Reformulation: Nonsmooth, Piecewise Smooth,
Semismooth and Smoothing Methods, pp. 355–369
Edited by M. Fukushima and L. Qi

A Globally Convergent Inexact Newton Method for Systems of Monotone Equations [1]

M. V. Solodov† and B. F. Svaiter†

Abstract We propose an algorithm for solving systems of monotone equations which combines Newton, proximal point, and projection methodologies. An important property of the algorithm is that the whole sequence of iterates is always globally convergent to a solution of the system without any additional regularity assumptions. Moreover, under standard assumptions the local superlinear rate of convergence is achieved. As opposed to classical globalization strategies for Newton methods, for computing the stepsize we do not use linesearch aimed at decreasing the value of some merit function. Instead, linesearch in the approximate Newton direction is used to construct an appropriate hyperplane which separates the current iterate from the solution set. This step is followed by projecting the current iterate onto this hyperplane, which ensures global convergence of the algorithm. Computational cost of each iteration of our method is of the same order as that of the classical damped Newton method. The crucial advantage is that our method is truly globally convergent. In particular, it cannot get trapped in a stationary point of a merit function. The presented algorithm is motivated by the hybrid projection-proximal point method proposed in [25].

Key Words nonlinear equations, Newton method, proximal point method, projection method, global convergence, superlinear convergence.

---

[1]Research of the first author is supported by CNPq Grant 300734/95-6 and by PRONEX–Optimization, research of the second author is supported by CNPq Grant 301200/93-9(RN) and by PRONEX–Optimization.
†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email : solodov@impa.br and benar@impa.br.

1 INTRODUCTION

We consider the problem of finding solutions of systems of nonlinear equations

$$F(x) = 0, \tag{1.1}$$

where $F : \Re^n \to \Re^n$ is continuous and monotone, i.e.

$$\langle F(x) - F(y), x - y\rangle \geq 0 \quad \text{for all} \quad x, y \in \Re^n.$$

Note that under this assumption, the solution set of (1.1) is convex. Systems of monotone equations arise in various applications. One important example is subproblems in the generalized proximal algorithms with Bregman distances [9, 7].

Among numerous algorithms for solving systems of equations, the Newton method and its variants are of particular importance [16, 5, 19, 2]. Given a point x, the Newton method generates the next iterate by

$$x^{new} = x + d,$$

where d is a (possibly approximate) solution of the system of linear equations

$$\nabla F(x)d = -F(x),$$

with $\nabla F(x)$ being the Jacobian of F at the point x (assuming F is differentiable). If the starting point is sufficiently close to some solution $\bar{x}$ of (1.1) where $\nabla F(\bar{x})$ is nonsingular, the sequence generated by the Newton method converges superlinearly or quadratically, depending on further assumptions. Similar local results hold in the more general case of semismooth equations (see [20, 21, 18, 15, 10]). Having in mind the problem under consideration, it is worth to mention that monotonicity of F is not needed for such local analysis.

To enlarge the domain of convergence of the Newton method, some globalization strategy has to be used. The most common globalization strategy is the damped Newton method

$$x^{new} = x + \alpha d$$

which employs a linesearch procedure along the Newton direction d to compute the stepsize $\alpha > 0$. This linesearch is typically based on an Armijo-type [1] sufficient descent condition for some merit function, usually the squared 2-norm merit function

$$f(x) := \|F(x)\|^2.$$

Sometimes also the damped Gauss-Newton method is used, where the search direction d is obtained by solving the regularized linear system

$$(\nabla F(x)^\top \nabla F(x) + \mu I)d = -\nabla F(x)^\top F(x), \quad \mu > 0.$$

The latter method has the advantage that the search direction always exists, even if $\nabla F(x)$ is singular. Note that in the case of semismooth equations, some additional assumptions are needed to apply the linesearch globalization strategy (see [11]).

To motivate the development of our algorithm, we emphasize the following drawbacks of damped Newton and damped Gauss-Newton methods :

- Either method can only ensure that all accumulation points of the generated sequence of iterates are stationary points of the merit function $f(x)$. In the absence of regularity, there is no guarantee that these stationary points are solutions of the original problem (1.1).

- Moreover, even to ensure the existence of such accumulation points, some additional assumptions are needed (for example, on boundedness of the level sets of f and, hence, of the solution set of the system of equations).

- Without some regularity conditions, one cannot prove convergence of the *whole* sequence of iterates even assuming that the sequence of merit function values actually goes to zero (so that the iterates do approach the solution set of the problem).

