^*)$ are nonsingular. Then the following two statements are valid:*

(i) *x^* is a locally unique solution of NCP(F).*

(ii) *If L is an infinite subset of $\mathcal{F}$ such that $\{x^k\}_{k\in L} \to x^*$, then for all $k \in L$ sufficiently large, the matrix $J\Phi_k(x^k)$ is nonsingular and satisfies the inequality*

$$\|J\Phi_k(x^k)^{-1}\| \le C_2$$

for a certain constant $C_2 > 0$.

Proof. (i) Let $V \in \partial_C\Phi(x^*)$, then V is nonsingular by assumption. Theorem 4.1 guarantees that x^* is a stationary point of $\theta(x)$, i.e., $\nabla\theta(x^*) = 0$. From Proposition 4.1, we have

$$0 = \nabla\theta(x^*) = V^T\Phi(x).$$

Then the nonsingularity of V implies $\Phi(x^*) = 0$. Therefore, Proposition 2.5 in [25], together with Lemma 5.1(i), shows that x^* is a locally unique solution of $\Phi(x) = 0$ and hence also of NCP(F). This proves (i).

(ii) Notice that for any $x \in R^n$, the set $\partial_C\Phi(x)$ is nonempty and compact. So, for all $k \in \mathfrak{N}$, there exists $\widetilde{V}_k \in \partial_C\Phi(x^k)$ such that

$$dist(J\Phi_k(x^k), \partial_C\Phi(x^k)) = \|J\Phi_k(x^k) - \widetilde{V}_k\|.$$

Then, by (3.8), we have

$$\|J\Phi_k(x^k) - \widetilde{V}_k\| \le \nu\beta_k, \tag{5.1}$$

for all $k \in \mathcal{F}$.

Since $\{x^k\}_L \to x^*$ and $L \subseteq \mathcal{F}$, the nonsingularity of all elements of $\partial_C\Phi(x^*)$ and the upper semicontinuity of the C-subdifferential imply that, for all $k \in L$ sufficiently large, all matrices $V_k \in \partial_C\Phi(x^k)$ are nonsingular with $\|V_k^{-1}\| \le \bar{c}$ for some constant $\bar{c} > 0$. In particular,

$$\|\widetilde{V}_k^{-1}\| \le \bar{c}. \tag{5.2}$$

In view of the updating rules of β_k in Algorithm 3.1, $\{x^k\}_L \to x^*$ implies that $\beta_k \to 0$ for $k \in L$. Thereby, for $k \in L$ large enough such that $\nu\beta_k\bar{c} \le \frac{1}{2}$, we obtain from (5.1) and (5.2)

$$\begin{aligned}\|I - \widetilde{V}_k^{-1}J\Phi_k(x^k)\| &= \|\widetilde{V}_k^{-1}(\widetilde{V}_k - J\Phi_k(x^k))\| \\ &\le \|\widetilde{V}_k^{-1}\|\|\widetilde{V}_k - J\Phi_k(x^k)\| \\ &\le \nu\beta_k\bar{c} \le \frac{1}{2} < 1.\end{aligned}$$

From the perturbation lemma (see Theorem 3.1.4 in [26]), we clearly see that the matrix $J\Phi_k(x^k)$ is nonsingular for all $k \in L$ sufficiently large with

$$\|J\Phi_k(x^k)^{-1}\| \leq C_2,$$

where $C_2 := 2\bar{c}$. We complete the proof. □

Lemma 5.4. *Let $\{x^k\}$ be a sequence generated by Algorithm* 3.1. *Assume that x^* is one of the accumulation points of $\{x^k\}$ and all elements of $\partial_C \Phi(x^*)$ are nonsingular. Let L be an infinite subset of $\mathcal{F}$ such that $\{x^k\}_{k\in L}$ converges to x^*. Then for sufficiently large $k \in L$, we have*

$$d^k = -J\Phi_k(x^k)^{-1}\Phi_k(x^k)$$

is the unique solution of the subproblem (1.4).

