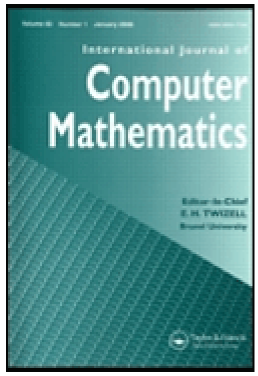
Inexact quasi-Newton global convergent method for solving constrained nonsmooth equations

Marek J. Śmietański



Inexact quasi-Newton global convergent method for solving constrained nonsmooth equations

Journal:	*International Journal of Computer Mathematics*
Manuscript ID:	GCOM-2006-0257.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	14-Feb-2007
Complete List of Authors:	Śmietański, Marek; University of Lodz, Faculty of Mathematics
Keywords:	nonsmooth equation, inexact-Newton method, quasi-Newton method, superlinear convergence, constrained system

Inexact quasi-Newton global convergent method for solving constrained nonsmooth equations

M. J. ŚMIETAŃSKI

Faculty of Mathematics, University of Lodz, Poland
90-238 Lodz, ul.Banacha 22
tel: (48)(42)635-58-77, fax: (48)(42)635-42-66

e-mail: smietan@math.uni.lodz.pl

In this work we introduce a new method for solving nonsmooth equations with simple constraints. The method is based on the inexact and quasi-Newton approaches with backtracking strategy. Some conditions are given that ensure global superlinear convergence to a solution of the equation. Moreover, we propose a nonmonotone scheme of algorithm. **Both versions of algorithm was constructed for the Lipschitz continuous equations.**

Keywords: nonsmooth equation, inexact-Newton method, quasi-Newton method, superlinear convergence, constrained system

2000 Mathematics Subject classification codes: 65H10 *CCS:* G.1.5, F.2.1

1 Introduction

The problem **considered** in this work is to find $x \in \Omega \subset R^n$, which is a solution of the system of nonlinear equations

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

where Ω is closed and convex, $F : R^n \to R^n$ is Lipschitz continuous on an open set that contains Ω.

Newton's method is the most known method for solving nonlinear systems, which arise from other important problems. Pang and Qi in [16] established a lot of the motivations of nonsmooth equations. **Constrained nonlinear systems appear in applications when we need to solve real-life problems. However, not all solutions of the mathematical model have physical meaning, only the ones belonging to a constraint set Ω.** Often, Ω is an n-dimensional box $\{x \in R^n : l \leq x \leq u\}$, where $l, u \in R^n$. A formulation as nonlinear programming problem using the squared norm of F as objective function can be inefficient in some cases.

An efficient method **for solving the equation** (1) without constraint is the inexact Newton method. The general idea of this method was presented by Dembo, Eisenstat and Steihaug in [7]. A sequence $\{x^{(k)}\}$ of **approximations** to solution x^* is generated as follows: find the step $s^{(k)} \in R^n$ which satisfies

$$\left\| J\left(x^{(k)}\right) s^{(k)} + F\left(x^{(k)}\right) \right\| \leq \eta_k \left\| F\left(x^{(k)}\right) \right\|$$

and set $x^{(k+1)} = x^{(k)} + s^{(k)}$, where $\eta_k \in [0, 1)$, $k = 0, 1, 2, ...$ are the scalar parameters and $J(x)$ denotes the Jacobian of F. The inexact-Newton iteration was analyzed from various points of view, e.g. global convergence by Eisenstat and Walker in [8], nonsmooth version by Martínez and Qi in [13], nonmonotone smooth version by Bonettini in [4]. Moreover, Kozakevich, Martínez and Santos introduced in [11] global convergent inexact-Newton algorithm for solving a smooth constrained equations. In quasi-Newton methods the direction (step) is computed by solving

$$B_k s^{(k)} = -F\left(x^{(k)}\right),$$

where, in general B_k is not the Jacobian, e.g. in generalized Newton method B_k is taken from Clarke's generalized Jacobian $\partial F(x)$ (see [20]) or from B-differential $\partial_B F(x)$ (see [18]). In [17] one type of globally convergent inexact generalized Newton's method to solve nonsmooth equations was proposed by Pu and Tian. The combination of the ideas of inexact-Newton and quasi-Newton method was described **in several papers. Some versions of the inexact quasi-Newton method for solving smooth equations was proposed e.g. by Bergamaschi, Moret, Zilli in [2] (inexact Newton-Cimmino method for sparse systems) and Birgin, Krejić, Martínez in [3] (inexact quasi-Newton algorithm with backtracking). Another study on the inexact quasi-Newton method with preconditioners can be found in Bergamaschi, Bru, Martínez, Putti [1]. Our approach is to generalize the smooth inexact quasi-Newton method for nonsmooth case and to modify the general framework in a nonmonotone way. Our proposed algorithms was constructed for solving the Lipschitz continuous equations for which some mild assumptions are fulfilled.**

In whole work we assume that function $F : R^n \to R^n$ is Lipschitz continuous, i.e. there exists $L > 0$ such that, for any $x, y \in R^n$ it holds

$$\|F(x) - F(y)\| \leq L \|x - y\|.$$

According to the Rademacher's theorem the Lipschitz continuity of F implies that F is differentiable almost everywhere. Let D_F be the set where F is differentiable. Then

$$\partial_B F(x) = \left\{ \lim_{x_i \to x} JF(x_i), x_i \in D_F \right\}$$

is called B-differential of F at x [18]. The generalized Jacobian of F at x in the sense of Clarke [6] is

$$\partial F(x) = \operatorname{conv} \partial_B F(x).$$

We have (see [6])
(a) $\partial F(x)$ is nonempty, convex and compact;
(b) ∂F is upper semicontinuous at x.

