J Glob Optim (2009) 44:283–296
DOI 10.1007/s10898-008-9324-8

A superlinearly convergent projection method for constrained systems of nonlinear equations

Chuanwei Wang · Yiju Wang

Received: 18 December 2007 / Accepted: 15 June 2008 / Published online: 3 July 2008

Abstract In this paper, a new projection method for solving a system of nonlinear equations with convex constraints is presented. Compared with the existing projection method for solving the problem, the projection region in this new algorithm is modified which makes an optimal stepsize available at each iteration and hence guarantees that the next iterate is more closer to the solution set. Under mild conditions, we show that the method is globally convergent, and if an error bound assumption holds in addition, it is shown to be superlinearly convergent. Preliminary numerical experiments also show that this method is more efficient and promising than the existing projection method.

Keywords Projection method · Constrained system of nonlinear equations · Superlinear convergence

Mathematics Subject Classifications (2000) 90C30 · 15A06

1 Introduction

Let F: $R^n \to R^n$ be a continuous nonlinear mapping and C be a nonempty closed convex set of R^n. Consider the problem of finding $x \in C$ such that

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

This work was done when Yiju Wang visited Chongqing Normal University.

C. Wang
College of Information Sciences and Engineering, Shandong Agricultural University, Tai'an 271018, Shandong, China
e-mail: tw516@163.com

Y. Wang (✉)
School of Operations Research and Management Science, Qufu Normal University, Rizhao 276800, Shandong, China
e-mail: wyiju@hotmail.com

This problem finds applications in many fields such as power engineering [3,12], chemical equilibrium systems and economic equilibrium problems [2,6].

For problem (1.1), a very popular method may be the Levenberg-Marquardt type methods [4,9,15] whose superlinear convergence rate can be established under an error bound estimation instead of the nonsingularity assumption. Different from the methods above, Maranas and Floudas proposed a new type of solution method for a certain class of the problem by introducing slack variables [5].

More recently, Wang et al. [11] established a projection type method for solving problem (1.1) motivated by the fact that the projection method has made a good success in solving such as linearly constrained optimization problems [1], variational inequalities [13] and nonlinear complementarity problems [8]. The numerical performances given in [11] show that the projection method for solving problem (1.1) is really efficient and has strong stability. To accelerate the convergence rate, in this paper, we would propose a modified version for the method inspired by Solodov and Svaiter's work for solving variational inequalities in [7]. The main difference between these two algorithms lies in that the projection region is modified, or more precisely, contracted, in the new version. Our theoretical analysis shows that this modification makes that an optimal step-size could be taken at each iteration and therefore guarantees that the next iterate is more closer to the solution set. Under the same conditions as those in [11], we establish the global convergence and the superlinear convergence of the proposed algorithm. Preliminary numerical experiments also show that this method is more efficient and promising than the projection method in [11].

The remaining part of this paper is distributed as follows. In Sect. 2, we will summarize some basic concepts and related properties which will be used in subsequent sections. The description of the modified projection method and the global convergence will be given in Sect. 3. The superlinear convergence of the method will be established in Sect. 4 and the last section will present some numerical experiments.

2 Preliminaries

A mapping $F\colon R^n \to R^n$ is said to be monotone if

$$\langle F(x) - F(y), x - y\rangle \geq 0 \quad \forall\, x, y \in R^n.$$

For a monotone mapping F, if $\langle F(x) - F(y), x - y\rangle = 0$ iff $x = y$, then it is said to be strictly monotone.

Let Ω be a nonempty closed convex subset of R^n. Then for any $x \in R^n$, its projection to Ω is defined as:

$$P_\Omega[x] = \arg\min\{\|y - x\| \mid y \in \Omega\}.$$

The mapping $P_\Omega\colon R^n \to \Omega$ is called a projection operator.

One well known property of the projection operator is that it is nonexpansive, i.e., for any $x, y \in R^n$, it holds that

$$\|P_\Omega[x] - P_\Omega[y]\| \leq \|x - y\|,$$

or more precisely,

$$\|P_\Omega[x] - P_\Omega[y]\|^2 \leq \|x - y\|^2 - \|P_\Omega[x] - x + y - P_\Omega[y]\|^2. \tag{2.1}$$

We also have the following properties on the projection operator (see [14,16]).

Lemma 2.1 *Let $\Omega \subseteq R^n$ be a nonempty closed convex subset. Then for all $x \in R^n$ and $y \in \Omega$,*

$$\langle P_\Omega[x] - x, y - P_\Omega[x]\rangle \geq 0.$$

Lemma 2.2 *Let $\Omega \subseteq R^n$ be a nonempty closed convex subset. For any $x, d \in R^n$ and $\alpha \geq 0$, define $x(\alpha) := P_\Omega[x - \alpha d]$. Then $\langle d, x(\alpha) - x\rangle$ is non-increasing with respect to $\alpha \geq 0$.*

Lemma 2.3 *Let $\Omega \subseteq R^n$ be a nonempty closed convex subset. For any $x \in \Omega$, $d \in R^n$ and $\alpha \geq 0$, define $\Psi(\alpha) := \min\{\|y - x + \alpha d\|^2 \mid y \in \Omega\}$. Then $\Psi'(\alpha) = 2\langle d, x(\alpha) - x + \alpha d\rangle$.*

3 Algorithm and convergence analysis

Now, we give a description of the modified projection method and then present its global convergence analysis.