Note that by itself, monotonicity of F does not help to resolve any of the above listed difficulties. For example, $f(x) = \|F(x)\|^2$ may still have stationary points which are not solutions of the system of equations $F(x) = 0$. In addition, if the solution set of (1.1) is not a singleton, regularity conditions do not hold.

In this paper, we present a Newton-type algorithm for solving systems of monotone equations which overcomes the above mentioned drawbacks. In particular, under the assumption of continuity and monotonicity of F, from any starting point the *whole* sequence of iterates converges to a solution of the problem, provided one exists (neither regularity nor boundedness of the solution set are needed for global convergence). Under the assumptions of differentiability and nonsingularity, we prove superlinear convergence of the inexact version of the algorithm, similar to the classical inexact Newton method [4].

Our algorithm is motivated, to some extent, by the hybrid projection-proximal point method proposed in [25] in the more general context of finding zeroes of set-valued maximal monotone operators in a Hilbert space. Let T be a maximal monotone operator on a real Hilbert space $\mathcal{H}$. And consider, for a moment, the problem of finding an $x \in \mathcal{H}$ such that $0 \in T(x)$.

Algorithm 1.1 (Hybrid Projection-Proximal Point Method) *[25]*
Choose any $x^0 \in \mathcal{H}$ and $\sigma \in [0,1)$; set $k := 0$.

Inexact proximal step. *Choose $\mu_k > 0$ and find $y^k \in \mathcal{H}$ and $v^k \in T(y^k)$ such that*

$$0 = v^k + \mu_k(y^k - x^k) + \varepsilon^k,$$

where

$$\|\varepsilon^k\| \leq \sigma \max\{\|v^k\|, \mu_k\|y^k - x^k\|\}.$$

Stop if $v^k = 0$ or $y^k = x^k$. Otherwise,

Projection step. *Compute*

$$x^{k+1} = x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k.$$

Set $k := k + 1$*; and repeat.*

If problem $0 \in T(x)$ has a solution and the sequence $\{\mu_k\}$ is bounded above, then the generated sequence $\{x^k\}$ either is finite and terminates at a solution, or it is infinite and converges (weakly) to a solution (for complete properties of the method, see [25]). The idea of Algorithm 1.1 is to use an approximate proximal iteration to construct a hyperplane

$$H_k := \{x \in \Re^n \mid \langle v^k, x - y^k \rangle = 0\},$$

which separates the current iterate x^k from the solutions of $0 \in T(x)$. The last step of Algorithm 1.1 is equivalent to projecting x^k onto this hyperplane. Separation arguments show that the distance to the solution set for thus constructed sequence monotonically decreases, which essentially ensures global convergence of the algorithm. The advantage of Algorithm 1.1 over the classical proximal point method [22] (of course, we are talking about the general context of operator equations here) is that the condition imposed on ε^k in the inexact proximal step is significantly less restrictive (and more constructive) than the corresponding tolerance requirements in the standard proximal point settings (see [22, 3]) :

$$\|\varepsilon^k\| \le \sigma_k \mu_k, \quad \sum_{k=0}^{\infty} \sigma_k < \infty$$

or

$$\|\varepsilon^k\| \le \sigma_k \mu_k \|y^k - x^k\|, \quad \sum_{k=0}^{\infty} \sigma_k < \infty.$$

Note that in the classical approach one further sets $x^{k+1} := y^k$ to obtain the next iterate. For one thing, the right-hand-side of the corresponding tolerance inequality in Algorithm 1.1 involves the *largest* of two quantities, $\|v^k\|$ and $\mu_k \|y^k - x^k\|$. And more importantly, the tolerance parameter σ can be fixed, which means that the *relative error* in solving the proximal subproblems has to be merely bounded, while in the classical setting it has to be summable (hence tend to zero). Thus Algorithm 1.1 has better robustness features while preserving computational costs and convergence properties of the classical proximal point method (computational costs are actually expected to be reduced, thanks to the relaxed tolerance requirements). We refer the readers to [25] for complete analysis of Algorithm 1.1 and a more detailed comparison with the classical proximal point method. Another related work using inexact proximal iterations with relative error bounded away from zero is [24].