Proof. Denote $s^k := -J\Phi_k(x^k)^{-1}\Phi_k(x^k)$. Lemma 5.3 shows that $\|\Phi(x^*)\| = 0$ and for all $k \in L$ sufficiently large, it holds that

$$\|J\Phi_k(x^k)^{-1}\| \leq C_2,$$

for a certain constant $C_2 > 0$. Hence $\{\|\Phi_k(x^k)\|\}_L \to 0$ and then, for all $k \in L$ sufficiently large,

$$\|s^k\| \leq \Delta_{\min} \leq \Delta_k,$$

which means that s^k is a feasible point of the subproblem (1.4). Note that $\Theta(s^k) = 0$, that is s^k is a global minimum point of $\Theta(d)$. Thus

$$d^k = -J\Phi_k(x^k)^{-1}\Phi_k(x^k)$$

is the unique solution of the subproblem (1.4). This completes the proof. □

In the proof of the main local convergence result, we will utilize the following two results; see Proposition 8.3.10 in [27] and Proposition 8 in [24], respectively.

Lemma 5.5. *Let $\{x^k\}$ be a sequence with a locally unique accumulation point x^*. Assume that for any subsequence $\{x^k\}_L$ converging to x^*, it holds that $\{\|x^{k+1} - x^k\|\}_L \to 0$. Then the whole sequence $\{x^k\}$ converges to x^*.*

Lemma 5.6. *Let $G : R^n \to R^n$ be locally Lipschitz continuous and $x^* \in R^n$ with $G(x^*) = 0$ such that all elements in $\partial G(x^*)$ are nonsingular, and assume that there are two sequences $\{x^k\} \subseteq R^n$ and $\{d^k\} \subseteq R^n$ with*

$$\lim_{k\to\infty} x^k = x^* \quad \textit{and} \quad \|x^k + d^k - x^*\| = o(\|x^k - x^*\|).$$

Then we have $\|G(x^k + d^k)\| = o(\|G(x^k)\|)$.

We are now able to prove our main local convergence result for Algorithm 3.1.

Theorem 5.1. *Let $\{x^k\}$ be a sequence generated by Algorithm* 3.1. *Assume that x^* is one of the accumulation points of $\{x^k\}$ and all elements of $\partial_C \Phi(x^*)$ are nonsingular. Then the entire sequence $\{x^k\}$ converges to x^* Q-superlinearly. Furthermore, if JF is locally Lipschitz continuous on R^n, then the convergence rate is Q-quadratic.*

Proof. (a) From Lemma 5.3 (i), it follows that x^* is a locally unique solution of NCP(F) by assumption. Then Lemma 5.2 shows that the index set $\mathcal{F}$ is infinite and $\{\mu_k\} \to 0$. Hence Proposition 4.2 (ii) shows that each accumulation point of the sequence $\{x^k\}$ must be a solution of NCP(F). Therefore, x^* is necessarily a locally unique accumulation point of the sequence $\{x^k\}$, and from Remark 4, it follows that the sequence $\{x^k : k \in \mathfrak{N} \setminus \mathcal{F}\}$ only contains finitely many members.

(b) In view of these, it therefore suffices to prove that the sequence $\{x^k : k \in \mathcal{F}\}$ converges to x^* Q-superlinearly.

Let L be an arbitrary subset of $\mathcal{F}$ such that $\{x^k\}_{k\in L}$ converges to x^*. From Lemma 5.4, we see that $d^k = -J\Phi_k(x^k)^{-1}\Phi_k(x^k)$ for sufficiently large $k \in L$ and $\{\|d^k\|\}_L \to 0$. On the other hand, we have

$$\|x^{k+1} - x^k\| = \left\{ \begin{array}{l} \|d^k\|, \\ \lambda^{i_k}\|d^k\| \end{array} \right\} \leq \|d^k\|$$

in view of the updating rule in Step 4 of Algorithm 3.1. Hence $\{\|x^{k+1} - x^k\|\}_L \to 0$. Then, by using Lemma 5.5, we obtain that the sequence $\{x^k : k \in \mathcal{F}\}$ converges to x^*.