We say that F is BD-regular at x if F is locally Lipschitz at x and if all $V \in \partial_B F(x)$ are nonsingular. Qi in [18] (Lemma 2.6) proved that if F is BD-regular at x, then there exist a neighborhood N of x and a constant $C > 0$ such that for any $y \in N$ and $V \in \partial_B F(y)$, V is nonsingular and

$$\left\|V^{-1}\right\| \leq C.$$

This paper is organized as follows. In section 2 we describe the model algorithm of our nonsmooth inexact quasi-Newton method, we prove the superlinear convergence and we state convergence theorem for main algorithm of method. In section 3 we present a nonmonotone version of algorithm. **Section 4 reports some numerical results concerning the application of the new algorithms to the different problems. Finally, we draw some conclusions in section 5.**

Notation. Throughout the paper, $x^* \in \Omega$ is a solution of (1). Moreover, $\|\cdot\|$ denotes the Euclidean norm. However, it is easy to verify that results are independent of this choice.

2 Algorithm and its properties

The main algorithm in this work is Algorithm 2. Before its statement, we define a more general method, that helps to understand the structure of the main algorithm.

Algorithm 1. (Model algorithm)
Assume that $\sigma \in (0,1)$, $\gamma \in (0,1]$, $\tau_1, \tau_2 \in (0,1)$, $\tau_1 < \tau_2$ are given independently of k. Let $x^{(0)} \in R^n$ be an arbitrary initial point and $\alpha_0 = 1$. Given a point $x^{(k)}$, the steps for obtaining $x^{(k+1)}$ are:
Step 1. Find some $s^{(k)} \in R^n$ such that

$$x^{(k)} + s^{(k)} \in \Omega.$$

Step 2. If

$$\left\|F\left(x^{(k)} + \alpha_k s^{(k)}\right)\right\| \leq \left\|F\left(x^{(k)}\right)\right\| \tag{2}$$

define

$$x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}. \tag{3}$$

Otherwise set $x^{(k+1)} = x^{(k)}$.
Step 3. If

$$\left\|F\left(x^{(k)} + \alpha_k s^{(k)}\right)\right\| \leq \left(1 - \frac{\sigma\gamma\alpha_k}{2}\right)\left\|F\left(x^{(k)}\right)\right\| \tag{4}$$

set $\alpha_{k+1} = 1$. Otherwise choose

$$\alpha_{k+1} \in [\tau_1\alpha_k, \tau_2\alpha_k], \tag{5}$$

Let us denote $K_1 = \{k \in N : (4) \text{ holds}\}$.

Lemma 1 *Let $\left\{x^{(k)}\right\}$ be the sequence generated by Algorithm 1. If K_1 is infinite and $\limsup_{k\in K_1} \alpha_k > 0$ then*

$$\lim_{k\to\infty} \left\|F\left(x^{(k)}\right)\right\| = 0.$$

Proof. Assume that K_2 is an infinite subset of K_1 such that

$$\alpha_k \geq \bar{\alpha} > 0 \text{ for all } k \in K_2.$$

Then
$$1-\frac{\sigma\gamma\alpha_k}{2}\leq 1-\frac{\sigma\gamma\bar{\alpha}}{2}\equiv c<1$$
for all $k\in K_2$. Therefore $\left\{\left\|F\left(x^{(k)}\right)\right\|\right\}$ is a nonincreasing sequence such that $\left\|F\left(x^{(k+1)}\right)\right\|\leq c\left\|F\left(x^{(k)}\right)\right\|$ for all $k\in K_2$. This implies that $\left\|F\left(x^{(k)}\right)\right\|\to 0$. ■

For convenience, the sum of squares of $F(x)$ as merit function
$$f(x)=\frac{1}{2}\left\|F(x)\right\|^2$$
will be used. Note that $F\left(x^{(k)}\right)$ is reduced monotonically in algorithm. The sufficient reduction criterion imposed depends on the norm of F not on its generalized Jacobian.

Assumption A: Assume that function F is Lipschitz continuous. We say that F satisfies A at x if for any $y\in R^n$ and any $V_y\in\partial_B F(y)$, the following equality holds
$$F(y)-F(x)=V_y\left(y-x\right)+o\left(\left\|y-x\right\|\right).$$
Moreover, we say that F satisfies A at x with degree ρ if F is Lipschitz continuous and the following equality holds
$$F(y)-F(x)=V_y\left(y-x\right)+O\left(\left\|y-x\right\|^{\rho}\right).$$

Remarks: (i) Pu and Tian in [17] established three classes of functions that satisfied assumption A. Semismoothness (introduced by Mifflin in [15]), second order C-differentiability (introduced by Qi in [19]) and H-differentiability (introduced by Gowda and Ravindran in [9]) are properties that imply A.
(ii) If F is BD-regular at x and satisfies A at x, then there exist a neighborhood N of x and a constant $C>0$ such that for any $y\in N$ and $V\in\partial_B F(y)$
$$\left\|y-x\right\|\leq C\left\|V_y\left(y-x\right)\right\|. \tag{6}$$

Lemma 2 *Let*
$$l=\max\left\{2\beta,\frac{1}{2\beta}+\left\|V_y\right\|\right\},$$
where $\beta=\left\|V_y^{-1}\right\|$, $V_y\in\partial_B F(y)$. *If* F *is BD-regular at* x^* *and satisfies assumption A at* x^* *then*
$$\frac{1}{l}\left\|y-x^*\right\|\leq\left\|F\left(y\right)\right\|\leq l\left\|y-x^*\right\|$$
for all $y\in N_{x^*}$, *where* N_{x^*} *is some neighborhood of* x^*.