Now let us go back to our problem (1.1). To ensure global convergence to a solution, one can apply Algorithm 1.1, setting $T = F$. However, it is clear that this approach, without modifications, is not quite practical for the following reasons. The resulting proximal subproblem $F(x) + \mu_k(x - x^k) = 0$ is still a rather difficult nonlinear problem, albeit with better nondegeneracy properties compared to the original equation. Solving this nonlinear equation even with the more realistic error tolerance of Algorithm 1.1, can be quite difficult. Therefore in this paper, we propose an algorithm where an approximate proximal subproblem solution is obtained by a single Newton step. Moreover, the (linear) regularized Newton equation itself can be solved inexactly (see Algorithm 2.1). Because such step may fail to generate an appropriate separating hyperplane (i.e. the tolerance requirements of Algorithm 1.1 need not be met), a linesearch procedure similar to the one in [23] is employed. Finally, as in Algorithm 1.1, a projection step is made. This hybrid algorithm is globally convergent to a solution of the system of equations, provided one exists, under no assumptions on F other than continuity and monotonicity (see Theorem 2.1). When F is differentiable at this solution and the Jacobian is nonsingular, the rate of convergence is superlinear (see Theorem 2.2). Thus attractive global convergence properties of proximal-like methods are successfully combined with fast local convergence of Newton-type methods. Globalization is achieved via a linesearch and projection strategy. The total computational cost of one iteration of the algorithm is of the same order as that of the standard damped Newton method.

2 THE ALGORITHM AND ITS CONVERGENCE

We now describe our algorithm in detail. Given a current iterate x^k and a regularization parameter $\mu_k > 0$, consider the proximal point subproblem (see [22, 6, 8, 12, 13]),

$$0 = F(x) + \mu_k(x - x^k).$$

The first step consists of solving the linearization of this subproblem at the point x^k

$$0 = F(x^k) + G_k(x - x^k) + \mu_k(x - x^k),$$

where G_k is a positive semidefinite matrix (this is similar to standard Newton-proximal point approaches, for example, [9]). We allow this Newton-type linear equation to be solved approximately (see Algorithm 2.1), much in the spirit of inexact Newton methods (see [4, 26, 14, 17, 15]). This feature is of particular importance for large-scale problems.

Because a full step in the obtained Newton direction may not satisfy the tolerance conditions imposed on solving the proximal subproblems in Algorithm 1.1, we cannot immediately perform the projection step. Thus the second step of the method is a linesearch procedure in the Newton direction d^k (see Algorithm 2.1) which computes a point $y^k = x^k + \alpha_k d^k$ such that

$$0 < \langle F(y^k), x^k - y^k \rangle.$$

A similar linesearch technique was used in [23]. Note that by monotonicity of F, for any $\bar{x}$ such that $F(\bar{x}) = 0$ we have

$$0 \geq \langle F(y^k), \bar{x} - y^k \rangle.$$

Thus the hyperplane

$$H_k := \{x \in \Re^n \mid \langle F(y^k), x - y^k \rangle = 0\}$$

strictly separates the current iterate x^k from zeroes of the system of equations. Once the separating hyperplane is obtained, the next iterate x^{k+1} is computed by projecting x^k onto it.

We now formally state the algorithm.

Algorithm 2.1 *Choose any* $x^0 \in \Re^n$, $\beta \in (0,1)$ *and* $\lambda \in (0,1)$*; set* $k := 0$.

Inexact Newton step. *Choose a positive semidefinite matrix* G_k*. Choose* $\mu_k > 0$ *and* $\rho_k \in [0,1)$*. Compute* $d^k \in \Re^n$ *such that*

$$0 = F(x^k) + (G_k + \mu_k I)d^k + e^k, \tag{2.1}$$

where

$$\|e^k\| \leq \rho_k \mu_k \|d^k\|.$$

Stop if $d^k = 0$*. Otherwise,*

Linesearch step. *Find* $y^k = x^k + \alpha_k d^k$*, where* $\alpha_k = \beta^{m_k}$ *with* m_k *being the smallest nonnegative integer* m *such that*

$$-\langle F(x^k + \beta^m d^k), d^k \rangle \geq \lambda(1 - \rho_k)\mu_k \|d^k\|^2. \tag{2.2}$$

Projection step. *Compute*

$$x^{k+1} = x^k - \frac{\langle F(y^k), x^k - y^k \rangle}{\|F(y^k)\|^2} F(y^k). \tag{2.3}$$

Set $k := k + 1$*; and repeat.*

We note that Algorithm 2.1 has computational costs per iteration comparable to those of damped (Gauss-) Newton method : solving a system of linear equations followed by a linesearch procedure (the projection step is explicit and hence computationally negligible). Algorithm 2.1 has the advantage that the whole sequence of iterates is globally convergent to a solution of the system of equations under no regularity assumptions (Theorem 2.1). When F is differentiable at this solution, ∇F is nonsingular, and the parameters of the algorithm are set properly, fast superlinear rate of convergence is obtained (Theorem 2.2).

We start with a preliminary result.