Next, we show that the convergence rate is Q-superlinear. Let $V_k \in \partial_C \Phi(x^k)$ satisfy

$$dist(J\Phi_k(x^k), \partial_C\Phi(x^k)) = \|J\Phi_k(x^k) - V_k\|.$$

Then it follows from Lemma 5.1(i), (3.8) and (2.7) that

$$\begin{aligned} \|x^k + d^k - x^*\| =& \|x^k - J\Phi_k(x^k)^{-1}\Phi_k(x^k) - x^*\| \\ =& \| - J\Phi_k(x^k)^{-1}\big(\Phi_k(x^k) - J\Phi_k(x^k)(x^k - x^*)\big)\| \\ \leq& C_2\big(\|\Phi(x^k) - \Phi(x^*) - V_k(x^k - x^*)\| \\ &\quad + \|(J\Phi_k(x^k) - V_k)(x^k - x^*)\| + \|\Phi_k(x^k) - \Phi(x^k)\|\big) \\ \leq& C_2\big(o(\|x^k - x^*\|) + \nu\|\Phi(x^k)\|\|x^k - x^*\| + \kappa\mu_k\big) \end{aligned} \tag{5.3}$$

for sufficiently large $k \in \mathcal{F}$. The locally Lipschitz continuity of Φ implies that

$$\|\Phi(x^k)\| = \mathcal{O}(\|x^k - x^*\|), \qquad \text{as } x^k \to x^*.$$

And from the updating rule (3.4) of μ_k, we get

$$\mu_k \leq \theta(x^k) = \frac{1}{2}\|\Phi(x^k)\|^2.$$

Therefore, from (5.3), we have

$$\|x^k + d^k - x^*\| \leq o(\|x^k - x^*\|) + \mathcal{O}(\|x^k - x^*\|^2) = o(\|x^k - x^*\|), \tag{5.4}$$

for $k \to \infty$, $k \in \mathcal{F}$. Combining with Lemma 5.6, it holds that

$$\|\Phi(x^k + d^k)\| = o(\|\Phi(x^k)\|), \qquad \text{for } k \to \infty, k \in \mathcal{F}. \tag{5.5}$$

Consequently, we get

$$\begin{aligned} \|\Phi_k(x^k + d^k)\| \leq& \|\Phi(x^k + d^k)\| + \kappa\mu_k \\ \leq& \|\Phi(x^k + d^k)\| + \frac{\kappa}{2}\|\Phi(x^k)\|^2 = o(\|\Phi(x^k)\|). \end{aligned} \tag{5.6}$$

We also obtain

$$\begin{aligned}\|\Phi_k(x^k)\| \geq& \|\Phi(x^k)\| - \|\Phi_k(x^k) - \Phi(x^k)\| \geq \|\Phi(x^k)\| - \kappa\mu_k \\ \geq& \|\Phi(x^k)\| - \frac{\alpha}{2}\beta_k \geq \left(1 - \frac{\alpha}{2}\right)\|\Phi(x^k)\| > 0.\end{aligned} \tag{5.7}$$

Taking into account that $\Theta(d^k) = 0$, we deduce from (5.6) and (5.7) that

$$\begin{aligned}|r_k - 1| =& \left|\frac{\theta_k(x^k) - \theta_k(x^k + d^k)}{\Theta(0) - \Theta(d^k)} - 1\right| = \left|\frac{\Theta(d^k) - \theta_k(x^k + d^k)}{\Theta(0) - \Theta(d^k)}\right| \\ =& \frac{\|\Phi_k(x^k + d^k)\|^2}{\|\Phi_k(x^k)\|^2} \leq \frac{o(\|\Phi(x^k)\|^2)}{(1 - \frac{\alpha}{2})^2\|\Phi(x^k)\|^2} \to 0,\end{aligned} \tag{5.8}$$

which indicates that there exists an index $\tilde{k} \in \mathcal{F}$ such that $r_k \geq \eta_1$ for all $k \in \mathcal{F}$ with $k \geq \tilde{k}$. So the full step size of 1 will eventually be accepted for all $k \geq \tilde{k}$, $k \in \mathcal{F}$. In particular, $x^{\tilde{k}+1} = x^{\tilde{k}} + d^{\tilde{k}}$. On the other hand, (5.5) implies that

$$\|\Phi(x^{\tilde{k}} + d^{\tilde{k}})\| \leq \eta\|\Phi(x^{\tilde{k}})\| = \eta\beta_{\tilde{k}}.$$

Hence $\tilde{k} + 1 \in \mathcal{F}$; cf. (3.7). Repeating the above process, we may prove that for all $k \geq \tilde{k}$, we have $k \in \mathcal{F}$ and $x^{k+1} = x^k + d^k$. Then by using (5.4), we have proved that $\{x^k : k \in \mathcal{F}\}$ converges to x^* superlinearly.