Proof. By assumption A, there exists a neighborhood N_{x^*} of x^* such that for any $y\in N_{x^*}$ and any $V_y\in\partial_B F(y)$
$$\left\|F(y)-F(x^*)-V_y\left(y-x^*\right)\right\|\leq\frac{1}{2\beta}\left\|y-x^*\right\|.$$

Since for all $V_y \in \partial_B F(y)$

$$F(y) = V_y (y - x^*) + [F(y) - F(x^*) - V_y (y - x^*)],$$

taking norms,

$$\begin{aligned} \|F(y)\| &\leq \|V_y\| \|y - x^*\| + \|F(y) - F(x^*) - V_y (y - x^*)\| \leq \\ &\leq \left(\|V_y\| + \frac{1}{2\beta}\right) \|y - x^*\|, \end{aligned}$$

and

$$\begin{aligned} \|F(y)\| &\geq \left\|V_y^{-1}\right\|^{-1} \|y - x^*\| - \|F(y) - F(x^*) - V_y (y - x^*)\| \geq \\ &\geq \left(\left\|V_y^{-1}\right\|^{-1} - \frac{1}{2\beta}\right) \|y - x^*\| = \frac{1}{2\beta} \|y - x^*\|, \end{aligned}$$

whenever $y \in N_{x^*}$. ■

Remark: A similar lemma as the above one was established by Dembo, Eisenstat and **Steihaug** in [7] for continuously differentiable functions to prove superlinear convergence of the classical inexact Newton method.

Theorem 3 *Assume that $L_F = \{x \in \Omega : \|F(x)\| \leq \|F(x_0)\|\}$ is bounded. Let $\theta \in [0, 1)$ and $\{x^{(k)}\}$ be the sequence generated by Algorithm 1 with $\gamma = 1 - \theta^2$. Assume that there exists $M > 0$ such that for all $k = 0, 1, 2, ...$*

$$\left\|s^{(k)}\right\| \leq M \tag{7}$$

and

$$\left\|V_k s^{(k)} + F\left(x^{(k)}\right)\right\| \leq \theta \left\|F\left(x^{(k)}\right)\right\|, \tag{8}$$

where $V_k \in \partial_B F\left(x^{(k)}\right)$.
If F is BD-regular at x^*, satisfies assumption A at x^* and for every sequence $\{x^{(k)}\}$ converging to x^*, every convergent sequence $\{s^{(k)}\}$ and every sequence $\{\lambda_k\}$ of positive scalars converging to 0

$$\limsup_{k\to\infty} \frac{f\left(x^{(k)} + \lambda_k s^{(k)}\right) - f\left(x^{(k)}\right)}{\lambda_k} \leq \lim_{k\to\infty} F\left(x^{(k)}\right)^T V_k s^{(k)}, \tag{9}$$

whenever the limit in the left-hand side exists, then every limit point of the sequence $\{x^{(k)}\}$ is a solution of equation (1) and $\{x^{(k)}\}$ converges superlinearly to x^*.

Proof. If K_1 is infinite and $\limsup_{k \in K_1} \alpha_k > 0$ the result follows from Lemma 1 and Lipschitz continuity of F. Now, assume that

$$K_1 \textbf{ is } \text{infinite and } \lim_{k \in K_1} \alpha_k = 0. \tag{10}$$

Because L_F is bounded, there exist $x^* \in \Omega$ and K_2, an infinite subset of K_1, such that

$$\lim_{k \in K_2} x^{(k)} = x^*.$$

Assume, by contradiction, that $F(x^*) \neq 0$. So, $F\left(x^{(k)}\right) \neq 0$ for all $k = 0, 1, 2, \ldots$. We may assume that $\alpha_k < 1$ for all $k \in K_2$ without loss of generality. By (5) and (4) we have that, for all $k \in K_2$

$$\alpha_k \in [\tau_1 \alpha_{k-1}, \tau_2 \alpha_{k-1}] \tag{11}$$

and

$$\left\| F\left(x^{(k-1)} + \alpha_{k-1} s^{(k-1)}\right) \right\| > \left(1 - \frac{\sigma \gamma \alpha_{k-1}}{2}\right) \left\| F\left(x^{(k-1)}\right) \right\|. \tag{12}$$

