CrossMark

Analysis on Newton projection method for the split feasibility problem

Biao Qu[1] · Changyu Wang[1] · Naihua Xiu[2]

Abstract In this paper, based on a merit function of the split feasibility problem (SFP), we present a Newton projection method for solving it and analyze the convergence properties of the method. The merit function is differentiable and convex. But its gradient is a linear composite function of the projection operator, so it is nonsmooth in general. We prove that the sequence of iterates converges globally to a solution of the SFP as long as the regularization parameter matrix in the algorithm is chosen properly. Especially, under some local assumptions which are necessary for the case where the projection operator is nonsmooth, we prove that the sequence of iterates generated by the algorithm superlinearly converges to a regular solution of the SFP. Finally, some numerical results are presented.

Keywords The split feasibility problem · Projection operator · Generalized Jacobian · Newton projection method · Global convergence and convergence rate

Mathematics Subject Classification 90C30 · 90C90 · 65K05

✉ Biao Qu
   qubiao001@163.com

   Changyu Wang
   wcy0537@126.com

   Naihua Xiu
   nhxiu@bjtu.edu.cn

[1] Institute of Operations Research, Qufu Normal University, Rizhao 276826, Shandong, People's Republic of China

[2] Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, People's Republic of China

✌ Springer

1 Introduction

In this paper, we consider the following problem: find $x \in C$ such that

$$Ax \in Q,$$

where, C and Q are nonempty closed convex sets in $\Re^n$ and $\Re^m$, respectively, and A is an m by n real matrix. This problem was called the split feasibility problem (SFP) by Censor and Elfving [1]. It has important and wide applications in signal processing, image reconstruction and so on. Hence, it attracts many researchers attention in recent 20 years.

In [1], the authors used their multidistance idea to obtain iterative algorithms for solving the SFP. Their algorithms, as well as others obtained later involve matrix inverses at each iteration. In [2], Byrne presented a projection method called the CQ algorithm for solving the SFP that does not involve matrix inverses, but they assumed that the orthogonal projections onto C and Q are easily calculated. However, in some cases it is impossible or needs too much work to exactly compute the orthogonal projection. Therefore, if this case appears, the efficiency of projection-type methods, including the CQ algorithm, will be seriously affected. In [3], by using the relaxed projection technology, Yang presented a relaxed CQ algorithm for solving the SFP, where he used two halfspaces C_k and Q_k in place of C and Q, respectively, at the kth iteration and the orthogonal projections onto C_k and Q_k are easily executed. Recently, Qu and Xiu [4] gave a new reformulation for the SFP and proposed a new halfspace-relaxation projection method based on the new reformulation to solve the SFP. For more effective algorithms and the extensions of the SFP, the readers can see [5–8] and the survey papers [9,10].

The global convergence of all the algorithms mentioned above were proved. It is worth noting that most of the algorithms are subject to the gradient projection method [11]. Projection-type methods represent important tools for finding the approximate solution of nonlinear programming and variational inequalities. The main idea of in this technique is to establish the equivalence between the problems related and the fixed-point problem by using the concept of projection. This alternative formulation has played a significant part in developing various projection-type methods for solving nonlinear programming and variational inequalities. It is well known that the convergence rate of such projection-type methods is linear at best. To our knowledge, superlinearly convergent algorithm for solving the SFP is not found in the literature. Studying the algorithm which has a fast convergence rate for solving the SFP is a very meaningful topic. It is the main purpose of this paper.

Basing on a merit function of the SFP, a Newton projection algorithm is presented for solving the SFP in this paper. The subproblem of this algorithm is to find an inexact solution of a convex quadratic programming, where the Newton step is judged by an inequality satisfied by the projection operator onto the set Q. We analyze the convergence properties of the method more completely. Under the condition that the solution set of the SFP is nonempty, we prove that the sequence of iterates converges globally to a solution of the SFP as long as the regularization parameter matrix is chosen properly. Further, in the analysis of the convergence rate, we must use the

second-order term of the merit function. Because the gradient of the merit function is the linear composite function of the projection operator, it is nonsmooth in general. So, we have two choices with respect to techniques: (i) we can choose directly a generalized Hessian from its generalized Jacobian defined by Clarke for the reason that its gradient is globally Lipschitzian (see, e.g., [12–17]). (ii) We can get its second-order term by regularing and smoothing its gradient (see, e.g., [18–22]). In this paper, we adopt the first choice. Under some local assumptions which are necessary for the case where the projection operator is nonsmooth, we prove that the sequence of iterates generated by the algorithm superlinearly converges to a regular solution of the SFP. It is worth noting that the condition of the regular solution is weaker than the conditions in [23–32]. Because those conditions are the sufficient condition for the regular solution.

The rest of this paper is organized as follows. Section 2 introduces the merit function of the SFP and its some related properties and concepts. Section 3 gives the Newton projection algorithm and proves its global convergence. Under some local assumptions which are necessary for the case where the gradient of the merit function is nonsmooth, Sect. 4 proves that the iterative sequence generated by the algorithm converges super-linearly to a regular solution of the SFP. Section 5 contains some numerical results. Section 6 is concluding remarks in which we point out that the method presented in this paper can be extended to solve a class of convex optimization which is more generalized than the SFP or a class of nonsmooth variational inequality problem.

2 A merit function of SFP and its properties

In this section, we introduce a merit function of the SFP and review some properties of it.

Byrne [10] gave a merit function of the SFP as follows:

$$f(x) = \frac{1}{2}\|Ax - P_Q(Ax)\|^2. \tag{2.1}$$

The function $f(x)$ is convex, continuously differentiable on $\Re^n$ and its derivative is the operator

$$g(x) = A^T(I - P_Q)Ax, \tag{2.2}$$

where, $P_Q(x)$ denotes the orthogonal projection from x onto Q, that is,

$$P_Q(x) = \operatorname{argmin}\{\|x - y\| \mid y \in Q\}.$$

Because Q is a closed convex set, $P_Q(x)$ is single valued. For the projection operator, the following properties are well-known.

Proposition 2.1 *For any $x, y \in \Re^n$, we have*

(i) $\|P_Q(y) - P_Q(x)\| \le \|y - x\|$;

(ii) $\langle P_Q(y) - P_Q(x), y - x \rangle \ge \|P_Q(y) - P_Q(x)\|^2$.

From part (i) of Proposition 2.1, we know that P_Q is a globally Lipschitzian operator. Part (ii) tells us that P_Q is co-coercive, thereby is monotone. From (i) and according to Rademacher's Theorem, we know that the generalized Jacobian defined by Clarke $\partial P_Q(x)$ exist almost everywhere and has the following property.