Algorithm 3.1

Step 0. Choose an arbitrary initial point $x^0 \in C$, parameters $\gamma_1, \gamma_2 > 0$, $\lambda, \beta \in (0, 1)$, $\kappa_0 \in [0, 1)$, and set $k := 0$.
Step 1. If $F(x^k) = 0$, stop. Otherwise, let $\mu_k = \gamma_1\|F(x^k)\|^{1/2}$, $\sigma_k = \min\{\kappa_0, \gamma_2\|F(x^k)\|^{1/2}\}$. Take a positive semi-definite matrix $G_k \in R^{n\times n}$ and solve the following linear equations with respect to $x \in R^n$

$$F(x^k) + (G_k + \mu_k I)(x - x^k) = 0 \tag{3.1}$$

approximately, i.e., find an approximate solution $\bar{x}^k \in R^n$ to (3.1) such that the residual r^k on the left-hand-side satisfies

$$\|r^k\| \leq \sigma_k\mu_k\|x^k - \bar{x}^k\|. \tag{3.2}$$

Step 2. Find $y^k = x^k + t_k(\bar{x}^k - x^k)$ satisfying that

$$\langle F(y^k), x^k - \bar{x}^k\rangle \geq \lambda(1 - \sigma_k)\mu_k\|x^k - \bar{x}^k\|^2, \tag{3.3}$$

where $t_k = \beta^{m_k}$ and m_k is the smallest nonnegative integer such that (3.3) holds.
Step 3. Compute x^{k+1} via

$$x^{k+1} = P_{C\cap H_k}[x^k - \alpha_k^1 F(y^k)], \tag{3.4}$$

where

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

and

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

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

Remark 3.1

(1) In the algorithm, a projection from R^n onto the intersected set $C \cap H_k$ needs to be computed, i.e., procedure (3.4), at each iteration. Surely, if the domain set C has a special structure such as a box or a ball, then the next iterate x_{k+1} can easily be computed. If the

domain set C is defined by a set of linear (in)equalities, then computing the projection is equivalent to solving a strictly convex quadratic optimization problem. Furthermore, the hyperplane H_k can be replaced by the following half space just as done in [7]:

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

However, this would increase a little more cost in computation for the case that the domain set C is defined by a set of linear equalities.

(2) It can readily be verified that the hyperplane H_k strictly separates the current point x^k from the solution set, denoted by S, if x^k is not a solution of the problem. That is, $S \subset H_k^-$.

(3) Compared with the projection given in [11], besides the major modification made in the projection procedure in the last step, the values of some parameters involved in the algorithm are also adjusted.

Before establishing the global convergence of Algorithm 3.1, we first give its theoretical analysis and its theoretical comparison to the algorithm given in [11].

For Algorithm 3.1, if it terminates within finite steps, then we can obtain a solution of (1.1). So, in the following analysis, we assume that Algorithm 3.1 always generates an infinite sequence.

For $k \ge 0$ and $\alpha \ge 0$, we introduce the following projection point

$$x^k(\alpha) := P_C[x^k - \alpha F(y^k)].$$

Then, for any solution point $x^* \in S$, by (2.1) and the monotonicity of F, we have

$$\begin{aligned}\|x^k(\alpha) - x^*\|^2 &= \|P_C[x^k - \alpha F(y^k)] - x^*\|^2\\ &\le \|x^k - x^* - \alpha F(y^k)\|^2 - \|x^k - x^k(\alpha) - \alpha F(y^k)\|^2\\ &\le \|x^k - x^*\|^2 - 2\alpha\langle F(y^k), x^k - y^k\rangle + \alpha^2\|F(y^k)\|^2\\ &\quad -\|x^k - x^k(\alpha) - \alpha F(y^k)\|^2\end{aligned}$$

If we define

$$\phi_k(\alpha) := 2\alpha\langle F(y^k), x^k - y^k\rangle + \|x^k - x^k(\alpha) - \alpha F(y^k)\|^2 - \alpha^2\|F(y^k)\|^2,$$

then

$$\|x^k(\alpha) - x^*\|^2 \le \|x^k - x^*\|^2 - \phi_k(\alpha).$$

This means that if we want to make the candidate iterate $x(\alpha)$ more closer to the solution set, we can take the maximizer of function $\phi_k(\alpha)$ as the step size.

For the function $\phi_k(\alpha)$, by Lemma 2.3, one has

$$\begin{aligned}\phi_k'(\alpha) &= 2\langle F(y^k), x^k - y^k\rangle + 2\langle F(y^k), x^k(\alpha) - x^k + \alpha F(y^k)\rangle - 2\alpha\|F(y^k)\|^2\\ &= 2\langle F(y^k), x^k - y^k\rangle + 2\langle F(y^k), x^k(\alpha) - x^k\rangle\\ &= 2\langle F(y^k), x^k(\alpha) - y^k\rangle.\end{aligned}$$

Thus, by the linear search procedure in Step 2, we have

$$\begin{aligned}\phi_k'(0) &= 2\langle F(y^k), x^k - y^k\rangle\\ &= 2t_k\langle F(y^k), x^k - \bar{x}^k\rangle\\ &\ge 2t_k\lambda(1-\sigma_k)\mu_k\|x^k - \bar{x}^k\|^2\\ &\ge 2\lambda(1-\sigma_k)\mu_k\|x^k - y^k\|^2 > 0,\end{aligned} \tag{3.5}$$

where the last non-strict inequality uses the fact that $t_k \le 1$.

Consider the optimization problem

$$\max\{\phi_k(\alpha) \mid \alpha \ge 0\} \tag{3.6}$$

Since $\phi_k(0) = 0$ and $\phi_k'(0) > 0$ by (3.5), we know that

$$\max\{\phi_k(\alpha) \mid \alpha \geq 0\} > 0.$$

By Lemma 2.2, we know that function $\phi_k'(\alpha)$ is nonincreasing and continuous with respect to $\alpha \geq 0$. So, if the equation

$$\phi_k'(\alpha) = 0$$

is solvable on $\alpha \geq 0$ then its any solution coincides with the maximizer of problem (3.6).