By (10) and (11) we have $\lim_{k \in K_2} \alpha_{k-1} = 0$.
So, using (7), we obtain that

$$\lim_{k \in K_2} x^{(k-1)} = x^*.$$

Moreover, by (12)

$$\frac{f\left(x^{(k-1)} + \alpha_{k-1} s^{(k-1)}\right) - f\left(x^{(k-1)}\right)}{\alpha_{k-1}} > \left[\frac{\sigma^2 \gamma^2 \alpha_{k-1}}{4} - \sigma\gamma\right] f\left(x^{(k-1)}\right) \tag{13}$$

for all $k \in K_2$.
Since $\left\| s^{(k-1)} \right\| \leq M$ for all k, there exists K_3, an infinite subset of K_2, such that

$$\lim_{k \in K_3} s^{(k-1)} = s.$$

Taking limits for $k \in K_3$ on both sides of (13), we obtain

$$\lim_{k \in K_3} \frac{f\left(x^{(k-1)} + \alpha_{k-1} s^{(k-1)}\right) - f\left(x^{(k-1)}\right)}{\alpha_{k-1}} \geq -\sigma\gamma f(x^*).$$

Exploiting the assumption (9), we obtain

$$\lim_{k \in K_3} F\left(x^{(k-1)}\right)^T V_{k-1} s^{(k-1)} \geq -\sigma\gamma f(x^*).$$

So, for large enough $k \in K_3$

$$\left\langle V_{k-1} s^{(k-1)}, F\left(x^{(k-1)}\right) \right\rangle > -\gamma f\left(x^{(k-1)}\right) \tag{14}$$

Now, observe that (8) implies that for all $k = 0, 1, 2, \ldots$

$$\left\langle V_k s^{(k)}, F\left(x^{(k)}\right) \right\rangle \leq -\gamma f\left(x^{(k)}\right)$$

where $V_k \in \partial_B F\left(x^{(k)}\right)$, which **contradicts (14)**. This proves that original assumption $F(x^*) \neq 0$ is false. Since $\left\{ \left\| F\left(x^{(k)}\right) \right\| \right\}$ is monotone, any other

limit point of $\{x^{(k)}\}$ **has to be** solution of (1).
Now assume that K_1 is finite. Hence, there exists $k_0 \in \mathbb{N}$ such that (4) does not hold for all $k \geq k_0$. Therefore, $\alpha_k \to 0$ and we can repeat the former proof with some minor modifications.
Observe that, since F is BD-regular and $\theta \in [0,1)$, then (8) implies that, for k large enough,

$$\begin{aligned}\left\|s^{(k)}\right\| &= \left\|V_k^{-1}V_k s^{(k)}\right\| = \left\|V_k^{-1}\left[V_k s^{(k)} + F\left(x^{(k)}\right)\right] - V_k^{-1}F\left(x^{(k)}\right)\right\| \leq \\ &\leq \|V_k^{-1}\|\,\theta \left\|F\left(x^{(k)}\right)\right\| + \|V_k^{-1}\| \left\|F\left(x^{(k)}\right)\right\| = \\ &= (\theta+1)\|V_k^{-1}\| \left\|F\left(x^{(k)}\right)\right\| \leq 2\|V_k^{-1}\| \left\|F\left(x^{(k)}\right)\right\|.\end{aligned}$$

Therefore, by the above inequality, assumption A and (8), we have for k large enough

$$\begin{aligned}\left\|F\left(x^{(k)}+s^{(k)}\right)\right\| &\leq \left\|F\left(x^{(k)}\right) + V_k s^{(k)}\right\| + o\left(\left\|s^{(k)}\right\|\right) \leq \\ &\leq \theta\left\|F\left(x^{(k)}\right)\right\| + o\left(\left\|F\left(x^{(k)}\right)\right\|\right).\end{aligned}$$

Since $\theta \in [0,1)$ and $\left\|F\left(x^{(k)}\right)\right\| \to 0$, this implies that

$$\lim_{k\to\infty} \frac{\left\|F\left(x^{(k)}+s^{(k)}\right)\right\|}{\left\|F\left(x^{(k)}\right)\right\|} = 0.$$

So, for k large enough

$$\left\|F\left(x^{(k)}+s^{(k)}\right)\right\| \leq \left(1-\frac{\sigma\gamma}{2}\right)\left\|F\left(x^{(k)}\right)\right\|.$$

Therefore (4) holds with $\alpha_k = 1$. Hence for k large enough $x^{(k+1)} = x^{(k)} + s^{(k)}$. Then we have

$$\lim_{k\to\infty} \frac{\left\|F\left(x^{(k+1)}\right)\right\|}{\left\|F\left(x^{(k)}\right)\right\|} = 0. \tag{15}$$

By the Lemma 2 there exists a number $l > 0$ such that

$$\frac{1}{l}\|y - x^*\| \leq \|F(y)\| \leq l\,\|y - x^*\|$$

for all y in a neighborhood of x^*. Then, by (15)

$$\lim_{k\to\infty} \frac{\left\|x^{(k+1)} - x^*\right\|}{\left\|x^{(k)} - x^*\right\|} = 0.$$

■

Remark: If F satisfies A at x^* with degree 2, then we obtain **quadratic** convergence of the algorithm.

Now, we will present Algorithm 2 **which is a particular case** of Algorithm 1, where either conditions (7) and (8) are fulfilled or the execution is stopped. If θ is close to 1 and M is large then failure in **satisfying** (7) and (8) reflects near-stationarity of the current point.