From parts (a) and (b), it follows that the entire sequence $\{x^k\}$ converges to x^* Q-superlinearly.

Furthermore, if JF is locally Lipschitz continuous on R^n, then Lemma 5.1 (ii) shows that

$$\|\Phi(x^k) - \Phi(x^*) - V_k(x^k - x^*)\| = \mathcal{O}(\|x^k - x^*\|^2).$$

Hence the Q-quadratic rate of convergence of $\{x^k\}$ to x^* follows from (5.3) by using similar arguments as for the proof of the local Q-superlinear convergence. The proof is then complete. □

6. Numerical Experiments

In this section, we present some numerical experiments for the smoothing trust region method from Algorithm 3.1. The program code was written in MATLAB and run in MATLAB 7.0 environment.

In our implementation, we used the following parameter settings:

• Termination criterion:

We terminated our iteration if one of the following conditions was satisfied:

$$\min\left\{\|\nabla\theta(x^k)\|,\ \|\min\{x^k, F(x^k)\}\|_\infty\right\} \leq 10^{-6}, \quad k > 300.$$

Notice that the first term on the left-hand side of the first stopping condition above is used as a safeguard against the case that an accumulation point of the sequence generated by Algorithm 3.1 is a mere stationary point of θ, which is not a solution of the NCP(F).

• Trust region parameters:

$$\Delta_{\min} = 1, \quad \Delta_0 = 100, \quad \eta_1 = 10^{-4}, \quad \eta_2 = 0.75, \quad \alpha_1 = 0.5, \quad \alpha_2 = 2.$$

Table 6.1: Numerical results for Example 1.

	$p=1.2$		$p=2$		$p=5$		$p=10$	
n	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
200	5	1.12e-11	5	1.93e-22	3	2.83e-16	3	1.72e-30
512	5	2.88e-11	5	4.97e-22	3	7.26e-16	3	4.30e-30
800	5	4.51e-11	5	7.77e-22	3	1.13e-15	3	6.71e-30
1024	5	5.77e-11	5	9.95e-22	3	1.45e-15	3	8.25e-30

Table 6.2: Numerical results for Example 2.

	$p=1.2$		$p=2$		$p=5$		$p=10$	
SP	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
$(0,0,0,0)^T$	12^*	3.43e-13	10	1.07e-14	9	3.24e-16	9	1.63e-14
$(1,1,1,1)^T$	8	3.13e-15	7	1.65e-15	6	5.07e-20	6	5.27e-24
$(10,10,10,10)^T$	10^*	1.47e-13	10	2.43e-19	7	2.25e-13	8	1.66e-22
$(100,100,100,100)^T$	12^*	4.28e-16	8^*	2.45e-19	11	3.25e-16	11	1.62e-14
$(-100,-100,-100,-100)^T$	14^*	1.25e-14	8^*	8.40e-19	11	3.24e-16	11	1.63e-14

- Other parameters:

$$\lambda = 0.5, \quad \eta = 0.9, \quad \alpha = 0.05, \quad \sigma = 0.1, \quad \tau = 2, \quad \nu = 30.$$

In the following tables of numerical results, SP denotes the starting point, k indicates the number of iterations, and $\theta(x^f)$ represents the value of the merit function $\theta(x)$ at the final iterate $x = x^f$. The test problems are introduced as follows:

Example 1. This test problem is from Ahn [28]. Let $F(x) = Mx + q$, where

$$M = \begin{pmatrix} 4 & -2 & 0 & \cdots & 0 & 0 \\ 1 & 4 & -2 & \cdots & 0 & 0 \\ 0 & 1 & 4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 4 & -2 \\ 0 & 0 & 0 & \cdots & 1 & 4 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ -1 \\ -1 \end{pmatrix}.$$

Table 1 gives the results for this example with starting point $x^0 = (0, 0, \cdots, 0)^T$ for different dimensions n.

Example 2. Kojshin Problem. This example was used by Pang and Gabriel [29], and Kanzow [30] with four variables. Let

$$F(x) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

This problem has one degenerate solution $(\frac{\sqrt{6}}{2}, 0, 0, \frac{1}{2})^T$ and one nondegenerate solution $(1, 0, 3, 0)^T$. The numerical results are listed in Table 2 using different initial points. The asterisk (*) denotes that the limit point generated by the algorithms is the degenerate solution, otherwise it is the nondegenerate solution.