The following conclusion tells us that the maximizer to problem (3.6) really exists.

Lemma 3.1 *Suppose that the underlying mapping F is monotone. Then the equation $\phi_k'(\alpha) = 0$ is solvable with respect to $\alpha > 0$.*

Proof It is easy to verify that $(x^k - \alpha_k^1 F(y^k))$ is the projection of x^k to hyperplane H_k. Thus, for any $\alpha > \alpha_k^1$, it holds that

$$x^k - \alpha F(y^k) \in \{x \in R^n \mid \langle F(y^k), x - y^k\rangle < 0\}.$$

From (2) in Remark 3.1, we know that the set $C \cap H_k^-$ is nonempty. Hence, it can be readily shown that there exists α_k' satisfying $\alpha_k' \geq \alpha_k^1 > 0$ such that

$$P_C[x^k - \alpha_k' F(y^k)] \in \{x \in R^n \mid \langle F(y^k), x - y^k\rangle \leq 0\}.$$

From (3.5), we conclude that

$$P_C[x^k - 0 \cdot F(y^k)] \in \{x \in R^n \mid \langle F(y^k), x - y^k\rangle > 0\}.$$

Thus, by the continuity of the projection, there exists $\alpha_k^2 \in (0, \alpha_k')$ such that

$$x^k(\alpha_k^2) = P_C[x^k - \alpha_k^2 F(y^k)] \in H_k \cap C,$$

which implies that $\phi_k'(\alpha_k^2) = 0$. This completes the proof. □

For the candidate iterate $x^k(\alpha)$, if we denote the smallest positive solution of equation $\phi_k'(\alpha) = 0$ by α_k^2, then it can be taken as an optimal stepsize for $x(\alpha)$. Compared with stepsize α_k^1 taken in Algorithm 3.1 in [11], α_k^2 is longer and iterate $x(\alpha_k^2)$ is more closer to the solution set than iterate $x(\alpha_k^1)$ in theory. In this sense, α_k^2 is an optimal stepsize. However, it is difficult to take unless the domain set C has a special structure. The following proposition shows this can be realized via the projection made in the last step of Algorithm 3.1, which also means that the algorithm given in [11] and Algorithm 3.1 can be unified via $x(\alpha)$.

Proposition 3.1 [10] *Suppose that the underlying mapping F is monotone. Then*

$$P_C[x^k - \alpha_k^2 F(y^k)] = P_{C\cap H_k}[x^k - \alpha_k^1 F(y^k)].$$

Now, we turn to establishing the global convergence of Algorithm 3.1. To this end, the following assumptions are needed, which are the same as those given in [11] for the convergence of the algorithm in that paper.

Assumption:

(A1) The solution set S of (1.1) is nonempty;
(A2) The underlying mapping F is monotone;
(A3) For the positive semi-definite matrix sequence $\{G_k\}$, it holds that $\sup_k \|G_k\| < \infty$.

Theorem 3.1 *Suppose that Assumptions (A1)–(A3) hold. Then the sequence* $\{x^k\}$ *generated by Algorithm 3.1 converges to a solution of (1.1).*

Proof First we show the sequences $\{x^k\}$ and $\{y^k\}$ are both bounded.

From the above analysis, for any $k \geq 0$ and any $x^* \in S$, it holds that

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|P_C[x^k - \alpha_k^2 F(y^k)] - x^*\|^2 \\
&\leq \|x^k - x^*\|^2 - \phi(\alpha_k^2) \\
&\leq \|x^k - x^*\|^2 - \phi(\alpha_k^1) \\
&= \|x^k - x^*\|^2 - 2\alpha_k^1 \langle F(y^k), x^k - y^k\rangle + (\alpha_k^1)^2 \|F(y^k)\|^2 \\
&\quad - \|x^k - x^k(\alpha_k^1) - \alpha_k^1 F(y^k)\|^2 \\
&\leq \|x^k - x^*\|^2 - 2\alpha_k^1 \langle F(y^k), x^k - y^k\rangle + (\alpha_k^1)^2 \|F(y^k)\|^2.
\end{aligned}$$

Thus by the definition of α_k^1, we have

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \frac{\langle F(y^k), x^k - y^k\rangle^2}{\|F(y^k)\|^2}, \tag{3.7}$$

which means that sequence $\{\|x^k - x^*\|\}$ is contractive and thus sequence $\{x^k\}$ is bounded.

Due to the monotonicity of F, (3.5) yields

$$\langle F(x^k), x^k - y^k\rangle \geq \lambda(1 - \sigma_k)\mu_k \|x^k - y^k\|^2.$$

By the Cauchy–Schwartz inequality and the choices of μ_k and σ_k, the inequality above reads

$$\|F(x^k)\|^{1/2} \geq \lambda(1 - \kappa_0)\gamma_1 \|x^k - y^k\|.$$

From the boundedness of $\{x^k\}$ and the continuity of F, we know that sequence $\{y^k\}$ is also bounded.

Now, we can show the global convergence of the sequence $\{x^k\}$.

Since F is continuous and sequence $\{y^k\}$ is bounded, there exists a positive constant M such that $\|F(y^k)\| \leq M$ for $k \geq 0$. It follows from (3.7) that

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \frac{\langle F(y^k), x^k - y^k\rangle^2}{M^2},$$

from which we deduce that

$$\lim_{k\to\infty} \langle F(y^k), x^k - y^k\rangle = 0. \tag{3.8}$$

On the other hand, by (3.5) and the choices of σ_k and λ, there exists a constant $\kappa > 0$ such that

$$\langle F(y^k), x^k - y^k\rangle \geq \kappa t_k \mu_k \|x^k - \bar{x}^k\|^2.$$

This, together with (3.8), yields that

$$\lim_{k\to\infty} t_k \mu_k \|x^k - \bar{x}^k\| = 0.$$

Following the latter part of the proof of Theorem 2.1 in [11], we can obtain the desired result. □

4 Convergence rate

Throughout this section, we assume that $x^k \to x^*$ as $k \to \infty$, where $x^* \in S$. To analyze the convergence rate of Algorithm 3.1, the following additional assumption is needed.