Algorithm 2. (Inexact quasi-Newton method)
Assume that $\theta \in [0,1)$, $\sigma \in (0,1)$, $\gamma = 1-\theta^2$, $\tau_1, \tau_2 \in (0,1)$, $\tau_1 < \tau_2$, $M > 0$ are given independently of k. Let $x^{(0)} \in R^n$ be an arbitrary initial point and $\alpha_0 = 1$. Given a point $x^{(k)}$, the steps for obtaining $x^{(k+1)}$ are:
Step 1. Find some $s_k \in R^n$ such that

$$x^{(k)} + s^{(k)} \in \Omega \text{ and } \left\| s^{(k)} \right\| \le M \tag{16}$$

and

$$\left\| V_k s^{(k)} + F\left(x^{(k)}\right) \right\| \le \theta \left\| F\left(x^{(k)}\right) \right\|, \tag{17}$$

where $V_k \in \partial_B F\left(x^{(k)}\right)$.
If such choice is not possible, the algorithm breaks down.
Step 2. If

$$\left\| F\left(x^{(k)} + \alpha_k s^{(k)}\right) \right\| \le \left\| F\left(x^{(k)}\right) \right\| \tag{18}$$

define

$$x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}. \tag{19}$$

Otherwise set $x^{(k+1)} = x^{(k)}$.
Step 3. If

$$\left\| F\left(x^{(k)} + \alpha_k s^{(k)}\right) \right\| \le \left(1 - \frac{\sigma\gamma\alpha_k}{2}\right) \left\| F\left(x^{(k)}\right) \right\| \tag{20}$$

set $\alpha_{k+1} = 1$. Otherwise choose

$$\alpha_{k+1} \in [\tau_1 \alpha_k, \tau_2 \alpha_k].$$

The proof of the **below theorem follows straightforward from** Theorem 3 and definition of Algorithm 2.

Theorem 4 *Assume that $L_F = \{x \in \Omega : \|F(x)\| \le \|F(x_0)\|\}$ is bounded. Let $\left\{x^{(k)}\right\}$ be the sequence generated by Algorithm 2. Then every limit point of the sequence $\left\{x^{(k)}\right\}$ is a solution of (1). Moreover, if F is BD-regular at x^*, satisfies A at x^* and (9) holds, then $\left\{x^{(k)}\right\}$ converges superlinearly to x^*.*

Remark: Assumption with condition (9) is a weaker version of the assumption (A4) in [13]. Such condition is not required in the smooth case, because the function F and its Jacobian have the strong properties.

3 Nonmonotone version of algorithm

In this section we describe a nonmonotone version of algorithm, modificating the general framework (17) and (20) by substituting these conditions (17) and (20) with the inequalities in which an element $x^{\ell(k)}$ is used. $x^{\ell(k)}$ is the point with the following property

$$\left\| F\left(x^{\ell(k)}\right)\right\| = \max_{0\le j\le \min(\bar{n},k)} \left\| F\left(x^{(k-j)}\right)\right\| \tag{21}$$

for given $\bar{n} \in \mathbb{N}$. Note that $k - \min(\bar{n}, k) \le \ell(k) \le k$.

The nonmonotone approach is well known **for their** effectiveness in the choice of the step in many linesearch procedures (see e.g. [10]). The smooth nonmonotone inexact Newton method was proposed by Bonettini in [4]. Bonettini and Tinti in [5] modified the general inexact Newton algorithm in a nonmonotone way for a semismooth equations. Our approach is similar as both **versions** presented in [4] (smooth) and [5] (semismooth). **In these papers the authors are mainly concerned with convergence of inexact Newton method for solving unconstrained semismooth equations while we study** the inexact quasi-Newton method for the Lipschitz continuous equations with simple constraint. **Moreover, the particular case of the below algorithm that corresponds to $\Omega = R^n$ consists a feature of the acceptance of $x^{(k)}+\alpha_k s^{(k)}$ as new iterate whenever $\left\|F\left(x^{(k)}+\alpha_k s^{(k)}\right)\right\| \le \left\|F\left(x^{\ell(k)}\right)\right\|$. Let us note that this approach allow to alleviate the tendency to taking "smaller than necessary" steps in backtracking.**

Algorithm 3. (Nonmonotone inexact quasi-Newton method)
Assume that $\theta \in [0, 1)$, $\sigma \in (0, 1)$, $\gamma = 1 - \theta^2$, $\tau_1, \tau_2 \in (0, 1)$, $\tau_1 < \tau_2$, $M > 0$ are given independently of k. Let $x^{(0)} \in R^n$ be an arbitrary initial point and $\alpha_0 = 1$. Given a point $x^{(k)}$, the steps for obtaining $x^{(k+1)}$ are:
Step 1. Find some $s_k \in R^n$ such that

$$x^{(k)} + s^{(k)} \in \Omega \text{ and } \left\| s^{(k)} \right\| \le M \tag{22}$$

and

$$\left\| V_k s^{(k)} + F\left(x^{(k)}\right)\right\| \le \theta \left\| F\left(x^{\ell(k)}\right)\right\|, \tag{23}$$

where $V_k \in \partial_B F\left(x^{(k)}\right)$.
If such choice is not possible, the algorithm breaks down.
Step 2. If

$$\left\| F\left(x^{(k)} + \alpha_k s^{(k)}\right)\right\| \le \left\| F\left(x^{\ell(k)}\right)\right\| \tag{24}$$

define

$$x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}. \tag{25}$$

Otherwise set $x^{(k+1)} = x^{(k)}$.
Step 3. If

$$\left\| F\left(x^{(k)} + \alpha_k s^{(k)}\right)\right\| \le \left(1 - \frac{\sigma\gamma\alpha_k}{2}\right) \left\| F\left(x^{\ell(k)}\right)\right\| \tag{26}$$

set $\alpha_{k+1} = 1$. Otherwise choose

$$\alpha_{k+1} \in [\tau_1 \alpha_k, \tau_2 \alpha_k] .$$

Remark: The sequence $\left\{\left\|F\left(x^{(k)}\right)\right\|\right\}$ satisfying (23) and (26) is nonmonotone but $\left\{\left\|F\left(x^{\ell(k)}\right)\right\|\right\}$ is a monotone nonincreasing subsequence of it.