Proposition 2.2 ([20]) *For $x \in \Re^n$, all $V \in \partial P_Q(x)$ are symmetric, positive semi-definite and $\|V\| \leq 1$.*

Using Clarke's generalized Jacobian knowledge ([33,34]), we deduce the following from Proposition 2.2 and (2.2):

$$\partial g(x)d = A^T(I - \partial P_Q(Ax))Ad \tag{2.3}$$

for all $d \in \Re^n$.

In addition, we also have

Proposition 2.3 ([7]) *The following two conclusions hold.*

(i) $g(x)$ is globally Lipschitz continuous on $\Re^n$;
(ii) $g(x)$ is co-coercive on $\Re^n$, thereby is monotone.

From (2.1) and (2.2), we can prove easily the following proposition.

Proposition 2.4 ([7]) *The following statements are equivalent:*

(i) x^* is a solution of the SFP;
(ii) $x^* \in C$ and $f(x^*) = 0$;
(iii) $x^* \in C$ and $g(x^*) = 0$.

From Proposition 2.4, we know that if the solution set of the SFP is nonempty, then we have

$$\text{The solution set of the SFP} = \operatorname{argmin}\{f(x)|x \in C\}. \tag{2.4}$$

Thus, solving the SFP is equivalent to solve the convex optimization problem:

$$\min_{x \in C} f(x).$$

Next, we introduce an important concept about nonsmooth analysis used in this paper .

Definition 2.5 ([15,35]) Suppose $F : \Re^n \to \Re^m$ is a locally Lipschitzian function. We say that F is semismooth at x if for any $h \in \Re^n$, $h \neq 0$, the limit

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \to h, t \downarrow 0}} \{Vh'\}$$

exists. F is called semismooth on a set $S \subseteq \Re^n$ if F is semismooth at each $x \in S$.

Proposition 2.6 ([15]) *Suppose* $F : \Re^n \to \Re^m$ *is a locally Lipschitzian function. Then* F *is semismooth at* x *if and only if for any* $h \in \Re^n$, $h \neq 0$, *and* $\forall \, V \in \partial F(x + h)$, *we have*

$$F(x + h) - F(x) - Vh = o(\|h\|).$$

It was pointed out in [15] that sums, differences, scalar products and compositions of semismooth functions are still semismooth functions. Especially, when Q is a polyhedral, the projection operator $P_Q(\cdot)$ is semismooth. Thus, from (2.2), we have

Proposition 2.7 *Let Q be a polyhedral. Then*

(i) $P_Q(x)$ *is semismooth on* $\Re^m$;
(ii) $g(x)$ *is semismooth on* $\Re^n$.

3 Algorithm and its global convergence

In this section, we present an algorithm for solving the SFP and prove its global convergence. From (2.4), we know that under the condition that the solution set of the SFP is nonempty, we only need to design algorithm to solve the optimization problem $\min\{f(x)|x \in C\}$. As in [2], we assume all the projections in the algorithm are easily calculated.

Let the natural residual be $r(x) = x - P_C(x - g(x))$. For the current iterate $x^k \in C$, select a positive semidefinite matrix G_k and a regularization parameter $\mu_k > 0$, and let

$$\varphi_k(z) = g(x^k) + (G_k + \mu_k I)(z - x^k).$$

Consider a linearized subproblem: find $\hat{z}^k \in C$ such that for all $z \in C$,

$$\left\langle \varphi_k(\hat{z}^k), z - \hat{z}^k \right\rangle \geq 0. \tag{3.1}$$

As $G_k + \mu_k I$ is positive definite, the subproblem (3.1) always has a unique solution $\hat{z}^k$, hence with a zero residual

$$\hat{z}^k - P_C(\hat{z}^k - \varphi_k(\hat{z}^k)) = 0.$$

So, for any $\epsilon_k > 0$, the subproblem (3.1) always has inexact solutions z^k satisfying

$$\|e^k\| \leq \epsilon_k,$$

where

$$e^k = z^k - P_C(z^k - \varphi_k(z^k)).$$

We now formally state the algorithm.

Algorithm 3.1 Step 0. Take an $x^0 \in C$, and choose parameters $\eta, \alpha, \beta \in (0, 1)$, $\rho \in (0, \eta)$. Set $k := 0$.

Step 1. Choose a positive semidefinite matrix $G_k = A^T(I - V_k)A$, where V_k is a positive semidefinite matrix with $\|V_k\| \leq 1$, a regularization parameter $\mu_k > 0$ and $\rho_k \in (0, \frac{\rho}{M_k}]$, where $M_k = \|g(x^k)\| + \|G_k\| + 2\mu_k + 1$. Compute z^k, an inexact solution of the subproblem (3.1), such that

$$\|e^k\| \leq \rho_k \mu_k \|z^k - x^k\| \cdot \min\{\|z^k - x^k\|, 1\}. \tag{3.2}$$

If $z^k = x^k$, then stop.

Step 2. Let $y_C^k = P_C(z^k - \varphi_k(z^k))$, $p^k = P_Q(Ay_C^k) - P_Q(Ax^k) - V_k A(y_C^k - x^k)$. If

$$\langle p^k, Ax^k - Ay_C^k \rangle \leq (1 - \eta)\mu_k \|y_C^k - x^k\|^2, \tag{3.3}$$

then set $y^k = y_C^k$, $\hat{p}^k = g(y^k) - \varphi_k(z^k) + e^k$ and go to Step 4.

Step 3. Let $y^k = x^k + \lambda_k(z^k - x^k)$, where $\lambda_k = \beta^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$\|g(x^k + \beta^m(z^k - x^k)) - g(x^k)\| \leq \alpha(1 - \rho)\mu_k \|z^k - x^k\|. \tag{3.4}$$

Set $\hat{p}^k = g(y^k)$.

Step 4. Compute

$$\hat{x}^k = x^k - \frac{\langle \hat{p}^k, x^k - y^k \rangle}{\|\hat{p}^k\|^2} \hat{p}^k;$$
$$x^{k+1} = P_C(\hat{x}^k).$$

Set $k := k + 1$, go back to Step 1.

In order to explain the stop criterion and the feasibility of the algorithm, we give the following two lemmas, which will also be used in both the global convergence and the convergence rate analysis.