Assumption:

(A4) For $x^* \in S$ and sufficiently large k, there exist positive constants δ, c_1 and c_2 such that

$$c_1 \text{dist}(x, S) \le \|F(x)\|, \quad \forall\, x \in N(x^*, \delta), \tag{4.1}$$

and

$$\|F(x) - F(y) - G_k(x-y)\| \le c_2\|x-y\|^2, \quad \forall\, x, y \in N(x^*, \delta), \tag{4.2}$$

where $\text{dist}(x, S)$ denotes the distance from x to solution set S, and

$$N(x^*, \delta) := \{x \in R^n \mid \|x - x^*\| \le \delta\}.$$

Under Assumption (A4), it is readily shown that there exists $L > 0$ such that

$$\|F(x) - F(y)\| \le L\|x-y\|, \quad \forall\, x, y \in N(x^*, \delta), \tag{4.3}$$

which means that F is locally Lipschitz continuous. Also, if F is continuously differentiable and F' is locally Lipschitz continuous, then (4.2) holds with $x = x^k$ and $G_k = F'(x^k)$.

In order to prove the convergence rate of the algorithm, we need several technical lemmas.

Lemma 4.1 [17] *Let $G \in R^{n\times n}$ be a positive semi-definite matrix and $\mu > 0$. Then*

(1) $\|(G+\mu I)^{-1}\| \le \frac{1}{\mu}$;

(2) $\|(G+\mu I)^{-1}G\| \le 2$.

Lemma 4.2 *Suppose that Assumption (A4) holds. For points x^k and $\bar{x}^k$ generated by Algorithm 3.1, if $x^k \in N(x^*, \frac{1}{2}\delta)$, then there exists a positive constant c_3 such that*

$$\|x^k - \bar{x}^k\| \le c_3 dist(x^k, S).$$

Proof Let $\hat{x}^k \in S$ be the closest solution to x^k. Then from $x^k \in N(x^*, \frac{1}{2}\delta)$, we conclude that

$$\|\hat{x}^k - x^*\| \le \|\hat{x}^k - x^k\| + \|x^k - x^*\| \le \delta,$$

which means that $\hat{x}^k \in N(x^*, \delta)$.

Thus, by (3.1), (3.2), (4.2), Lemma 4.1 and Assumption (A4), we have

$$\begin{aligned}\|x^k - \bar{x}^k\| &\le \|(G_k + \mu_k I)^{-1}F(x^k)\| + \|(G_k+\mu_k I)^{-1}r^k\| \\ &\le \|(G_k+\mu_k I)^{-1}[F(\hat{x}^k) - F(x^k) - G_k(\hat{x}^k - x^k)]\| \\ &\quad + \|(G_k+\mu_k I)^{-1}G_k(x^k - \hat{x}^k)\| + \tfrac{1}{\mu_k}\|r^k\| \\ &\le \tfrac{c_2}{\mu_k}\|x^k - \hat{x}^k\|^2 + 2\|x^k - \hat{x}^k\| + \sigma_k\|x^k - \bar{x}^k\|,\end{aligned}$$

Hence,

$$(1-\sigma_k)\|x^k - \bar{x}^k\| \le \left(\frac{c_2}{\mu_k}\|x^k - \hat{x}^k\| + 2\right)\text{dist}(x^k, S). \tag{4.4}$$

From (4.1) and the choice of μ_k, it holds that

$$\frac{c_2}{\mu_k}\|x^k - \hat{x}^k\| = \frac{c_2}{\mu_k}\text{dist}(x^k, S) \le \frac{c_2 M_1}{c_1\gamma_1}, \tag{4.5}$$

where M_1 is the upper bound of $\{\|F(x^k)\|^{1/2}\}$.

Set $c_3 := \frac{c_2 M_1 + 2c_1\gamma_1}{c_1\gamma_1(1-\kappa_0)}$. Then by (4.4) and (4.5), one has

$$\|x^k - \bar{x}^k\| \leq c_3 \mathrm{dist}(x^k, S).$$

This completes the proof. □

Lemma 4.3 *Suppose that Assumptions (A1)–(A4) hold. Then for all k sufficiently large,*

(1) $c_4\|x^k - \bar{x}^k\| \leq \|F(x^k)\| \leq c_5\|x^k - \bar{x}^k\|$;
(2) $\|F(x^k) - G_k(x^k - \bar{x}^k)\| \leq c_6\|x^k - \bar{x}^k\|^{3/2}$;

where c_4, c_5, c_6 are all positive constants.

Proof For (1), the left-hand-side of the inequality follows directly from Lemma 4.2 and (4.1) by setting $c_4 := c_1/c_3$.

For the right-hand-side part, from (3.1), (3.2) and the triangle inequality, we have

$$\begin{aligned}\|F(x^k)\| &\leq \|(G_k + \mu_k I)(\bar{x}^k - x^k)\| + \|r^k\| \\ &\leq \|(G_k + \mu_k I)\|\|(\bar{x}^k - x^k)\| + \sigma_k\mu_k\|\bar{x}^k - x^k\| \\ &\leq c_5\|\bar{x}^k - x^k\|,\end{aligned}$$

where the last inequality follows from Assumption (A3) and the choices of σ_k, μ_k.