We will assume that at each iteration step k it is possible to obtain the vector $s^{(k)}$ which is an inexact quasi-Newton step for some $V_k \in \partial_B F\left(x^{(k)}\right)$. We can use the following sufficient condition, which is the special case of assumption (A1) given by Martinez, Qi [13]: there exists $\delta \geq 0$ such that for all $x, v \in R^n$, the intersection of the ball $N(v, \delta) = \{u \in R^n : \|u - v\| \leq \delta\}$ and the range set $R(x) = \{u \in R^n : u = V_k s \text{ for some } s \in R^n \text{ and } V_k \in \partial_B F(x)\}$ is not empty.

The **below lemma** shows that the sequence generated by Algorithm 3 satisfies conditions (23) and (24).

Lemma 5 *Let $\theta \in [0, 1)$, $\sigma \in (0, 1)$, $\gamma = 1 - \theta^2$ and $M > 0$. Suppose that exists $\bar{s}$ satisfying*

$$x^{(k)} + \bar{s} \in \Omega \text{ and } \|\bar{s}\| \leq M$$

and

$$\left\|V_k \bar{s} + F\left(x^{(k)}\right)\right\| \leq \theta \left\|F\left(x^{\ell(k)}\right)\right\|$$

for some $V_k \in \partial_B F\left(x^{(k)}\right)$.
Then, there exist $\alpha_{\max} \in (0, 1]$ and a vector s such that

$$\left\|F\left(x^{(k)} + \alpha s\right)\right\| \leq \left(1 - \frac{\sigma\gamma\alpha}{2}\right) \left\|F\left(x^{\ell(k)}\right)\right\|$$

holds for any $\alpha \in (0, \alpha_{\max}]$.

Proof. Let $s = \alpha \bar{s}$. Then we have

$$\begin{aligned}
\left\|V_k s + F(x^{(k)})\right\| &= \left\|\alpha V_k \bar{s} + \alpha F(x^{(k)}) - \alpha F(x^{(k)}) + F(x^{(k)})\right\| \leq \\
&\leq \alpha \left\|V_k \bar{s} + F(x^{(k)})\right\| + (1 - \alpha) \left\|F(x^{(k)})\right\| \leq \\
&\leq \alpha\theta \left\|F(x^{\ell(k)})\right\| + (1 - \alpha) \left\|F(x^{\ell(k)})\right\| = \\
&= [1 - \alpha(1 - \theta)] \left\|F(x^{\ell(k)})\right\| .
\end{aligned}$$

Now, let

$$\varepsilon = \frac{(1 - \theta)[2 - \delta(1 + \theta)]}{2\alpha \|\bar{s}\|} \left\|F(x^{\ell(k)})\right\| \tag{27}$$

and $\delta > 0$ be sufficiently small that

$$\left\|F\left(x^{(k)} + \alpha s\right) - F\left(x^{(k)}\right) - V_k s\right\| \leq \varepsilon\alpha \|s\| \tag{28}$$

for all $V_k \in \partial_B F\left(x^{(k)}\right)$ whenever $\|s\| \leq \delta$.
Choosing $\alpha_{\max} = \min\left(1, \frac{\delta}{\|\bar{s}\|}\right)$, for any $\alpha \in (0, \alpha_{\max}]$ we have $\|s\| \leq \delta$ **and, using** (27) and (28) we obtain the following inequality

$$\begin{aligned}\left\|F\left(x^{(k)}+\alpha s\right)\right\| &\leq \left\|F\left(x^{(k)}+\alpha s\right)-F\left(x^{(k)}\right)-V_k s\right\|+\left\|F\left(x^{(k)}\right)+V_k s\right\| \leq \\ &\leq \varepsilon\alpha^2\|\bar{s}\|+[1-\alpha(1-\theta)]\left\|F(x^{\ell(k)})\right\| = \\ &= \left(1-\frac{\sigma\gamma\alpha}{2}\right)\left\|F\left(x^{\ell(k)}\right)\right\|,\end{aligned}$$

which completes the proof. ■

So, the above lemma yields that Algorithm 3 breaks down **if and only if it is** impossible to find a nonmonotone inexact quasi-Newton step.

Since $\left\{x^{\ell(k)}\right\}$ is a subsequence of $\left\{x^{(k)}\right\}$, also the sequence $\left\{\left\|F\left(x^{\ell(k)}\right)\right\|\right\}$ converges to 0 when k **goes to infinity, the proof** of the below theorem **follows from** Theorem 4 and definition of Algorithm 3.