Lemma 3.2 *The following inequality holds for all k*

$$\langle g(x^k), x^k - z^k \rangle \geq \langle (G_k + (1 - \rho)\mu_k I)(z^k - x^k), z^k - x^k \rangle.$$

Proof From the definition of y_C^k, we have

$$\langle y_C^k - z^k + \varphi_k(z^k), x^k - y_C^k \rangle \geq 0,$$

which, together with (3.2) and the definition of ρ_k, implies that

$$
\begin{aligned}
\langle g(x^k), x^k - z^k \rangle &\geq \langle (G_k + \mu_k I)(z^k - x^k), z^k - x^k \rangle \\
&\quad + \|e^k\|^2 - \langle \varphi_k(z^k) + z^k - x^k, e^k \rangle \\
&\geq \langle (G_k + \mu_k I)(z^k - x^k), z^k - x^k \rangle - \langle \varphi_k(z^k) + z^k - x^k, e^k \rangle \\
&\geq \langle (G_k + \mu_k I)(z^k - x^k), z^k - x^k \rangle \\
&\quad - (\|g(x^k)\| \|e^k\| + (\|G_k\| + \mu_k + 1)\|z^k - x^k\| \|e^k\|) \\
&\geq \langle (G_k + \mu_k I)(z^k - x^k), z^k - x^k \rangle \\
&\quad - (\|g(x^k)\| + \|G_k\| + \mu_k + 1)\rho_k \mu_k \|z^k - x^k\|^2 \\
&\geq \langle (G_k + (1 - \rho)\mu_k I)(z^k - x^k), z^k - x^k \rangle.
\end{aligned}
$$

$\square$

Lemma 3.3 *For all* k, *the following inequality is true*

$$
\langle \hat{p}^k, x^k - y^k \rangle \geq \begin{cases} (\eta - \rho)\mu_k \|y^k - x^k\|^2, & \text{if (3.3) holds;} \\ \lambda_k (1 - \alpha)(1 - \rho)\mu_k \|z^k - x^k\|^2, & \text{otherwise.} \end{cases}
$$

Proof If (3.3) holds, then from (2.2) and the choice of G_k, we have

$$
\begin{aligned}
\langle \hat{p}^k, x^k - y^k \rangle &= \left\langle g(y^k) - \varphi_k(z^k) + e^k, x^k - y^k \right\rangle \\
&= \left\langle g(y^k) - g(x^k) - (G_k + \mu_k I)(z^k - x^k) + e^k, x^k - y^k \right\rangle \\
&= \langle A^T(I - P_Q)Ay^k - A^T(I - P_Q)Ax^k \\
&\quad - (A^T(I - V_k)A + \mu_k I)(z^k - x^k) + e^k, x^k - y^k \rangle \\
&= \mu_k \|y^k - x^k\|^2 - \langle P_Q(Ay^k) \\
&\quad - P_Q(Ax^k) - V_k A(y^k - x^k), A(x^k - y^k) \rangle \\
&\quad - \left\langle (G_k + \mu_k I - I)e^k, x^k - y^k \right\rangle \\
&= \mu_k \|y^k - x^k\|^2 - \langle p^k, Ax^k - Ay^k \rangle \\
&\quad - \langle (G_k + \mu_k I - I)e^k, x^k - y^k \rangle,
\end{aligned}
$$

which, together with (3.2) and (3.3) and the definition of ρ_k, can derive that

$$
\begin{aligned}
\langle \hat{p}^k, x^k - y^k \rangle &\geq \eta \mu_k \|y^k - x^k\|^2 - (\|G_k\| + \mu_k + 1)\|e^k\| \|y^k - x^k\| \\
&\geq [\eta - (\|G_k\| + \mu_k + 1)\frac{\rho_k}{1 - \rho_k \mu_k}]\mu_k \|y^k - x^k\|^2 \\
&\geq \frac{\eta - (\|G_k\| + 2\mu_k + 1)\rho_k}{1 - \rho_k \mu_k}\mu_k \|y^k - x^k\|^2 \\
&\geq (\eta - \rho)\mu_k \|y^k - x^k\|^2.
\end{aligned}
$$

Thus, the first inequality is proved. If (3.3) does not hold, then (3.4) holds. From Lemma 3.2, we have

$$
\begin{aligned}
\langle \hat{p}^k, x^k - y^k \rangle &= \langle g(y^k), x^k - y^k \rangle \\
&= \lambda_k [\langle g(x^k), x^k - z^k \rangle + \langle g(y^k) - g(x^k), x^k - z^k \rangle] \\
&\geq \lambda_k [(1 - \rho)\mu_k \|z^k - x^k\|^2 - \|g(y^k) - g(x^k)\| \|z^k - x^k\|] \\
&\geq \lambda_k (1 - \alpha)(1 - \rho)\mu_k \|z^k - x^k\|^2.
\end{aligned}
$$

This proves the second inequality. $\square$

Remark 3.4 From Lemmas (3.2) and (3.2), we know that the stop criterion is sufficient and necessary, that is, $x^k \in \arg\min_{x \in C} f(x)$ if and only if $z^k = x^k$. From now on, we assume that $z^k \neq x^k$ for all k, and an infinite sequence $\{x^k\}$ is generated. Next, we are ready to show that the algorithm is well defined. Firstly, if (3.3) holds, then $\hat{p}^k \neq 0$. Otherwise, from the first inequality of Lemmas (3.3) and (3.2), we obtain that $z^k = x^k$. Secondly, it is obvious that line search rule (3.4) is feasible and $\hat{p}^k = g(y^k) \neq 0$. In fact, if $g(y^k) = 0$, then from (3.4) we can know that

$$
\|g(x^k)\| \leq \alpha(1 - \rho)\mu_k \|z^k - x^k\|.
$$

Because $\alpha \in (0, 1)$ and matrix G_k is positive semidefinite, the above inequality contradicts Lemma 3.2. Overall we have proved that Algorithm 3.1 is well defined.

Lemma 3.5 *If* $\arg\min_{x \in C} f(x) \neq \emptyset$, *then for* $\forall x^* \in \arg\min_{x \in C} f(x)$ *and* $\forall k$, *we have*

$$
\frac{\langle \hat{p}^k, x^k - y^k \rangle^2}{\|\hat{p}^k\|^2} \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2.
$$

Proof It is easy to prove that for $\forall \bar{x} \in \Re^n$, the following equality is true

$$
\begin{aligned}
\|\hat{x}^k - \bar{x}\|^2 &= \|x^k - \bar{x}\|^2 - \frac{\langle \hat{p}^k, x^k - y^k \rangle^2}{\|\hat{p}^k\|^2} \\
&\quad + 2\frac{\langle \hat{p}^k, x^k - y^k \rangle}{\|\hat{p}^k\|^2} \langle \hat{p}^k, \bar{x} - y^k \rangle.
\end{aligned}
\tag{3.5}
$$

Now take an arbitrary $x^* \in \arg\min_{x \in C} f(x)$, since $y^k \in C$, we have

$$
\langle g(x^*), y^k - x^* \rangle \geq 0.
$$

Using the monotonicity of $g(\cdot)$, we obtain

$$
\langle g(y^k), y^k - x^* \rangle \geq 0.
\tag{3.6}
$$

Let $\bar{x} = x^*$ in (3.5), we consider the last term on the right hand side. Let

$$F_k = \frac{\langle \hat{p}^k, x^k - y^k \rangle}{\|\hat{p}^k\|^2} \langle \hat{p}^k, x^* - y^k \rangle.$$

First suppose (3.3) holds. From the first inequality of Lemma 3.3, we have

$$\langle \hat{p}^k, x^k - y^k \rangle \geq 0. \tag{3.7}$$

As $y^k = P_C(z^k - \varphi_k(z^k))$ and $x^* \in C$, from the property of projection, we have

$$\langle \varphi_k(z^k) - e^k, x^* - y^k \rangle \geq 0. \tag{3.8}$$

Thus, form (3.6) and (3.8) and the definition of $\hat{p}^k$, we see that

$$\langle \hat{p}^k, x^* - y^k \rangle \leq 0. \tag{3.9}$$

Hence, from (3.7) and (3.9), we obtain $F_k \leq 0$.