Lemma 2.1 *Let* F *be monotone and* $x, y \in \Re^n$ *be such that*

$$\langle F(y), x - y \rangle > 0.$$

Let

$$x^+ = x - \frac{\langle F(y), x - y \rangle}{\|F(y)\|^2} F(y).$$

Then for any $\bar{x} \in \Re^n$ *such that* $F(\bar{x}) = 0$*, it holds that*

$$\|x^+ - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - \|x^+ - x\|^2.$$

Proof. Let $\bar{x} \in \Re^n$ be any point such that $F(\bar{x}) = 0$. By monotonicity of F,

$$\langle F(y), \bar{x} - y \rangle \leq 0.$$

It follows from the hypothesis that the hyperplane

$$H := \{s \in \Re^n \mid \langle F(y), s - y \rangle = 0\}$$

strictly separates x from $\bar{x}$. It is also easy to verify that x^+ is the projection of x onto the halfspace $\{s \in \Re^n \mid \langle F(y), s - y \rangle \leq 0\}$. Since $\bar{x}$ belongs to this halfspace, it follows from the basic properties of the projection operator (see [19, p.121]) that $\langle x - x^+, x^+ - \bar{x} \rangle \geq 0$. Therefore

$$\begin{aligned}
\|x - \bar{x}\|^2 &= \|x - x^+\|^2 + \|x^+ - \bar{x}\|^2 + 2\langle x - x^+, x^+ - \bar{x} \rangle \\
&\geq \|x - x^+\|^2 + \|x^+ - \bar{x}\|^2 \\
&= \left(\frac{\langle F(y), x - y \rangle}{\|F(y)\|} \right)^2 + \|x^+ - \bar{x}\|^2.
\end{aligned}$$

■

We are now ready to prove our main global convergence result. Throughout we assume that the solution set of the problem is nonempty.

Theorem 2.1 *Suppose that* F *is continuous and monotone and let* $\{x^k\}$ *be any sequence generated by Algorithm 2.1.*

For any $\bar{x}$ *such that* $F(\bar{x}) = 0$*, it holds that*

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2.$$

In particular, $\{x^k\}$ *is bounded. Furthermore, it holds that either* $\{x^k\}$ *is finite and the last iterate is a solution, or the sequence is infinite and* $0 = \lim_{k\to\infty} \|x^{k+1} - x^k\|$.

Suppose the sequence $\{x^k\}$ *is infinite. Suppose that* $1 > \limsup_{k\to\infty} \rho_k$ *and there exist constants* $C_1, C_2, C_3 > 0$ *such that* $\|G_k\| \leq C_1$ *for all* k *and, starting with some index* k_0, $C_2 \geq \mu_k \geq C_3\|F(x^k)\|$. *Then* $\{x^k\}$ *converges to some* $\bar{x}$ *such that* $F(\bar{x}) = 0$.

Proof. First note that if the algorithm terminates at some iteration k then $d^k = 0$, and consequently, $e^k = 0$. It follows from (2.1) that $F(x^k) = 0$, so that x^k is a solution. From now on, we assume that $d^k \neq 0$ for all k. We next show that the method is well-defined and an infinite sequence $\{x^k\}$ is generated.

Since G_k is positive semidefinite, the regularized Newton equation is always solvable. Hence the inexact Newton step (2.1) is well-defined.

We now show that the linesearch procedure (2.2) always terminates with a positive stepsize α_k. Suppose that for some iteration index k this is not the case. That is, for all integers m we have

$$-\langle F(x^k + \beta^m d^k), d^k \rangle < \lambda(1-\rho_k)\mu_k \|d^k\|^2. \tag{2.4}$$

We further obtain

$$\begin{aligned} -\lim_{m\to\infty} \langle F(x^k + \beta^m d^k), d^k \rangle &= -\langle F(x^k), d^k \rangle \\ &= \langle (G_k + \mu_k I)d^k + e^k, d^k \rangle \\ &\geq \mu_k \|d^k\|^2 - \|e^k\| \|d^k\| \\ &\geq (1-\rho_k)\mu_k \|d^k\|^2, \end{aligned}$$

where the second equality and the last inequality follow from (2.1), and the first inequality follows from positive semidefiniteness of G_k and the Cauchy-Schwarz inequality. Now taking the limits as $m \to \infty$ in both sides of (2.4) implies that $\lambda \geq 1$, which contradicts the choice of $\lambda \in (0,1)$. It follows that the linesearch step (and hence the whole algorithm) is well-defined.