Table 6.3: Numerical results for Example 3.

	$p = 1.2$		$p = 2$		$p = 5$		$p = 10$	
SP	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
$(0,0,0,0,0)^T$	29	4.37e-13	25	4.06e-26	22	4.58e-30	21	6.96e-13
$(1,2,3,1,2)^T$	18	2.56e-14	21	2.50e-16	28	1.48e-20	28	4.01e-31
$(2,2,2,2,2)^T$	30	5.40e-15	30	3.60e-23	33	4.00e-31	28	4.32e-13
$(1,2,3,4,5)^T$	8	3.16e-14	11	1.31e-20	13	9.80e-21	12	5.62e-17
$(1,0,1,3,5)^T$	7	4.18e-15	6	1.48e-16	7	3.94e-31	7	1.36e-38

Table 6.4: Numerical results for Example 4.

	$p = 1.2$		$p = 2$		$p = 5$		$p = 10$	
SP	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
$(1,1,1,1)^T$	5	9.65e-13	4	3.12e-16	3	1.46e-19	3	2.58e-30
$(2,2,2,2)^T$	10	8.38e-15	4	1.78e-21	3	2.29e-31	3	3.93e-61
$(-2,-2,-2,-2)^T$	7	9.02e-16	5	3.28e-17	3	3.63e-18	3	1.12e-30
$(-4,-4,-4,-4)^T$	7	3.00e-14	4	4.42e-16	3	7.07e-14	3	2.91e-30
$(9,9,9,9)^T$	9	9.95e-14	7	8.64e-16	5	2.64e-23	6	1.23e-31

Example 3. This problem was tested by Kanzow [30] with five variables defined by

$$F_i(x) = 2(x_i - i + 2)\exp\left\{\sum_{i=1}^{5}(x_i - i + 2)^2\right\}, \qquad 1 \le i \le 5.$$

This example has one degenerate solution $x^* = (0,0,1,2,3)^T$. The numerical results are given in Table 3 using different initial points.

Example 4. Modified Mathiesen Problem. This test problem is the fifth example of Jiang and Qi [31] with four variables, which was also tested by Kanzow [30]. Let

$$F(x) = \begin{pmatrix} -x_2 + x_3 + x_4, \\ x_1 - \frac{(4.5x_3+2.7x_4)}{(x_2+1)}, \\ 5 - x_1 - \frac{(0.5x_3+0.3x_4)}{(x_3+1)}, \\ 3 - x_1 \end{pmatrix}$$

This example has infinitely many solutions $(\lambda, 0, 0, 0)$, where $\lambda \in [0,3]$. For $\lambda = 0, 3$, the solutions are degenerate, and for $\lambda \in (0,3)$ nondegenerate. The test results are listed in Table 4 using different starting points.

We next test some economic equilibrium problems with larger sizes, but with only standard starting points. They are from the MCPLIB collection [32].

Example 5. Hanshoop Problem. The test function is

$$F(x,y,u) = \begin{bmatrix} -\nabla v(x) \\ 0 \\ w \end{bmatrix} + \begin{bmatrix} 0 & A^T - \alpha B^T & C^T \\ B - A & 0 & 0 \\ -C & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ u \end{bmatrix},$$

where $v(x) = (x_1 + 2.5x_2)^p(2.5x_3 + x_4)^p(2x_5 + 3x_6)^p$, $\alpha = 0.7$, $p = 0.2$, $w = (0.8, 0.8)^T$, and

$$A = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 2 & 2 & 1 & 1 & 1 & 0.5 & 1.5 & 0.5 \end{bmatrix},$$

$$B = \begin{bmatrix} 1.5 & 1.5 & 1.5 & 1.5 & 1.5 & 1.5 & 4 & 3 & 1.5 & 1.5 \\ 2.7 & 2.7 & 1.8 & 1.8 & 0.9 & 0.9 & 0.9 & 0.4 & 2 & 1.5 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.5 & 1.5 & 1.5 & 0.5 & 0.5 & 1.5 & 1.5 & 0.5 & 0.5 & 1.5 \end{bmatrix}.$$

We take the starting points $y^0 = (0,0)^T$, $u^0 = (0,0)^T$ and x^0: (1) $0.3e$; (2) $0.5e$; (3) e; (4) $(0.3, 0, 0.3, 0, 0.3, 0, 0.3, 0, 0.3, 0)^T$, where e is the 10-dimensional vector of ones. The numerical results for Example 5 are given in Table 5 using these starting points.