Second, assume (3.4) holds. In this case, $\hat{p}^k = g(y^k)$, hence, from (3.6), we have

$$\langle \hat{p}^k, x^* - y^k \rangle \leq 0.$$

So, from the second inequality of Lemma (3.3), we see that $\langle \hat{p}^k, x^k - y^k \rangle \geq 0$. Thus we obtain $F_k \leq 0$.

In conclusion, for each k, $F_k \leq 0$. Therefore, from the property of projection and (3.5), we have

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &\leq \|\hat{x}^k - x^*\|^2 \\
&\leq \|x^k - x^*\|^2 - \frac{\langle \hat{p}^k, x^k - y^k \rangle^2}{\|\hat{p}^k\|^2},
\end{aligned}$$

i.e., the lemma holds. □

Lemma 3.6 *Suppose that* $\arg\min_{x \in C} f(x) \neq \emptyset$, *and there exist constants* $0 < m < M$ *such that for all k and, starting with some index k_0,* $\mu_k \in [m, M]$. *Then*

$$\lim_{k \to \infty} \|z^k - x^k\| = 0, \quad \lim_{k \to \infty} \|y^k - x^k\| = 0.$$

Proof From Lemma 3.5, we know that the sequence $\{x^k\}$ is bounded and

$$\lim_{k \to \infty} \frac{\langle \hat{p}^k, x^k - y^k \rangle}{\|\hat{p}^k\|} = 0. \tag{3.10}$$

From Lemma 3.2 and the definition of μ_k, we have

$$\|g(x^k)\| \geq m(1 - \rho)\|z^k - x^k\|.$$

As $g(\cdot)$ is continuous and $\{x^k\}$ is bounded, from the above equality, we see that $\{z^k\}$ is bounded. Using (3.2) and the boundedness of $\{G_k\}$, we know that $\{y^k\}$ and $\{\hat{p}^k\}$ are also bounded. Hence, from (3.10) we obtain

$$\lim_{k\to\infty} \langle \hat{p}^k, x^k - y^k \rangle = 0. \tag{3.11}$$

Below we consider some possible cases. To this end, we divide all indices k into two classes.

$$K_N = \{k \mid y^k = y_C^k, \text{ i.e., (3.3) holds}\}.$$
$$K_A = \{k \mid y^k = x^k + \lambda_k(z^k - x^k), \lambda_k \text{ is obtained by (3.4)}\}.$$

It is obvious that $K_N \bigcup K_A = \{1, 2, \ldots\}$. From Lemma 3.3 and the definition of μ_k, we have

$$\langle \hat{p}^k, x^k - y^k \rangle \geq \begin{cases} (\eta - \rho)m\|y^k - x^k\|^2, & \text{if } k \in K_N; \\ \lambda_k(1 - \alpha)(1 - \rho)m\|z^k - x^k\|^2, & \text{if } k \in K_A. \end{cases} \tag{3.12}$$

If K_N is an infinite sequence, then from (3.11) and the first inequality of (3.12), we obtain $\lim\limits_{k\in K_N, k\to\infty} \|y^k - x^k\| = 0$. From (3.2), we have

$$\|z^k - x^k\| \leq \frac{1}{1 - \rho_k\mu_k}\|y^k - x^k\|$$
$$\leq \frac{1}{1 - \rho}\|y^k - x^k\|.$$

Thus, $\lim\limits_{k\in K_N, k\to\infty} \|z^k - x^k\| = 0$.

If K_A is an infinite sequence, then from (3.11) and the second inequality of (3.12), we obtain

$$\lim_{k\in K_A, k\to\infty} \lambda_k\|z^k - x^k\|^2 = 0. \tag{3.13}$$

Suppose that there exists an infinite sequence $K_0 \subseteq K_A$, such that

$$\lim_{k\in K_0, k\to\infty} \|z^k - x^k\| > 0. \tag{3.14}$$

As $\{z^k\}$ and $\{x^k\}$ are bounded, without loss of generality, we may assume that

$$\lim_{k\in K_0, k\to\infty} z^k = \bar{z}, \quad \lim_{k\in K_0, k\to\infty} x^k = \bar{x}.$$

From (3.14) we know that $\bar{z} \neq \bar{x}$. Then by (3.13) and (3.14),

$$\lim_{k\in K_0, k\to\infty} \lambda_k = 0. \tag{3.15}$$

By (3.15), line search rule (3.4) and the definition of μ_k, when $k \in K_0$ and k is sufficiently large,

$$\|g(x^k + \beta^{-1}\lambda_k(z^k - x^k)) - g(x^k)\| \geq \alpha(1 - \rho)\mu_k\|z^k - x^k\|$$
$$\geq \alpha(1 - \rho)m\|z^k - x^k\|.$$

Since (3.15) holds, passing onto the limit as $k \to \infty$, we obtain $\|\bar{z} - \bar{x}\| = 0$, which is a contradiction. Thus we obtain $\lim\limits_{k \in K_A, k \to \infty} \|z^k - x^k\| = 0$ and $\lim\limits_{k \in K_A, k \to \infty} \|y^k - x^k\| = 0$.

□

We now give the global convergence theorem of Algorithm 3.1.