By (2.2) we have

$$\langle F(y^k), x^k - y^k \rangle = -\alpha_k \langle F(y^k), d^k \rangle \geq \lambda(1-\rho_k)\mu_k \alpha_k \|d^k\|^2 > 0. \tag{2.5}$$

Let $\bar{x}$ be any point such that $F(\bar{x}) = 0$. By (2.3), (2.5) and Lemma 2.1, it follows that

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2. \tag{2.6}$$

Hence the sequence $\{\|x^k - \bar{x}\|\}$ is nonincreasing and convergent, therefore the sequence $\{x^k\}$ is bounded, and also

$$0 = \lim_{k\to\infty} \|x^{k+1} - x^k\|. \tag{2.7}$$

By (2.1), the triangle inequality and the choice of μ_k for $k \geq k_0$, we have

$$\begin{aligned} \|F(x^k)\| &\geq \|(G_k + \mu_k I)d^k\| - \|e^k\| \\ &\geq (1-\rho_k)\mu_k \|d^k\| \\ &\geq (1-\rho_k)C_3 \|F(x^k)\| \|d^k\|. \end{aligned}$$

It follows that the sequence $\{d^k\}$ is bounded, hence so is $\{y^k\}$. Now by continuity of F and condition $1 > \limsup_{k\to\infty} \rho_k$, there exists $C_4 > 0$ such that $\lambda(1-\rho_k)\|F(y^k)\|^{-1} \geq C_4$. Using (2.3) and (2.5), we thus obtain

$$\begin{aligned} \|x^{k+1} - x^k\| &= \frac{\langle F(y^k), x^k - y^k \rangle}{\|F(y^k)\|} \\ &\geq C_4 \mu_k \alpha_k \|d^k\|^2. \end{aligned}$$

From the latter relation and (2.7) it follows that

$$0 = \lim_{k\to\infty} \mu_k \alpha_k \|d^k\|^2. \tag{2.8}$$

We consider the two possible cases :

$$0 = \liminf_{k\to\infty} \|F(x^k)\| \quad \text{and} \quad 0 < \liminf_{k\to\infty} \|F(x^k)\|.$$

In the first case, continuity of F implies that the sequence $\{x^k\}$ has some accumulation point $\hat{x}$ such that $F(\hat{x}) = 0$ (recall that $\{x^k\}$ is bounded). Since $\bar{x}$ was an *arbitrary* solution, we can choose $\bar{x} = \hat{x}$ in (2.6). The sequence $\{\|x^k - \hat{x}\|\}$ converges, and since $\hat{x}$ is an accumulation point of $\{x^k\}$, it must be the case that $\{x^k\}$ converges to $\hat{x}$.

Consider now the second case. From the choice of μ_k for $k \geq k_0$,

$$\liminf_{k\to\infty} \mu_k \geq C_3 \liminf_{k\to\infty} \|F(x^k)\| > 0.$$

Furthermore, by (2.1) and the triangle inequality, we have

$$\begin{aligned} \|F(x^k)\| &\leq \|(G_k + \mu_k I)d^k\| + \|e^k\| \\ &\leq (\|G_k\| + \mu_k + \rho_k\mu_k)\, \|d^k\| \\ &\leq (C_1 + 2C_2)\|d^k\|. \end{aligned}$$

Hence,

$$0 < \liminf_{k\to\infty} \|d^k\|.$$

Then, by (2.8), it must hold that

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

The latter is equivalent to saying that $m_k \to \infty$. By the stepsize rule we have that (2.2) is not satisfied for the value of $\beta^{m_k - 1}$, i.e.

$$-\langle F(x^k + \beta^{m_k-1} d^k), d^k\rangle < \lambda(1-\rho_k)\mu_k\|d^k\|^2.$$

Taking into account boundedness of the sequences $\{x^k\}$, $\{\mu_k\}$, $\{\rho_k\}$ and $\{d^k\}$, and passing onto a subsequence if necessary, as $k \to \infty$ we obtain

$$-\langle F(\hat{x}), \hat{d}\rangle \leq \lambda(1-\hat{\rho})\hat{\mu}\|\hat{d}\|^2,$$

where $\hat{x}$, $\hat{\mu}$, $\hat{\rho}$ and $\hat{d}$ are limits of corresponding subsequences. On the other hand, by (2.1) and already familiar argument,

$$-\langle F(\hat{x}), \hat{d}\rangle \geq (1-\hat{\rho})\hat{\mu}\|\hat{d}\|^2.$$

Taking into account that $\hat{\mu}\|\hat{d}\| > 0$ and $\hat{\rho} \leq \limsup_{k\to\infty} \rho_k < 1$, the last two relations are a contradiction because $\lambda \in (0,1)$. Hence the case $0 < \liminf_{k\to\infty} \|F(x^k)\|$ is not possible.