For (2), from (3.1), (3.2) and the triangle inequality, we have

$$\begin{aligned}\|F(x^k) - G_k(x^k - \bar{x}^k)\| &\leq \mu_k\|x^k - \bar{x}^k\| + \|r^k\| \\ &\leq (1+\sigma_k)\mu_k\|x^k - \bar{x}^k\| \\ &\leq (1+\kappa_0)\gamma_1\|F(x^k)\|^{1/2}\|x^k - \bar{x}^k\|.\end{aligned}$$

Using the right-hand-side inequality of (1) and setting $c_6 := (1+\kappa_0)\gamma_1 c_5^{1/2}$ yield

$$\|F(x^k) - G_k(x^k - \bar{x}^k)\| \leq c_6\|x^k - \bar{x}^k\|^{3/2}.$$

This completes the proof. □

Lemma 4.4 *Suppose that Assumptions (A1)–(A4) hold. Then for all k sufficiently large, it holds that $t_k = 1$. That is, $y^k = \bar{x}^k$.*

Proof From the fact that $x^k \to x^*$ as $k \to \infty$, we know that $\{\|F(x^k)\|\}$ converges to 0 as $k \to \infty$. From Lemma 4.3, we know that the sequence $\{\|x^k - \bar{x}^k\|\}$ tends to 0 as $k \to \infty$. So, $\bar{x}^k \in N(x^*, \delta)$ for k sufficiently large. Hence it follows from (4.2) that

$$F(\bar{x}^k) = F(x^k) + G_k(\bar{x}^k - x^k) + R^k$$

with $\|R^k\| \leq c_2\|x^k - \bar{x}^k\|^2$. From (3.1), the above equality can be written as

$$F(\bar{x}^k) = \mu_k(x^k - \bar{x}^k) + r^k + R^k. \tag{4.6}$$

Hence,

$$\begin{aligned}\langle F(\bar{x}^k), x^k - \bar{x}^k\rangle &= \langle \mu_k(x^k - \bar{x}^k) + r^k + R^k, x^k - \bar{x}^k\rangle \\ &\geq \mu_k\|x^k - \bar{x}^k\|^2 - \sigma_k\mu_k\|x^k - \bar{x}^k\|^2 - c_2(\|x^k - \bar{x}^k\|^3) \\ &\geq \left(1 - \frac{c_2(\|x^k - \bar{x}^k\|)}{\mu_k(1-\sigma_k)}\right)(1-\sigma_k)\mu_k\|x^k - \bar{x}^k\|^2.\end{aligned} \tag{4.7}$$

By the first inequality in Lemma 4.3 and the choices of μ_k, σ_k, we know that for k sufficiently large,

$$1 \geq 1 - \frac{c_2\|x^k - \bar{x}^k\|}{\mu_k(1-\sigma_k)} \geq 1 - \frac{c_2\|F(x^k)\|^{1/2}}{c_4\gamma_1(1-\kappa_0)}.$$

Since $\|f(x^k)\| \to 0$ as $k \to \infty$, then for k sufficiently large, we have

$$1 - \frac{c_2\|x^k - \bar{x}^k\|}{\mu_k(1-\sigma_k)} \geq \lambda,$$

which, in junction with inequality (4.7), implies that (3.3) holds with $t_k = 1$ for all k sufficiently large. This completes the proof. □

This conclusion tells us that we can assume that $y^k = \bar{x}^k$ for sufficiently large k in the subsequent analysis.

Lemma 4.5 *Suppose that Assumptions (A1)–(A4) hold. Set $\tilde{x}^k := x^k - \alpha_k^1 F(y^k)$. Then for all k sufficiently large, there exists a positive constant c_7 such that*

$$\|\tilde{x}^k - y^k\| \leq c_7\|x^k - \bar{x}^k\|^{3/2}.$$

Proof Since $\tilde{x}^k$ is the orthogonal projection of x^k onto H_k, then we have

$$\|\tilde{x}^k - y^k\| = \|y^k - x^k\| \sin\theta_k = \|x^k - \bar{x}^k\| \sin\theta_k, \tag{4.8}$$

where θ_k is the angle composed by vectors $(\tilde{x}^k - x^k)$ and $(y^k - x^k)$.

Since $\tilde{x}^k - x^k = -\alpha_k^1 F(y^k)$ and $y^k - x^k = \bar{x}^k - x^k$, so the angle composed by $\alpha_k^1 F(y^k)$ and $\mu_k(x^k - \bar{x}^k)$ is also θ_k. Now, we will give a bound estimation to the sine function.

From (4.6), we know that the three edges $F(y^k)$, $\mu_k(x^k - \bar{x}^k)$ and $(r_k + R_k)$ constitute a triangle. From the geometry knowledge of sine function, we conclude that

$$\begin{aligned}
\sin\theta_k &\leq \frac{\|r^k + R^k\|}{\mu_k\|x^k - \bar{x}^k\|} \\
&\leq \frac{\sigma_k\mu_k\|x^k - \bar{x}^k\| + c_2\|x^k - \bar{x}^k\|^2}{\mu_k\|x^k - \bar{x}^k\|} \\
&= \sigma_k + \frac{c_2\|x^k - \bar{x}^k\|}{\mu_k} \\
&\leq \gamma_2\|F(x^k)\|^{1/2} + \frac{c_2\|F(x^k)\|}{c_4\gamma_1\|F(x^k)\|^{1/2}} \\
&= \eta\|F(x^k)\|^{1/2},
\end{aligned}$$

where $\eta := \gamma_2 + \frac{c_2}{c_4\gamma_1}$.