Theorem 3.7 *The following two conclusions hold for Algorithm 3.1.*

(1) *If* $\arg\min\limits_{x \in C} f(x) \neq \emptyset$, *then* $\{x^k\}$ *is bounded. Suppose that there exists a constant* $M > 0$ *such that for all k and, starting with some index k_0, $\mu_k \in [\|r(x^k)\|, M]$. Then*

$$\lim_{k \to \infty} x^k = x^* \in \arg\min_{x \in C} f(x). \tag{3.16}$$

(2) *If the sequence* $\{x^k\}$ *is bounded and starting with some index k_0, $\mu_k \in [\|r(x^k)\|, M]$, then, $\arg\min\limits_{x \in C} f(x) \neq \emptyset$ and (3.16) holds.*

Proof (1) If $\arg\min\limits_{x \in C} f(x) \neq \emptyset$, then from Lemma 3.5, $\{x^k\}$ is bounded and for any $x^* \in \arg\min\limits_{x \in C} f(x)$, $\{\|x^k - x^*\|\}$ is monotonically decreasing. So, in order to prove conclusion (1), it suffices to show that $\liminf\limits_{k \to \infty} \|r(x^k)\| = 0$. Suppose the conclusion does not hold, then it would be

$$\liminf_{k \to \infty} \|r(x^k)\| > 0. \tag{3.17}$$

From the definition of μ_k, there exists a constant $m > 0$ such that for all sufficiently large k, $\mu_k \in [m, M]$. Thus, from Lemmas (3.6) and (3.2), we have

$$\lim_{k \to \infty} \|z^k - x^k\| = 0, \ \lim_{k \to \infty} \|y_C^k - x^k\| = 0. \tag{3.18}$$

As $\{x^k\}$ is bounded, there is a subsequence $K \subseteq \{1, 2, \cdots\}$ such that

$$\lim_{k \in K, k \to \infty} x^k = \bar{x}. \tag{3.19}$$

From (3.18) and (3.19) and the fact that $\{G_k\}$ and $\{\mu_k\}$ are bounded, we have

$$\lim_{k \in K, k \to \infty} \varphi_k(z^k) = g(\bar{x}). \tag{3.20}$$

From the definition of y_C^k and the property of projection, for all $x \in C$, we have

$$\langle \varphi_k(z^k) - e^k, x - y_C^k \rangle \geq 0.$$

Using (3.18), (3.19) and (3.20) and taking limits on both sides of the above inequality when $k \in K$ and $k \to \infty$,

$$\langle g(\bar{x}), x - \bar{x} \rangle \geq 0.$$

As x can be any point in C, the above inequality means that $\bar{x} \in \arg\min_{x \in C} f(x)$. From this we can conclude that $\lim_{k \to \infty} \|r(x^k)\| = 0$, which contradicts (3.17). So part (1) is proved.

(2) Now suppose that $\{x^k\}$ is bounded and $\mu_k \in [\|r(x^k)\|, M]$. In this case we only need to prove that

$$\liminf_{k \to \infty} \|r(x^k)\| = 0,$$

which implies that $\arg\min_{x \in C} f(x) \neq \emptyset$, then by conclusion (1) we know that conclusion (2) also holds. Again we prove it by contradiction. Suppose (3.17) holds. Then from the definition of μ_k, there exists $m > 0$ such that for all sufficiently large k, $\mu_k \in [m, M]$. From Lemma 3.2, following the same line as in the proof for Lemma 3.6, we see that $\{z^k\}, \{y^k\}$ are bounded. Hence, we can select a constant $M_0 > 0$, such that for all k, we have

$$\max\{\|x^k\|, \|z^k\|, \|y^k\|\} \leq M_0.$$

We arbitrarily take an $\bar{x} \in C$, and let

$$Y = \{x \in \Re^n \,|\, \|x\| \leq M_0 + \|\bar{x}\|\} \bigcap C.$$

As $\bar{x} \in Y$, Y is nonempty. Also, Y is a compact set and to be contained in C. Hence $\arg\min_{x \in Y} f(x) \neq \emptyset$. Since $\{x^k\}, \{z^k\}$ and $\{y^k\} \subseteq Y$, we can regard $\{x^k\}$ as the iterative sequence obtained by using the Algorithm 3.1 for the problem $\min_{x \in Y} f(x)$. As $\arg\min_{x \in Y} f(x) \neq \emptyset$, by the proved conclusion (1), we have

$$\lim_{k \to \infty} x^k = x^* \in \arg\min_{x \in Y} f(x). \tag{3.21}$$

As $\bar{x} \in Y$, from (3.21), we have

$$\langle g(x^*), \bar{x} - x^* \rangle \geq 0. \tag{3.22}$$

Notice that $x^* \in C$ and $\bar{x}$ can be any point in C, (3.22) implies that $x^* \in \arg\min_{x \in C} f(x)$. Hence we see that $\lim_{k \to \infty} \|r(x^k)\| = 0$, which contradicts (3.17). $\qquad \square$

In Algorithm 3.1, we can choose a special case of V_k, i.e., $V_k = 0$. Then, from the monotonicity of $P_Q(\cdot)$, we have

$$\langle p^k, Ax^k - Ay_C^k \rangle = \langle P_Q(Ay_C^k) - P_Q(Ax^k), Ax^k - Ay_C^k \rangle \leq 0.$$

So, (3.3) is obvious true. Thus we can drive a special case of Algorithm 3.1, that is, the following algorithm.

Algorithm 3.8 Step 0. Take an $x^0 \in C$, and choose parameters $\eta \in (0, 1)$, $\rho \in (0, \eta)$. Set $k := 0$.

Step 1. Choose a positive semidefinite matrix $G_k = A^T A$, a regularization parameter $\mu_k > 0$ and $\rho_k \in (0, \frac{\rho}{M_k}]$, where $M_k = \|g(x^k)\| + \|A^T A\| + 2\mu_k + 1$. Compute z^k, an inexact solution of the subproblem (3.1), such that (3.2) holds. If $z^k = x^k$, then stop.

Step 2. Set $y^k = P_C(z^k - \varphi_k(z^k))$, $\hat{p}^k = g(y^k) - \varphi_k(z^k) + e^k$. Compute

$$\hat{x}^k = x^k - \frac{\langle \hat{p}^k, x^k - y^k \rangle}{\|\hat{p}^k\|^2} \hat{p}^k;$$

$$x^{k+1} = P_C(\hat{x}^k).$$

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

From Theorem 3.7, we can drive the global convergence of Algorithm 3.8.

4 Convergence rate analysis

In this section, we will discuss the convergence rate of the algorithm. To this end, for the nonsmooth $g(\cdot)$, we first introduce the definition of regular solution [28,36]. Consider a perturbed linear variational inequality problem ($VIP(C, g^\omega)$), where

$$g^\omega(x) = g(x^*) + \omega + G^*(x - x^*), \tag{4.1}$$

and $G^* = A^T(I - V(x^*))A$, where $V(x^*) \in \partial P_Q(Ax^*)$.

Definition 4.1 Let $x^* \in \arg\min_{x \in C} f(x)$. We call x^* a regular solution if there exist a neighborhood $U(x^*)$ of x^* and a neighborhood $W(0)$ of the origin, such that for all $\omega \in W(0)$, problem ($VIP(C, g^\omega)$) has a unique solution $x(\omega)$ over $U(x^*)$, and furthermore, $x(\omega)$ is Lipschitz continuous over $W(0)$, i.e., there is a constant $L^* > 0$, such that for all ω' and $\omega'' \in W(0)$, we have

$$\|x(\omega') - x(\omega'')\| \leq L^* \|\omega' - \omega''\|,$$

where L^* is called a regular Lipschitz constant for the solution x^*.