Theorem 6 *Assume that* $L_F = \{x \in \Omega : \|F(x)\| \leq \|F(x_0)\|\}$ *is bounded. Let* $\theta \in [0,1)$ *and* $\left\{x^{(k)}\right\}$ *be the sequence generated by Algorithm 3 with* $\gamma = 1-\theta^2$. *Assume that there exists* $M > 0$ *such that for all* $k = 0, 1, 2, \ldots$

$$\left\|s^{(k)}\right\| \leq M \tag{29}$$

and

$$\left\|V_k s^{(k)} + F\left(x^{(k)}\right)\right\| \leq \theta\left\|F\left(x^{\ell(k)}\right)\right\|, \tag{30}$$

where $V_k \in \partial_B F\left(x^{(k)}\right)$.
If F *is BD-regular at* x^*, *satisfies assumption A at* x^*, *at each iteration step* k *it is possible to find a vector* $s^{(k)}$ *such that condition (29) is satisfied and for every sequence* $\left\{x^{(k)}\right\}$ *converging to* x^*, *every convergent sequence* $\left\{s^{(k)}\right\}$ *and every sequence* $\{\lambda_k\}$ *of positive scalars converging to* 0

$$\limsup_{k\to\infty}\frac{f\left(x^{(k)}+\lambda_k s^{(k)}\right)-f\left(x^{\ell(k)}\right)}{\lambda_k} \leq \lim_{k\to\infty} F\left(x^{(k)}\right)^T V_k s^{(k)}, \tag{31}$$

whenever the limit in the left-hand side exists, then every limit point of the sequence $\left\{x^{(k)}\right\}$ *is a solution of equation (1) and* $\left\{x^{(k)}\right\}$ *converges superlinearly to* x^*.

4 Numerical examples

In this section, we present some preliminary numerical results for constructed algorithms. We solved some nonsmooth equation from Spedicato [21] and the box-constrained nonlinear system related to the computation of singular points of homotopic paths (defined in [11]). In the last one we used second test problem taken from collection of Melhem and Rheinboldt [14].

All the experiments were performed on a Pentium IV 2.4 GHz using Dev-C++ and double precision arithmetic. The parameters used in Algorithms 2 and 3 are specified as follows:

$$\theta = 0.999,\ \sigma = 10^{-3},\ \tau_1 = \tau_2 = 0.5 \text{ and } M = 10.$$

Moreover, we declare a failure of the algorithm when the stopping criterion $\left\|F\left(x^{(k)}\right)\right\| \leq 10^{-10}$ is not reached after 1000 iterations or when, in order to satisfy the backtracking condition (20) or (26), more than 25 reductions of the parameter α_k have been performed. Tables 1 and 2 summarize the results in terms of number of iterations and of backtracking reductions, reported in the rows with the "iter" and "back" symbols, respectively. Our aim is to compare the performances of Algorithm 2 (monotone case) and the ones of Algorithm 3 with different nonmonotonicity degrees. For the nonmonotone algorithm the parameter $\bar{n}$ has been chosen equal 2, 5 and 8.

Example 1. Consider the equation (1) with function $F : R^n \to R^n$ defined by

$$F^i(x) = \begin{cases} c_1 g_i(x) \text{ for } g_i(x) \geq 0, \\ c_2 g_i(x) \text{ for } g_i(x) \leq 0, \end{cases}$$

where

$$g_i(x) = i - \sum_{j=1}^{i} \left\{\cos(x_j - 1) + j\left[1 - \cos(x_j - 1)\right] - \sin(x_j - 1)\right\}\ .$$

If $c_1 = c_2$, F is differentiable. Therefore $|c_1 - c_2|$ may be interpreted as the degree of nondifferentiability of F. See [21]. The system $F(x) = 0$ has the solutions $(1 + 2k_1\pi, ..., 1 + 2k_n\pi)^T$,where $k_1, ..., k_n$ are arbitrary integers. We executed both algorithms for three nonsmooth cases: $c_1 = -c_2 = 1, 10, 100$ with $\Omega = \{-100 \leq x_i \leq 100, i = 1, ..., n\}$. Table 2 shows the nonmonotone scheme differs to the monotone one only on the backtracking rule for larger systems.

Example 2. Given $H : R^{m+1} \to R^m$, $H = H(y, t)$, we say that (y_*, t_*) is a singular point of $H(y, t) = 0$ if $H(y_*, t_*) = 0$ and $H_y(y_*, t_*)$ is singular. Singular points are solutions of system

$$\begin{aligned} H(y,t) &= 0 \\ H_y(y,t)\,v &= 0 \\ \|v\|_2 &= 1 \end{aligned}$$

which has $2m{+}1$ equations and unknowns. We used the problem with Freudenstein-Roth function ($m = 2$, $n = 5$)

$$\begin{aligned} h_1(y,t) &= y_1 - y_2^3 + 5y_2^2 - 2y_2 - 13 + 34\,(t-1) \\ h_2(y,t) &= y_1 + y_2^3 + y_2^2 - 14y_2 - 29 + 10\,(t-1) \end{aligned}$$

with $\Omega = \{-100 \leq y_1, y_2 \leq 100 \text{ and } -10 \leq t \leq 10\}$. Table 3 shows that the general reduction of the number of iteration can be observed in nonmonotone approach.

Table 1. Numerical results for Example 1.