Therefore, by the right-hand-side inequality (1) in Lemma 4.3, we can deduce from (4.8) that

$$\begin{aligned}
\|\tilde{x}^k - y^k\| &\leq \eta\|F(x^k)\|^{1/2}\|x^k - \bar{x}^k\| \\
&\leq c_7\|x^k - \bar{x}^k\|^{3/2},
\end{aligned}$$

where $c_7 := \eta c_5^{\frac{1}{2}}$. This completes the proof. □

Lemma 4.6 *Suppose that Assumptions (A1)–(A4) hold. Then for $x^* \in S$ and all $k \geq 0$,*

$$\|x^{k+1} - x^*\| \leq \|\tilde{x}^k - x^*\|.$$

Proof If $\tilde{x}^k \in C$, then $x^{k+1} = \tilde{x}^k$ and the assertion follows immediately. So, we can only consider the case that $\tilde{x}^k \notin C$.

From Proposition 3.1,

$$x^{k+1} = P_C[x^k - \alpha_k^2 F(y^k)].$$

Then for the hyperplane

$$U_k := \{x \in R^n \mid \langle x^k - \alpha_k^2 F(y^k) - x^{k+1}, x - x^{k+1}\rangle = 0\}.$$

From Lemma 2.1, we know that

$$C \subset U_k^- := \{x \in R^n \mid \langle x^k - \alpha_k^2 F(y^k) - x^{k+1}, x - x^{k+1}\rangle \leq 0\}.$$

Combining this with the fact that $S \subset H_k^- \cap C$ yields that $S \subset U_k^- \cap H_k^- \cap C$.

Define hyperplane

$$V_k := \{x \in R^n \mid \langle \tilde{x}^k - x^{k+1}, x - x^{k+1}\rangle = 0\}.$$

For any $x \in R^n$ such that

$$\langle F(y^k), x - x^{k+1}\rangle \leq 0, \quad \langle x^k - \alpha_k^2 F(y^k) - x^{k+1}, x - x^{k+1}\rangle \leq 0,$$

using inequality $\alpha_k^2 \geq \alpha_k^1$, we can readily deduce that

$$\langle x^k - \alpha_k^1 F(y^k) - x^{k+1}, x - x^{k+1}\rangle \leq 0,$$

i.e.,

$$\langle \tilde{x}^k - x^{k+1}, x - x^{k+1}\rangle \leq 0.$$

This implies that

$$U_k^- \cap H_k^- \cap C \subset V_k^- \cap H_k^- \cap C,$$

where V_k^- is a half space defined similarly to H_k^- and U_k^-.

Again from $S \subset U_k^- \cap H_k^- \cap C$, we conclude that $S \subset V_k^- \cap H_k^- \cap C$.

Since hyperplane H_k and hyperplane V_k are perpendicular, $\tilde{x}^k \in H_k \cap V_k^+$, $x^{k+1} \in H_k \cap V_k$ and $S \subset V_k^- \cap H_k^-$. Thus for any $x^* \in S$, we denote its projection to the line determined by $\tilde{x}^k$ and x^{k+1} by $\bar{x}^*$, then x^{k+1} lies within the segment $[\tilde{x}^k, \bar{x}^*]$ and three points $\tilde{x}^k, \bar{x}^*, x^*$ constitute a right triangle. By the triangle geometry property, we obtain the desired result and this completes the proof. □

Now, we are at a position to state the main result in this section.

Theorem 4.1 *Suppose that Assumptions (A1)–(A4) hold. Then the sequence $\{dist(x^k, S)\}$ Q-superlinearly converges to 0.*

Proof From Lemma 4.3 and (4.3), for sufficiently large k, it holds that

$$\begin{aligned}\|\tilde{x}^k - x^*\| &\leq \|x^k - x^*\| + \|\alpha_k^1 F(y^k)\| \\ &\leq \|x^k - x^*\| + \|x^k - \bar{x}^k\| \\ &\leq \|x^k - x^*\| + \|F(x^k)\|/c_4 \\ &\leq (1 + L/c_4)\|x^k - x^*\|,\end{aligned}$$

from which we deduce that $\|\tilde{x}^k - x^*\| \to 0$ as $k \to \infty$. Thus, $\tilde{x}^k \in N(x^*, \delta)$ for k sufficiently large, and from (4.2), we have

$$\begin{aligned}\|F(\tilde{x}^k)\| &\le \|F(x^k) - G_k(\tilde{x}^k - x^k)\| + c_2\|\tilde{x}^k - x^k\|^2\\ &\le \|F(x^k) - G_k(x^k - \bar{x}^k)\| + \|G_k(\tilde{x}^k - \bar{x}^k)\| + c_2\|\tilde{x}^k - x^k\|^2\\ &\le \|F(x^k) - G_k(x^k - \bar{x}^k)\| + \tau_{\max}\|\tilde{x}^k - y^k\| + c_2\|x^k - \bar{x}^k\|^2\\ &\le c_6\|x^k - \bar{x}^k\|^{3/2} + \tau_{\max}c_7\|x^k - \bar{x}^k\|^{3/2} + c_2\|x^k - \bar{x}^k\|^2,\end{aligned}$$

where $\tau_{\max} = \sup_k \|G_k\|$ and the last inequality follows from the second conclusion in Lemma 4.3 and from Lemma 4.5. So, there exists a positive constant c_8 such that

$$\|F(\tilde{x}^k)\| \le c_8\|x^k - \bar{x}^k\|^{3/2}. \tag{4.9}$$

On the other hand, by Lemma 4.6, one has

$$\begin{aligned}\mathrm{dist}(x^{k+1}, S) &= \inf_{s\in S}\|x^{k+1} - s\|\\ &\le \inf_{s\in S}\|\tilde{x}^k - s\|\\ &= \mathrm{dist}(\tilde{x}^k, S).\end{aligned}$$