It is noticed that [28,36] gave the regular solution for smooth analysis. Although this concept is applied to optimization, especially in the local convergence analysis of the algorithms for solving variational inequality problem, most references used the sufficient condition of regular solution. For example, the sufficient condition used in [25–27,29] and the references therein was given under the condition that the mapping F satisfies some local assumptions and some constraint qualifications. Armand et al., Kanzow and Fukushima, Peng and Fukushima, Solodov and Svaiter, Sun et al. and Taji et al. [23,24,28,30–32] used the stronger sufficient condition than those used in the references mentioned above. Unlike these references, in this section, under some local assumptions which are necessary for the case where $P_Q(\cdot)$ is nonsmooth, we prove that the sequence of iterates generated by the algorithm superlinearly converges to a regular solution of the SFP.

Let $\arg\min_{x \in C} f(x) \neq \emptyset$. We choose $G_k = A^T(I - V(x^k))A$, where $V(x^k) \in \partial P_Q(Ax^k)$ in Algorithm 3.1. From Proposition 2.2, we know that G_k is symmetric and positive semidefinite and $\{G_k\}$ is bounded. Choose $\rho_k \to 0(k \to \infty)$, $\mu_k = \max\{\|r(x^k)\|, S(x^k)^t\}$, $t \in (0, 1)$, where

$$S(x^k) = \sup \left\{ \frac{\|P_Q(Ax^k + u) - P_Q(Ax^k) - V(x^k)u\|}{\|u\|} \Big| u \neq 0, \|u\| \leq \|r(x^k)\|^t \right\}.$$

From Propositions 2.1 and 2.2, we know that $\{S(x^k)\}$ is bounded.

Under the above choice, from Theorems (3.7) and (3.16) holds, that is

$$\lim_{k \to \infty} x^k = x^* \in \arg\min_{x \in C} f(x).$$

All the discussions below are about the analysis of the convergence rate of the sequence $\{x^k\}$ converging to x^* which aimed to (3.16) under the above choice.

We need the following assumptions.

(A1): $\lim_{k \to \infty} z^k = x^*$;

(A2): $\lim_{k \to \infty} V(x^k) = V(x^*)$, where, $V(x^*)$ is chosen by (4.1);

(A3): $\lim_{k \to \infty} S(x^k) = 0$;

(A4): $P_Q(Ax^k) - P_Q(Ax^*) - V(x^k)A(x^k - x^*) = o(\|x^k - x^*\|)$.

The above assumptions, especially (A2), (A3) and (A4), are necessary for the projection operator $P_Q(\cdot)$ which is nonsmooth. They have obvious background, which we will state in detail latter.

Lemma 4.2 *Let x^* be a regular solution and the assumptions (A1), (A2) and (A3) hold. Then for all sufficiently large k, we have*

$$\|A(y_C^k - x^k)\| \leq \|r(x^k)\|^t.$$

Proof Let

$$\omega'_k = \varphi_k(z^k) - e^k - g(x^*) - G^*(y^k_C - x^*), \qquad (4.2)$$

$$\omega''_k = g(x^k) - r(x^k) - g(x^*) - G^*(x^k_C - x^*), \qquad (4.3)$$

where $x^k_C = P_C(x^k - g(x^k))$. From the definition of y^k_C and x^k_C, they are the solution of perturbation $VIP(C, g^{\omega'_k})$ and $VIP(C, g^{\omega''_k})$ respectively, where

$$g^{\omega'_k}(x) = g(x^*) + \omega'_k + G^*(x - x^*), \qquad (4.4)$$

$$g^{\omega''_k}(x) = g(x^*) + \omega''_k + G^*(x - x^*). \qquad (4.5)$$

From assumption (A1) and (3.16), we have

$$\lim_{k \to \infty} y^k_C = x^*, \ \lim_{k \to \infty} x^k_C = x^*. \qquad (4.6)$$

In addition, from (4.2) and (4.3), we have

$$\lim_{k \to \infty} \omega'_k = 0, \ \lim_{k \to \infty} \omega''_k = 0. \qquad (4.7)$$

From (4.6) and (4.7), for all sufficiently large k, we obtain $y^k_C, x^k_C \in U(x^*), \omega'_k, \omega''_k \in W(0)$, where $U(x^*)$ and $W(0)$ is defined as in Definition 4.1. From assumption, x^* is a regular solution. Thus, from Definition 4.1, y^k_C and x^k_C are the unique solution of $VIP(C, g^{\omega'_k})$ and $VIP(C, g^{\omega''_k})$ over $U(x^*)$ and satisfy

$$\|y^k_C - x^k_C\| \le L^*\|\omega'_k - \omega''_k\|. \qquad (4.8)$$

From (3.2) and (4.8), we have

$$\begin{aligned}
\|y^k_C - x^k\| &\le \|y^k_C - x^k_C\| + \|r(x^k)\| \\
&\le L^*\|\omega'_k - \omega''_k\| + \|r(x^k)\| \\
&= L^*\|\varphi_k(z^k) - e^k - G^*(y^k_C - x^*) \\
&\quad - g(x^k) + r(x^k) + G^*(x^k_C - x^*)\| + \|r(x^k)\| \\
&= L^*\|(G_k - G^* + \mu_k I)(z^k - x^k) \\
&\quad + (I - G^*)r(x^k) + (G^* - I)e^k\| + \|r(x^k)\| \\
&\le L^*[(\|G^*\| + 1)\|r(x^k)\| + (\|G^*\| + 1)\|e^k\| \\
&\quad + (\|G_k - G^*\| + \mu_k)\|z^k - x^k\|] + \|r(x^k)\| \\
&\le L^*[(\|G^*\| + 1)\rho_k\mu_k + \|G_k - G^*\| + \mu_k]\|z^k - x^k\| \\
&\quad + [L^*(\|G^*\| + 1) + 1]\|r(x^k)\| \\
&\le L^*[(\|G^*\| + 1)\rho_k\mu_k + (\|G_k - G^*\| + \mu_k)]\frac{1}{1 - \rho_k\mu_k}\|y^k_C - x^k\| \\
&\quad + [L^*(\|G^*\| + 1) + 1]\|r(x^k)\|. \qquad (4.9)
\end{aligned}$$

From the definition of G_k and μ_k, using assumption (A2) and (A3), we obtain $\|G_k - G^*\| \to 0, \mu_k \to 0(k \to \infty)$. Thus, we can choose sufficiently large k, such that

$$L^*[(\|G^*\| + 1)\rho_k\mu_k + (\|G_k - G^*\| + \mu_k)]\frac{1}{1 - \rho_k\mu_k} \leq \frac{1}{2 + L^*(\|G^*\| + 1)}.$$