This completes the proof. ■

Remark. It can be observed from the proof of Theorem 2.1 that conditions imposed on the choice of μ_k (even though they are not restrictive) can be replaced by more general ones. Let t_1 and t_2 be two nonnegative monotonically increasing functions (not necessarily continuous) such that $t_1(0) = t_2(0) = 0$ and $t_2(s) \geq t_1(s) > 0$ for all $s > 0$. If μ_k is such that

$$t_1(\|F(x^k)\|) \leq \mu_k \leq t_2(\|F(x^k)\|),$$

then the global convergence result still holds.

We now turn our attention to local rate of convergence analysis.

Theorem 2.2 *Let F be monotone and continuous on $\Re^n$. Let $\bar{x}$ be the (unique) solution of (1.1) at which F is differentiable with $\nabla F(\bar{x})$ nonsingular. Let ∇F be locally Lipschitz continuous around $\bar{x}$. Suppose that starting with some index k_0, $G_k = \nabla F(x^k)$. Suppose that*

$$0 = \lim_{k\to\infty} \mu_k = \lim_{k\to\infty} \rho_k = \lim_{k\to\infty} \mu_k^{-1}\|F(x^k)\|. \tag{2.9}$$

Then the sequence $\{x^k\}$ converges to $\bar{x}$ Q-superlinearly.

If starting with some index k_0, $\mu_k = \gamma_1\|F(x^k)\|^{1/2}$ and $\rho_k = \gamma_2\|F(x^k)\|^{1/2}$ for some positive constants γ_1 and γ_2, then the order of superlinear convergence is at least 1.5 .

Proof. By Theorem 2.1, we already know that the sequence $\{x^k\}$ converges to $\bar{x}$ (note that monotonicity of F and nonsingularity of $\nabla F(\bar{x})$ imply that $\bar{x}$ is the unique solution).

Since $\{x^k\}$ converges to $\bar{x}$ and $G_k = \nabla F(x^k)$ starting with some index k_0, nonsingularity of $\nabla F(\bar{x})$ implies that there exists a constant $C_5 > 0$ such that $\|G_k z\| \geq C_5\|z\|$ for all $z \in \Re^n$ and all k sufficiently large. By (2.1) and the triangle inequality, we have

$$\begin{aligned} \|F(x^k)\| &\geq \|(G_k + \mu_k I)d^k\| - \|e^k\| \\ &\geq \|G_k d^k\| - \|e^k\| \\ &\geq C_5\|d^k\| - \rho_k\mu_k\|d^k\|. \end{aligned}$$

Since $0 = \lim_{k\to\infty} \rho_k\mu_k$, it follows that for some $C_6 > 0$

$$\|d^k\| \leq C_6\|F(x^k)\|. \tag{2.10}$$

In particular, the sequence $\{d^k\}$ converges to zero. By Lipschitz continuity of ∇F in the neighbourhood of $\bar{x}$, it follows that there exists $C_7 > 0$ such that

$$\begin{aligned} &F(x^k + d^k) = F(x^k) + \nabla F(x^k)d^k + R^k, \\ &\|R^k\| \leq C_7\|d^k\|^2 \end{aligned} \tag{2.11}$$

for k large enough.

We next establish that for k large, the unit stepsize is always accepted. By (2.1) and (2.11) we have

$$F(x^k + d^k) = -\mu_k d^k - e^k + R^k. \tag{2.12}$$

Using the Cauchy-Schwarz inequality, for k large enough we obtain

$$\begin{aligned} -\langle F(x^k + d^k), d^k\rangle &= \langle \mu_k d^k + e^k - R^k, d^k\rangle \\ &\geq \mu_k \|d^k\|^2 - (\|e^k\| + \|R^k\|)\|d^k\| \\ &\geq (1-\rho_k)\mu_k\|d^k\|^2 - C_7\|d^k\|^3 \\ &= \left[1 - \frac{C_7\|d^k\|}{(1-\rho_k)\mu_k}\right](1-\rho_k)\mu_k\|d^k\|^2. \end{aligned}$$

Note that from (2.10) and (2.9), it follows that

$$\lim_{k\to\infty} \frac{C_7\|d^k\|}{(1-\rho_k)\mu_k} \leq C_7 C_6 \lim_{k\to\infty} \frac{\|F(x^k)\|}{(1-\rho_k)\mu_k} = 0.$$

Therefore (recall that $\lambda < 1$), for k sufficiently large it holds that

$$-\langle F(x^k + d^k), d^k\rangle > \lambda(1-\rho_k)\mu_k\|d^k\|^2,$$

i.e. condition (2.2) is satisfied for $m = 0$. Hence $\alpha_k = 1$ and $y^k = x^k + d^k$ for all indices k large enough.