Example 6. Nash Problem. This is a Nash equilibrium model with ten variables. The test function $F(x) = (F_1(x), \cdots, F_{10}(x))^T$ is defined by

$$F_i(x) = c_i + (L_i x_i)^{\frac{1}{\beta_i}} - \left[\frac{5000}{\sum_{k=1}^{10} x_k}\right]^{\frac{1}{\gamma}} + \frac{x_i}{\gamma \sum_{k=1}^{10} x_k}\left[\frac{5000}{\sum_{k=1}^{10} x_k}\right]^{\frac{1}{\gamma}}, \quad 1 \le i \le 10,$$

where $\gamma = 1.2$, $c = (5.0, 3.0, 8.0, 5.0, 1.0, 3.0, 7.0, 4.0, 6.0, 3.0)^T$, $L_i = 10$ $(1 \le i \le 10)$, and $\beta = (1.2, 1.0, 0.9, 0.6, 1.5, 1.0, 0.7, 1.1, 0.95, 0.75)^T$. The test results for Example 6 are summarized in Table 6 using the following standard starting points: (1) e; (2) $10e$; (3) $(1.0, 1.2, 1.4, 1.6, 1.8, 2.1, 2.3, 2.5, 2.7, 2.9)^T$; (4) $(7, 4, 3, 1, 8, 4, 1, 6, 3, 2)^T$.

Table 6.5: Numerical results for Example 5.

	$p=1.2$		$p=2$		$p=5$		$p=10$	
SP	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
(1)	7	2.42e-13	8	5.34e-20	7	3.12e-20	7	4.30e-22
(2)	12	8.48e-16	8	1.07e-17	9	2.54e-13	8	6.11e-19
(3)	9	1.10e-13	10	5.59e-22	9	1.94e-13	7	4.32e-15
(4)	10	1.08e-14	7	1.90e-16	7	2.80e-13	7	1.13e-22

Table 6.6: Numerical results for Example 6.

	$p=1.2$		$p=2$		$p=5$		$p=10$	
SP	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$	k	$\theta(x^f)$
(1)	23	6.00e-13	25	3.60e-13	23	5.89e-13	27	7.47e-13
(2)	24	5.08e-13	29	7.91e-13	32	5.11e-13	32	5.20e-13
(3)	23	3.40e-13	23	4.08e-13	33	3.74e-13	30	5.81e-13
(4)	23	4.68e-13	23	5.95e-13	25	7.87e-13	25	7.79e-13

7. Conclusions

In this paper, we have presented a new smoothing trust region-type method with line search for the solution of a general (i.e., not necessarily monotone) complementarity problem. This method is based on a reformulation of the complementarity problem as a system of nonsmooth equations by using the generalized Fischer-Burmeister function. With careful analyses, we are able to prove that our method is globally and superlinearly convergent. Moreover, the numerical performance is extremely promising. In particular, our algorithm associated with a bigger p would have better numerical behavior in terms of the number of iteration.

It would be interesting to see how our smoothing method within the trust region framework would work on mixed complementarity problems, general box constrained variational inequalities, or more general class of problems. We will leave it in a subsequent research topic.

Acknowledgments. The work was supported by the National Natural Science Foundation of China (11071041) and Fujian Natural Science Foundation (2009J01002).

References

[1] P.T. Harker, J.S. Pang, Finite dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Math. Program.*, **48**(1990), 161-220.

[2] M.C. Ferris, J.S. Pang, Engineering and economic application of complementarity problems, *SIAM Rev.*, **39** (1997), 669-713.

[3] H. Jiang, M. Fukushima, L. Qi, D. Sun, A trust region method for solving generalized complementarity problem, *SIAM J. Optimiz.*, **8** (1998), 140-157.

[4] J.-S. Chen, S. Pan, A regularization semismooth Newton method based on the generalized Fischer-Burmeister function for P_0-NCPs, *J. Comput. Appl. Math.*, **220** (2008), 464-479.

[5] J.-S. Chen, The semismooth-related properties of a merit function and a descent method for the nonlinear complementarity problem, *J. Global Optim.*, **36** (2006), 565-580.