Thus, from (4.9), we have

$$\|y_C^k - x^k\| \leq \frac{1}{2 + L^*(\|G^*\| + 1)}\|y_C^k - x^k\|$$
$$+ [L^*(\|G^*\| + 1) + 1]\|r(x^k)\|.$$

Therefore, we have

$$\|y_C^k - x^k\| \leq [2 + L^*(\|G^*\| + 1)]\|r(x^k)\|,$$
$$\|A(y_C^k - x^k)\| \leq \|A\|[2 + L^*(\|G^*\| + 1)]\|r(x^k)\|.$$

Because $\|r(x^k)\| \to 0(k \to \infty)$, for all sufficiently large k, we have

$$\|A\|[2 + L^*(\|G^*\| + 1)]\|r(x^k)\|^{1-t} \leq 1.$$

Thus, we can get

$$\|A(y_C^k - x^k)\| \leq \|r(x^k)\|^t.$$

□

Lemma 4.3 *The following inequality holds for* $\forall k$

$$\|\hat{x}^k - y^k\| \leq \frac{\|\hat{p}^k + \mu_k(y^k - x^k)\|}{\mu_k}.$$

Proof From the following equality, this conclusion can be proved.

$$\|\hat{p}^k + \mu_k(y^k - x^k)\|^2 - \mu_k^2\|\hat{x}^k - y^k\|^2$$
$$= \|\hat{p}^k\|^2 - 2\mu_k\langle\hat{p}^k, x^k - y^k\rangle + \mu_k^2\frac{\langle\hat{p}^k, x^k - y^k\rangle^2}{\|\hat{p}^k\|^2}$$
$$= \|\hat{p}^k - \mu_k\frac{\langle\hat{p}^k, x^k - y^k\rangle}{\|\hat{p}^k\|^2}\hat{p}^k\|^2.$$

□

Lemma 4.4 *Let* x^* *be a regular solution and the assumptions (A1), (A2) and (A3) hold. Then for all sufficiently large* k, *(3.3) holds.*

Proof If $A(x^k - y_C^k) = 0$, then (3.3) is obviously true. Now, we suppose that $A(x^k - y_C^k) \neq 0$. From Lemma 4.2, for all sufficiently large k, we have

$$\|A(y_C^k - x^k)\| \leq \|r(x^k)\|^t. \tag{4.10}$$

Thus, from (4.10) and the definition of μ_k, we have

$$
\begin{aligned}
&\langle p^k, A(x^k - y_C^k) \rangle \\
&= \left\langle P_Q(Ay_C^k) - P_Q(Ax^k) - V(x^k)A(y_C^k - x^k), A(x^k - y_C^k) \right\rangle \\
&\leq \|A\|^2 \frac{\|P_Q(Ax^k + A(y_C^k - x^k)) - P_Q(Ax^k) - V(x^k)A(y_C^k - x^k)\|}{\|A(y_C^k - x^k)\|} \|y_C^k - x^k\|^2 \\
&\leq \|A\|^2 \sup \left\{ \frac{\|P_Q(Ax^k + u) - P_Q(Ax^k) - V(x^k)u\|}{\|u\|} \,\middle|\, u \neq 0, \|u\| \leq \|r(x^k)\|^t \right\} \|y_C^k - x^k\|^2 \\
&= \|A\|^2 S(x^k) \|y_C^k - x^k\|^2 \\
&\leq \|A\|^2 S(x^k)^{1-t} \mu_k \|y_C^k - x^k\|^2.
\end{aligned}
$$

From assumption (A3), $S(x^k) \to 0 (k \to \infty)$. So, for all sufficiently large k, $\|A\|^2 S(x^k)^{1-t} \leq 1 - \eta$. Thus, we have

$$\langle p^k, A(x^k - y_C^k) \rangle \leq (1 - \eta)\mu_k \|y_C^k - x^k\|^2,$$

i.e., (3.3) holds.   □

Now, we give the theorem of the convergence rate analysis of Algorithm 3.1.

Theorem 4.5 *Let x^* be a regular solution and the assumptions (A1), (A2), (A3)and (A4) hold. Then $\{x^k\}$ superlinearly converges to x^*, that is,*

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

Proof By Lemma 4.4, for sufficiently large k, (3.3) holds. Thus, we see that $y^k = y_C^k$, $\hat{p}^k = g(y^k) - \varphi_k(z^k) + e^k$. Next, we will estimate $\|x^{k+1} - x^*\|$. We have

$$
\begin{aligned}
\|x^{k+1} - x^*\| &\leq \|\hat{x}^k - x^*\| \\
&\leq \|\hat{x}^k - y^k\| + \|y^k - x^*\|.
\end{aligned} \tag{4.11}
$$

We first estimate $\|y^k - x^*\|$. From assumption, x^* is a regular solution, hence is the unique solution when $\omega = 0$, i.e., the solution to the problem $VIP(C, g^0)$ over $U(x^*)$. Thus, from Definition 4.1, we have

$$\|y^k - x^*\| \leq L^* \|\omega_k'\|.$$

Noting the definition of G_k and G^*, by (3.2) and (4.2) and assumption (A4), we can get from the above inequality

$$
\begin{aligned}
\|y^k - x^*\| &\leq L^* \|\varphi_k(z^k) - e^k - g(x^*) - G^*(y^k - x^*)\| \\
&= L^* \|g(x^k) - g(x^*) - G^*(x^k - x^*) \\
&\quad + (G_k - G^* + \mu_k)(z^k - x^k) + (G^* - I)e^k\| \\
&= L^* \|A^T (P_Q(Ax^*) - P_Q(Ax^k) + V(x^*)A(x^k - x^*)) \\
&\quad + (G_k - G^* + \mu_k)(z^k - x^k) + (G^* - I)e^k\| \\
&\leq L^* [\|A\| \|P_Q(Ax^k) - P_Q(Ax^*) - V(x^*)A(x^k - x^*)\| \\
&\quad + (\|G_k - G^*\| + \mu_k)\|z^k - x^k\| \\
&\quad + (\|G^*\| + 1)\|e^k\|] \\
&\leq L^* \|A\| [\|P_Q(Ax^k) - P_Q(Ax^*) - V(x^k)A(x^k - x^*)\| \\
&\quad + \|V(x^k) - V(x^*)\| \|A\| \|x^k - x^*\| \\
&\quad + L^* [\|G_k - G^*\| + \mu_k + (\|G^*\| + 1)\rho_k\mu_k]\|z^k - x^k\| \\
&\leq L^* \|A\| \left[\frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} + \|V(x^k) - V(x^*)\| \|A\| \right] \|x^k - x^*\| \\
&\quad + L^* [\|G_k - G^*\| + \mu_k + (\|G^*\| + 1)\rho_k\mu_k] \frac{1}{1 - \rho_k\mu_k} \|y^k - x^k\| \\
&\leq S_{1k}\|x^k - x^*\| + S_{2k}\|y^k - x^*\|,
\end{aligned}
\tag{4.12}
$$

where,

$$
S_{1k} = L^* \|A\| [o(1) + \|V(x^k) - V(x^*)\| \|A\|],
$$
$$
S_{2k} = L^* [\|G_k - G^*\| + \mu_k + (\|G^*\| + 1)\rho_k\mu_k] \frac{1}{1 - \rho_k\mu_k}.
$$