From now on, we assume that k is large enough so that (2.11) and (2.12) hold, and $y^k = x^k + d^k$.

Denote

$$s^k = -(\nabla F(x^k) + \mu_k I)^{-1} F(x^k) \quad \text{and} \quad z^k = x^k + s^k.$$

By the triangle inequality,

$$\|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - y^k\| + \|y^k - z^k\| + \|z^k - \bar{x}\|. \tag{2.13}$$

We proceed to analyze the three terms in the right-hand-side of the above inequality separately.

We start with the first term. Since the point x^{k+1} is the projection of x^k onto the hyperplane H_k (defined previously), and $y^k \in H_k$, the vectors $x^{k+1} - x^k$ and $x^{k+1} - y^k$ are orthogonal. Hence

$$\|x^{k+1} - y^k\| = \|y^k - x^k\| \sin\theta_k = \|d^k\| \sin\theta_k,$$

where θ_k is the angle between $x^{k+1} - x^k$ and $y^k - x^k$. Because $x^{k+1} - x^k = -t_k F(y^k)$ for a certain $t_k > 0$ and $y^k - x^k = d^k$, the angle between the vectors $F(y^k)$ and $-\mu_k d^k$ is also θ_k. Since $y^k = x^k + d^k$, by (2.12) we have

$$F(y^k) = -\mu_k d^k - e^k + R^k.$$

Given the above relation, sin of the angle between $F(y^k)$ and $-\mu_k d^k$ can be easily bounded:

$$\sin\theta_k \le \frac{\|-e^k + R^k\|}{\mu_k\|d^k\|}.$$

By the triangle inequality, we further obtain

$$\sin\theta_k \le \rho_k + C_7\frac{\|d^k\|}{\mu_k}.$$

Therefore,

$$\|x^{k+1} - y^k\| \le (\rho_k + C_7\frac{\|d^k\|}{\mu_k})\|d^k\|. \tag{2.14}$$

For the second term in (2.13) we have

$$\begin{aligned}
\|y^k - z^k\| &= \|d^k - s^k\| \\
&= \|(\nabla F(x^k) + \mu_k I)^{-1} e^k\| \\
&\le C_8\|e^k\| \\
&\le C_8\rho_k\mu_k\|d^k\|,
\end{aligned} \tag{2.15}$$

where C_8 is some positive constant.

We now consider the third term in the right-hand-side of (2.13). By the Cauchy-Schwarz and triangle inequalities, and Lipschitz continuity of ∇F in the neighbourhood of $\bar{x}$, we have

$$\begin{aligned}
\|z^k - \bar{x}\| &= \|x^k - \bar{x} - (\nabla F(x^k) + \mu_k I)^{-1} F(x^k)\| \\
&= \|(\nabla F(x^k) + \mu_k I)^{-1}[F(\bar{x}) - F(x^k) + (\nabla F(x^k) + \mu_k I)(x^k - \bar{x})]\| \\
&\le C_8\|F(\bar{x}) - F(x^k) - \nabla F(x^k)(\bar{x} - x^k)\| + C_8\mu_k\|x^k - \bar{x}\| \\
&\le C_9\|x^k - \bar{x}\|^2 + C_8\mu_k\|x^k - \bar{x}\|,
\end{aligned} \tag{2.16}$$

with some $C_9 > 0$.

By (2.10) and Lipschitz continuity of F around $\bar{x}$, we have

$$\|d^k\| \le C_6\|F(x^k)\| \le C_{10}\|x^k - \bar{x}\|, \tag{2.17}$$

where $C_{10} > 0$. Combing (2.13)-(2.16) and taking into account (2.17), we obtain

$$\begin{aligned}
&\|x^{k+1} - \bar{x}\| \le \\
&\left(C_{10}(\rho_k + C_7C_6\tfrac{\|F(x^k)\|}{\mu_k} + C_8\rho_k\mu_k) + C_9\|x^k - \bar{x}\| + C_8\mu_k\right)\|x^k - \bar{x}\|.
\end{aligned} \tag{2.18}$$

From (2.9) it follows that the sequence $\{x^k\}$ converges to $\bar{x}$ Q-superlinearly.