		$c_1 = -c_2 = 1$			$c_1 = -c_2 = 10$			$c_1 = -c_2 = 100$		
n		$\bar{n}$=0	$\bar{n}$=2	$\bar{n}$=5	$\bar{n}$=0	$\bar{n}$=2	$\bar{n}$=5	$\bar{n}$=0	$\bar{n}$=2	$\bar{n}$=5
2	iter	6	6	6	6	6	6	6	6	6
	back	0	0	0	0	0	0	0	0	0
3	iter	6	6	6	6	6	6	7	7	7
	back	0	0	0	0	0	0	0	0	0
4	iter	7	7	7	7	7	7	8	8	8
	back	0	0	0	0	0	0	0	0	0
5	iter	8	8	8	8	8	8	8	8	8
	back	0	0	0	0	0	0	0	0	0
8	iter	9	9	9	9	9	8	9	9	8
	back	0	0	0	0	0	0	1	0	0
10	iter	10	10	9	10	10	9	10	10	9
	back	0	0	0	1	0	0	1	1	0
12	iter	10	10	9	10	10	9	10	9	9
	back	1	0	0	1	0	0	1	1	0
15	iter	11	12	10	11	10	10	11	10	10
	back	1	0	0	1	1	0	2	1	1
20	iter	12	12	11	12	11	11	12	11	11
	back	4	1	1	5	1	1	7	2	2

Table 2. Numerical results for Freudenstein-Roth function (Example 2)

x_0		$\bar{n}$=0	$\bar{n}$=2	$\bar{n}$=5	$\bar{n}$=8
1	iter	35	29	29	43
	back	1	1	1	2
2	iter	20	18	17	18
	back	0	0	0	0

5 Conclusions

A family of Newton-type methods is important for solving nonlinear equations. The are especially useful when the system has many variables and inexact approach is practical. In this paper, we have studied the new version of inexact quasi-Newton method for solving nonsmooth equations with simple constraint. We have first proved that under mild assumptions, the every limit point of sequence generated by the inexact quasi-Newton algorithm is solution of equation (1) and this sequence is globally and superlinearly convergent. Then we proposed the nonmonotone technique which can reduce the number of steps of iteration. The numerical experiments with the inexact quasi-Newton method for solving some constrained equations are promising. The numerical tests show that the nonmonotone approach can produce a sensible decrease both the number of iterations and of backtracking reductions. However, a degenerate behaviour of the algorithm can be observed in some problems for a too large value of parameter $\bar{n}$.

References

[1] Bergamaschi, L., Bru, R., Martínez, A., Putti, M., 2006, Quasi-Newton preconditioners for the inexact Newton method. *Electron. Trans. Numer. Anal.,* **23**, 76-87.

[2] Bergamaschi, L., Moret, I., Zilli, G., 2001, Inexact quasi-Newton methods for sparse systems of nonlinear equations. *Future Generat. Comput. Syst.*, **18**, 41-53.

[3] Birgin, E.G., Krejić, N., Martínez, J.M., 2003, Globally convergent inexact quasi-Newton methods for solving nonlinear systems. *Numer. Algorithms,* **32**, 249-260.

[4] Bonettini, S., 2005, A nonmonotone inexact Newton method. *Optim. Meth. Software*, **20**, 475-491.

[5] Bonettini, S., Tinti, F., 2005, A nonmonotone semismooth inexact Newton method. Technical Report no. 70, Department of Mathematics, University of Modena and Reggio Emilia, Italy.

[6] Clarke, F.H., 1990, *Optimization and Nonsmooth Analysis* (Philadelphia: SIAM).

[7] Dembo, R.S., Eisenstat, S.C., Steihaug, T., 1982, Inexact Newton methods. *SIAM J. Numer. Anal.*, 14, 400-408.

[8] Eisenstat, S.C., Walker, H.F., 1994, Globally convergent inexact Newton methods. *SIAM J. Optimiz.*, **4**, 393-422.

[9] Gowda, M.S., Ravindran, G., 2000, On the characterizations of P- and P_0-properties in nonsmooth functions. *Math. Oper. Res.*, **25**, 400-408.

[10] Grippo, L., Lampariello, F., Lucidi S., 1986, A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, **23**, 707-716.

[11] Kozakevich, D.N., Martínez, J.M., Santos, S.A., 1997, Solving nonlinear systems of equations with simple constraints. *Comput. Appl. Math.*, **16**, 215-235.

[12] Martínez, J.M., 1997, Quasi-inexact-Newton methods with global convergence for solving constrained nonlinear systems. *Nonlinear Analysis, Theory, Methods and Applications*, **30**, 1-8.

[13] Martínez, J.M., Qi, L., 1995, Inexact Newton method for solving nonsmooth equations. *J. Comput. Appl. Math.*, **60**, 127-145.

[14] Melhem, R.G., Rheinboldt, W.C., 1982, A comparison of methods for determining turning pons of nonlinear equations. *Computing,* **29**, 201-226.

[15] Mifflin, R., 1977, Semismooth and Semiconvex Functions in Constrained Optimization. *SIAM J. Control and Optimization,* **15**, 959-972.

[16] Pang, J.S., Qi, L., 1993, Nonsmooth equations: motivation and algorithms. *SIAM J. Optim.*, **3**, 443-465.

[17] Pu, D., Tian, W., 2002, Globally convergent inexact generalized Newton's methods for nonsmooth equations. *J. Comput. Appl. Math.*, **138**, 37-49.

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

[19] Qi, L., 1996, C-differential operators, C-differentiability and generalized Newton methods. Applied Mathematics Report AMR 96/5, University of New South Wales, Australia.

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

[21] Spedicato, E., 1975, Computational experience with quasi-Newton algorithms for minimization problems of moderately large size. Report CISE-N-175, Segrate, Milano.