Hence, from (4.1) we have

$$c_1\mathrm{dist}(x^{k+1}, S) \le \|F(\tilde{x}^k)\|. \tag{4.10}$$

Then from (4.9), (4.10) and Lemma 4.2 we have

$$c_1\mathrm{dist}(x^{k+1}, S) \le \|F(\tilde{x}^k)\| \le c_8\|x^k - \bar{x}^k\|^{3/2} \le c_9\mathrm{dist}^{3/2}(x^k, S),$$

where $c_9 := c_3^{3/2}c_8$. This means that the sequence $\{\mathrm{dist}(x^k, S)\}$ Q-superlinearly converges to 0. □

5 Preliminary numerical experiments

In our numerical experiments, all examples used in this section were tested in [11]. Just as done in [11], we take $G_k = F'(x^k)$ and use the left division operation in MATLAB to solve the system of linear equations (3.1) at each iteration. In this sense, the subproblem is solved with a higher accurate for all k (i.e., $\kappa_0 \approx 0$). Other parameters used in the algorithm are set as $\lambda = 0.95$, $\beta = 0.6$ and $\gamma_1 = 1$. We choose $\|F(x^k)\| \le 10^{-6}$ as the stop criterion. All codes are written in MATLAB 6.5 and run on a PIV 2.0 GHz personal computer.

For convenience, we denote the projection algorithm proposed in [11] by Alg-P and the modified projection algorithm in this paper by Alg-MP. Certainly, the numerical comparison of these two algorithms is the focus of this section. So, the computation of Alg-P will be repeated here.

Example 5.1 The mapping F is taken as $F(x) = (f_1(x), \dots, f_n(x))^T$, where $f_i(x) = e^{x_i} - 1,\ i = 1, \dots, n$ and $C = R^n_+$.

It is obvious that the mapping F is strictly monotone and this problem has a unique solution $x^* = (0, \dots, 0)$. For initial point $x^0 = (1, \dots, 1)$, Table 1 gives the numerical results by Alg-P and Alg-MP with different dimensions, where the unit of running time of CPU is second and Iter. denotes the iteration number when the algorithm terminates.

Table 1 Numerical results of Example 5.1

	Dimension	8	16	32	64	128	256
Alg-P	Iter.	43	45	46	47	48	49
	Running time	0.20	0.26	0.32	0.37	0.75	4.01
Alg-MP	Iter.	4	4	4	4	4	4
	Running time	0.04	0.04	0.04	0.04	0.15	1.11

Example 5.2 Let the domain set C and the mapping F be, respectively, taken as

$$C = \left\{ x \in R^5 \;\middle|\; \sum_{i=1}^{5} x_i \geq 10,\; x_i \geq 0,\; i = 1, 2, \ldots, 5 \right\},$$

and

$$F(x) = \rho D(x) + Mx + q + q_0,$$

where M is a 5×5 asymmetric positive definite matrix whose entries are randomly generated in $(-5, 5)$, the vector q is generated from a uniform distribution in the interval $(-10, 10)$ and $D_i(x) = \arctan(x_i - 2), i = 1, 2, \ldots, 5$. The parameter ρ is a constant. q_0 is a constant vector which guarantees that the equation $F(x) = 0$ has solutions over C.

For this problem, Table 2 gives the numerical results for $\rho = 100$ and $\rho = 200$ by Al-10, respectively, while Table 3 gives the numerical results for $\rho = 100$ and $\rho = 200$ by Alg-MP, respectively.

Table 2 Numerical results of Example 5.2 by Alg-P

	Initial point	Iter.	Running time	$\lVert F(x^*) \rVert$
	(25,0,0,0,0)	14	0.10	1.28×10^{-7}
$\rho = 100$	(10,0,0,0,0)	12	0.09	1.98×10^{-7}
	(10,0,10,0,10)	13	0.10	1.13×10^{-7}
	(0,2.5,2.5,2.5,2.5)	10	0.09	1.77×10^{-7}
	(25,0,0,0,0)	14	0.11	2.81×10^{-8}
$\rho = 200$	(10,0,0,0,0)	12	0.11	2.78×10^{-8}
	(10,0,10,0,10)	13	0.10	2.25×10^{-8}
	(0,2.5,2.5,2.5,2.5)	10	0.09	8.74×10^{-7}

Table 3 Numerical results of Example 5.2 by Alg-MP

	Initial point	Iter.	Running time	$\lVert F(x^*) \rVert$
	(25,0,0,0,0)	11	0.06	4.45×10^{-10}
$\rho = 100$	(10,0,0,0,0)	8	0.04	4.58×10^{-8}
	(10,0,10,0,10)	9	0.06	4.94×10^{-10}
	(0,2.5,2.5,2.5,2.5)	6	0.04	1.15×10^{-7}
	(25,0,0,0,0)	11	0.06	1.04×10^{-9}
$\rho = 200$	(10,0,0,0,0)	8	0.04	3.71×10^{-7}
	(10,0,10,0,10)	8	0.06	8.31×10^{-9}
	(0,2.5,2.5,2.5,2.5)	6	0.04	6.92×10^{-9}

Table 4 Numerical results of Example 5.3 by Alg-P

Dimension	10	20	30	40	50	60	70	80	90	100
Iter.	29	33	38	42	43	47	52	54	58	60
Running time	0.14	0.17	0.18	0.34	0.40	0.51	0.78	1.01	1.32	1.57