If $\mu_k = \gamma_1\|F(x^k)\|^{1/2}$ and $\rho_k = \gamma_2\|F(x^k)\|^{1/2}$ for some positive constants γ_1 and γ_2, then (2.17) and (2.18) yield

$$\|x^{k+1} - \bar{x}\| \le C_{11}\|x^k - \bar{x}\|^{3/2}, \quad C_{11} > 0,$$

i.e. the order of superlinear convergence is at least 1.5 . ∎

Remark. Inequality (2.18) also shows that if parameters μ_k and ρ_k are fixed and sufficiently small, then the linear rate of convergence is attained.

Remark. It can be observed that under the hypotheses of Theorem 2.2, for k large enough (so that the unit stepsize is always accepted), the steps of Algorithm 2.1 become a particular (practical) implementation of the more general Algorithm 1.1.

Remark. The superlinear convergence of order 1.5 implies at least 2-step Q-quadratic convergence.

Finally, we give one possible choice of parameters which satisfies the conditions of Theorems 2.1 and 2.2 :

$$\mu_k = \max\{C_3\|F(x^k)\|, \gamma_1\|F(x^k)\|^{1/2}\}$$

and

$$\rho_k = \min\{1/2, \gamma_2\|F(x^k)\|^{1/2}\}.$$

3 CONCLUDING REMARKS

A hybrid algorithm for solving systems of monotone equations was presented. The algorithm combines elements of Newton, proximal point and projection methods. An important property of the presented method is that it is truly globally convergent to a solution without any regularity assumptions. Under the assumption of differentiability and nonsingularity at the solution, locally superlinear rate of convergence was established.

References

[1] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16:1–3, 1966.

[2] D.P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmont, Massachusetts, 1995.

[3] J.V. Burke and Maijian Qian. The variable metric proximal point algorithm, I: Basic convergence theory, 1996. Department of Mathematics, University of Washington, Seattle, WA.

[4] R.S. Dembo, S.C Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM Journal of Numerical Analysis*, 19:400–408, 1982.

[5] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, N.J., 1983.

[6] J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[7] J. Eckstein and M.C. Ferris. Smooth methods of multipliers for complementarity problems. Technical Report RRR 27-96, Rutgers Center for Operations Research, Rutgers University, New Brunswick, New Jersey, August 1996. Revised February 1997.

[8] M.C. Ferris. Finite termination of the proximal point algorithm. *Mathematical Programming*, 50:359–366, 1991.

[9] A.N. Iusem and M.V. Solodov. Newton-type methods with generalized distances for constrained optimization. *Optimization*, 41:257–278, 1997.

[10] H. Jiang, L. Qi, X. Chen, and D. Sun. Semismoothness and superlinear copnvergence in nonsmooth optimization and nonsmooth equations. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 197–212. Plenum Press, 1996.

[11] H. Jiang and D. Ralph. Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares. Department of Mathematics, The University of Melbourne, Australia. July 1997.

[12] B. Lemaire. The proximal algorithm. In J.P. Penot, editor, *International Series of Numerical Mathematics*, pages 73–87. Birkhauser, Basel, 1989.

[13] F.J. Luque. Asymptotic convergence analysis of the proximal point algorithm. *SIAM Journal on Control and Optimization*, 22:277–293, 1984.

[14] J.M. Martínez. Local convergence theory for inexact Newton methods based on structural least-squares updates. *Mathematics of Computation*, 55:143–168, 1990.

[15] J.M. Martínez and L. Qi. Inexact Newton methods for solving nonsmoooth equations. *Journal of Computational and Applied Mathematics*, 60:127–145, 1995.

[16] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, 1970.

[17] J.-S. Pang and S.A. Gabriel. An inexact NE/SQP method for solving the nonlinear complementarity problem. *Computational Optimization and Applications*, 1:67–92, 1992.

[18] J.-S. Pang and L. Qi. Nonsmooth equations : Motivation and algorithms. *SIAM Journal on Optimization*, 3:443–465, 1995.

[19] B.T. Polyak. *Introduction to Optimization*. Optimization Software, Inc., Publications Division, New York, 1987.

[20] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18:227–244, 1993.

[21] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–367, 1993.

[22] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[23] M.V. Solodov and B.F. Svaiter. A new projection method for variational inequality problems. Technical Report B-109, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460, Brazil, November 1996. *SIAM Journal on Control and Optimization*, submitted.

[24] M.V. Solodov and B.F. Svaiter. Forcing strong convergence of proximal point iterations in a Hilbert space, 1997. *Mathematical Programming*, submitted.

[25] M.V. Solodov and B.F. Svaiter. A hybrid projection – proximal point algorithm. Technical Report B-115, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460, Brazil, January 1997. *Journal of Convex Analysis*, submitted.

[26] T.J. Ypma. Local convergence of inexact Newton methods. *SIAM Journal of Numerical Analysis*, 21:583–590, 1984.