[6] F. Facchinei, J. Soares , A new merit function for nonlinear complementarity problems and a related algorithm, *SIAM J. Optimiz.*, **7** (1997), 225-247.

[7] C. Kanzow , Nonlinear complementarity as unconstrained optimization, *J. Optimiz. Theory App.*, **88** (1996), 139-155.

[8] P. Tseng, Global behavior of a class of merit functions for the nonlinear complementarity problem, *J. Optimiz. Theory App.*, **89** (1996), 17-37.

[9] L. Qi, Trust region algorithms for solving nonsmooth equation, *SIAM J. Optimiz.*, **5** (1995), 219-230.

[10] O. L. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations, *SIAM J. Appl. Math.*, **31** (1976), 89-92.

[11] J.J. Moré, Global methods for nonlinear complementarity problems, *Math. Oper. Res.*, **21** (1996), 589-614.

[12] B.C. Eaves, On the basic theorem of complementarity, *Math. Program.*, **1** (1971), 68-75.

[13] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[14] P.T. Harker, B. Xiao, Newton method for the nonlinear complementarity problem: A B-differentiable equation approach, *Math. Program.*, **48** (1990), 339-357.

[15] S.M. Robinson, Homeomorphism conditions for normal maps of polyhedra, in Optimization and Nonlinear Analysis, A.Loffe, M.Marcus and S.Reich(Eds.), Pitman Research Notes in Mathematics Series 244, 1992, 240-248.

[16] S.M. Robinson, Sensitivity analysis of variational inequalities by normal-map techniques, in Nonlinear Variational Inequalities and Network Equilibrium Problems, F.Giannessi and A.Maugeri(Eds.), Plenum Press: New York, 1995, 257-269.

[17] L. Zhang, Z. Gao, Superlinear/quadratic one-step smoothing Newton method for P_0-NCP without strict complementarity, *Math. Method. Oper. Res.*, **56** (2002), 231-241.

[18] X. Chen, L. Qi, D. Sun, Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, *Math. Comput.*, **67** (1998), 519-540.

[19] J.-S. Pang, A B-differentiable equations based, globally, and locally quadratically convergent algorithm for nonlinear problems, complementarity and variational inequality problems, *Math. Program.*, **51** (1991), 101-131.

[20] C. Kanzow, H. Pieper, Jacobian smoothing methods for general nonlinear complementarity problems, *SIAM J. Optimiz.*, **9** (1999), 342-373.

[21] L. Qi, C-Differentiability, C-Differential Operators and Generalized Newton Methods, Tech. Report, School of Mathematics, The University of New South Wales, Sydney, Australia, January 1996.

[22] L. Qi, J. Sun, A nonsmooth version of Newton's method, *Math. Program.*, **58** (1993), 353-367.

[23] J. Nocedal, Y.X. Yuan, Combining trust region and line search techniques, in: Y. Yuan (Ed.), Advances in Nonlinear Programming, Kluwer Academic Publishers, Boston, MA, 1998, 153-175.

[24] C. Kanzow, H.D. Qi, A QP-free constrained Newton-type method for variational inequality problems, *Math. Program.*, **85** (1999), 81-106.

[25] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Math. Oper. Res.*, **18** (1993), 227-244.

[26] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983(reprinted by SIAM, Philadelphia, 1996).

[27] F. Facchinei, J.-S. Pang, Finite-dimensional Variational Inequalities and Complementarity Problems, Vol. I and II. New York: Springer, 2003.

[28] B.H. Ahn, Iterative methods for linear complementarity problem with upperbounds and lowerbounds, *Math. Program.*, **26** (1983), 265-315.

[29] J.-S. Pang, S.A. Gabriel, NE/SQP: A robust algorithm for the nonlinear Complementarity Problem, *Math. Program.*, **60** (1993), 295-337.

[30] C. Kanzow, Some equation-based methods for the nonlinear complementarity problem, *Optim. Method. Softw.*, **3** (1994), 327-340.

[31] H. Jiang, L. Qi, A new nonsmooth equations approach to nonlinear complementarity problems, *SIAM J. Control Optim.*, **35** (1997), 178-193.

[32] S.P. Dirkse, M.C. Ferris, MCPLIB: A collection of nonlinear mixed complementarity problems, *Optim. Method. Softw.*, **5** (1995), 123-156.