Table 5 Numerical results of Example 5.3 by Alg-MP

Dimension	10	20	30	40	50	60	70	80	90	100
Iter.	8	8	8	8	8	8	8	9	9	9
Running time	0.06	0.07	0.12	0.17	0.21	0.28	0.40	0.57	0.68	0.73

Table 6 Numerical results of Example 5.4

	Initial point	Iter.	Running time	$\|F(x^*)\|$
	(0,0,0,0)	16	0.14	6.12×10^{-7}
Alg-P	(3,0,0,0)	16	0.14	6.69×10^{-7}
	(1,1,1,0)	29	0.26	8.52×10^{-7}
	(0,1,1,1)	57	0.51	8.41×10^{-7}
	(0,0,0,0)	13	0.09	1.21×10^{-8}
Alg-MP	(3,0,0,0)	12	0.09	8.43×10^{-9}
	(1,1,1,0)	13	0.10	4.01×10^{-9}
	(0,1,1,1)	15	0.11	9.11×10^{-9}

Example 5.3 Let $C = R^n_+$ and the mapping

$$F(x) = D(x) + Mx + q + q_0,$$

where $D(x)$ and $Mx + q$ are the nonlinear part and the linear part of $F(x)$, respectively. The matrix $M = A^\top A + B$, where A is an $n \times n$ matrix whose entries are randomly generated in the interval $(-1, 1)$ and a skew-symmetric matrix B is generated in the same way. The vector q is generated from a uniform distribution in the interval $(-500, 500)$. In $D(x)$, the nonlinear part of $F(x)$, the components are $D_j(x) = a_j * \arctan(x_j)$ and a_j is a random variable in $(0, 100)$. q_0 is a regular vector.

Tables 4 and 5 report the average results for n from 10 to 100 with the initial points randomly generated in $(0, 1)$ by Alg-P and Alg-MP, respectively.

Example 5.4 Let

$$F(x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} x_1^3 \\ x_2^3 \\ 2x_3^3 \\ 2x_4^3 \end{pmatrix} + \begin{pmatrix} -10 \\ 1 \\ -3 \\ 0 \end{pmatrix},$$

and the constraint set C be taken as

$$C = \left\{ x \in R^4 \;\middle|\; \sum_{i=1}^{4} x_i \le 3,\ x_i \ge 0,\ i = 1, \ldots, 4 \right\}.$$

This problem has the unique but degenerate solution $x^* = (2, 0, 1, 0)^T$. The numerical results of Alg-P and Alg-MP are given in Table 6. Note that the algorithm in [11] requires more iterations for approximating a solution for some starting points than Algorithm 3.1.

Acknowledgements The authors wish to give their sincere thanks to the anonymous referees for their valuable suggestions and helpful comments which improved the presentation of the paper. This work was supported by the Natural Science Foundation of China (Grant No.10771120), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and Chongqing Key Lab. of Operations Research and System Engineering.

References

1. Calamai, P.H., Moré, J.J.: Projected gradient methods for linearly constrained problems. Math. Program. **39**(1), 93–116 (1987)
2. Dirkse, S.P., Ferris, M.C.: MCPLIB: a collection of nonlinear mixed complementarity problems. Optim. Methods Software **5**, 319–345 (1995)
3. El-Hawary, M.E.: Optimal Power Flow: Solution Techniques, Requirement and Challenges. IEEE Service Center, Piscataway, NJ (1996)
4. Kanzow, C., Yamashita, N., Fukushima, M.: Levenberg-Marquardt methods for constrained nonlinear equations with strong local convergence properties. J. Comput. Appl. Math. **172**, 375–397 (2004)
5. Maranas, C.D., Floudas, C.A.: Finding all solutions of nonlinearly constrained systems of equations. J. Glob. Optim. **7**(2), 143–182 (1995)
6. Meintjes, K., Morgan, A.P.: A methodology for solving chemical equilibrium systems. Appl. Math. Comput. **22**, 333–361 (1987)
7. Solodov, M.V., Svaiter, B.F.: A new projection method for variational inequality problems. SIAM J. Control Optim. **37**(3), 765–776 (1999)
8. Solodov, M.V., Svaiter, B.F.: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. SIAM J. Optim. **10**, 605–625 (2000)
9. Tong, X.J., Qi, L.: On the convergence of a trust-region method for solving constrained nonlinear equations with degenerate solution. J. Optim. Theory Appl. **123**, 187–211 (2004)
10. Wang, Y.J., Xiu, N.H., Zhang, J.Z.: Unified framework of extragradient-type methods for pseudomonotone variational inequalities. J. Optim. Theory Appl. **111**, 641–656 (2001)
11. Wang, C.W., Wang, Y.J., Xu, C.L.: A projection method for a system of nonlinear equtions with convex constraints. Math. Methods Oper. Res. **66**, 33–46 (2007)
12. Wood, A.J., Wollenberg, B.F.: Power Generations, Operations, and Control. Wiley, New York (1996)
13. Xiu, N.H., Zhang, J.Z.: Some recent advances in projection-type methods for variational inequalities. J. Comput. Appl. Math. **152**, 559–585 (2003)
14. Xiu, N.H., Wang, C.Y., Zhang, J.Z.: Convergence properties of projection and contraction methods for variational inequality problems. Appl. Math. Optim. **43**, 147–168 (2001)
15. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. Computing (Suppl.) **15**, 237–249 (2001)
16. Zarantonello, E.H.: Projections on convex sets in Hilbert space and spectral theory, contributions to nonlinear functional analysis. In: Zarantonello, E.H. (ed.) Contributions to Nonlinear Functional Analysis. Academic Press, New York, NY (1971)
17. Zhou, G.L., Toh, K.C.: Superlinear convergence of a Newton-type algorithm for monotone equations. J. Optim. Theory Appl. **125**, 205–221 (2005)