From assumption (A2) and (A3), $S_{1k} \to 0$, $S_{2k} \to 0 (k \to \infty)$. Thus, from (4.12), we have

$$
\|y^k - x^*\| \leq \frac{S_{1k}}{1 - S_{2k}} \|x^k - x^*\|.
\tag{4.13}
$$

We next estimate $\|\hat{x}^k - y^k\|$. By Lemmas (4.3), (3.2) and (4.13), we have

$$
\begin{aligned}
\|\hat{x}^k - y^k\| &\leq \frac{\|\hat{p}^k - \mu_k(y^k - x^k)\|}{\mu_k} \\
&= \frac{\|g(y^k) - \varphi_k(z^k) + e^k + \mu_k(y^k - x^k)\|}{\mu_k} \\
&= \frac{\|g(y^k) - g(x^k) - G_k(y^k - x^k) - (G_k + \mu_k I)e^k + e^k\|}{\mu_k} \\
&\leq \frac{\|A\| \|P_Q(Ay^k) - P_Q(Ax^k) - V(x^k)A(y^k - x^k)\|}{\mu_k \|A(y^k - x^k)\|} \|A(y^k - x^k)\|
\end{aligned}
$$

$$+ \frac{\|G_k\| + \mu_k + 1}{\mu_k} \|e^k\|$$

$$\leq \frac{\|A\|^2 S(x^k) \|y^k - x^k\|}{\mu_k} + \frac{\|G_k\| + \mu_k + 1}{\mu_k} \|e^k\|$$

$$\leq \left[\|A\|^2 S(x^k)^{1-t} + (\|G_k\| + \mu_k + 1) \frac{\rho_k}{1 - \rho_k \mu_k} \right] \|y^k - x^k\|$$

$$\leq \left[\|A\|^2 S(x^k)^{1-t} + (\|G_k\| + \mu_k + 1) \frac{\rho_k}{1 - \rho_k \mu_k} \right]$$

$$\left(\frac{S_{1k}}{1 - S_{2k}} + 1 \right) \|x^k - x^*\|$$

$$= S_{3k} \left(\frac{S_{1k}}{1 - S_{2k}} + 1 \right) \|x^k - x^*\|, \tag{4.14}$$

where, $S_{3k} = \|A\|^2 S(x^k)^{1-t} + (\|G_k\| + \mu_k + 1) \frac{\rho_k}{1 - \rho_k \mu_k}$. Because $\{G_k\}$ is bounded, $\rho_k \to 0 (k \to \infty)$, from (A3), we obtain $S(x^k) \to 0 (k \to \infty)$. By (4.11), (4.13) and (4.14), we have

$$\|x^{k+1} - x^*\| \leq \left[\frac{S_{1k}}{1 - S_{2k}} + S_{3k} \left(\frac{S_{1k}}{1 - S_{2k}} + 1 \right) \right] \|x^k - x^*\|.$$

As $S_{1k} \to 0$, $S_{2k} \to 0$, $S_{3k} \to 0 (k \to \infty)$, we have

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

This completes the proof of this theorem.   □

Remark 4.6 Now we are ready to analyze the background of assumptions (A1), (A2), (A3) and (A4), which can show that these assumptions are necessary to the nonsmooth of P_Q.

(1) If $P_Q(\cdot)$ is semismooth at x^* (for example, Q is a polyhedron), then from Proposition 2.5, assumption (A4) holds.
(2) If $P_Q(\cdot)$ is smooth at x^* (for example, $Q = \{z \in \Re^m | Bz = b\}$, where B is a $p \times m$ matrix and has full row rank), then assumptions (A2), (A3) and (A4) are satisfied.
(3) If $P_Q(\cdot)$ is smooth at x^* and $\bigtriangledown g(x^*)$ is positive definite, then assumptions (A2), (A3) and (A4) are satisfied. Furthermore, by Lemma 3.2, assumption (A1) is also satisfied. From Lemma 6 of [28], x^* is a regular solution. Thus, in this case, we only need to choose $G_k = \bigtriangledown g(x^k)$, $\mu_k = \max\{\|r(x^k)\|, S(x^k)^t\}$ in Theorem 4.5, then $\{x^k\}$ superlinearly converges to x^*. It is worth noting that the condition of this result is weaker than that used in the related literature. For example, in Theorem 4.3 of [30], it not only need that $\nabla F(x^*)$ is positive definite, but also need that $\nabla F(\cdot)$ is Hölder continuous around x^* with degree p. On the other hand, we must point out that the Algorithm 2.2 of [30] play an enlightenment role for the algorithms of this paper.

By an argument analogous to that used in [28], Lemma 6, in the case of nonsmooth, we can prove that if G^* is positive definite, then x^* is a regular solution. In addition, by Lemma 3.2, we know that (A1) holds. Thus, from Theorem 4.5, we can obtain the following corollary.

Corollary 4.7 *Suppose that G^* in (4.1) is positive definite and assumptions (A2), (A3) and (A4) hold. Then, $\{x^k\}$ superlinearly converges to x^*.*

Below we give the convergence analysis for the algorithm under the following assumption.

Assumption (A_1') : $\{z^k\}$ is bounded and $\inf_k \lambda_k > 0$.

If the assumption (A1) is replaced by (A_1') in Theorem 4.5, whether the convergence property of the algorithm can also hold? We first give the following lemma.

Lemma 4.8 *Suppose that $\arg\min_{x\in C} f(x) = \{x^*\}$ and assumption (A_1') holds. Then*

$$\lim_{k\to\infty} z^k = x^*.$$

Proof From the assumption, we know that $\{z^k\}$ is bounded. So, we only need to prove that any of its accumulation point $z^* = x^*$. Suppose there exists an infinite sequence $K \subseteq \{1, 2, \cdots\}$ such that

$$\lim_{k\in K, k\to\infty} z^k = z^*. \tag{4.15}$$

Because $\rho_k \to 0(k \to \infty)$, from (3.2) and (4.15), we have

$$\lim_{k\in K, k\to\infty} y_C^k = z^*. \tag{4.16}$$

In addition, as $\{\lambda_k\}$ is bounded, without loss of generality, we may assume

$$\lim_{k\in K, k\to\infty} \lambda_k = \lambda^*. \tag{4.17}$$

From (4.15) and Lemma 3.2, noting that $x^* \in \arg\min_{x\in C} f(x)$, we can conclude that

$$\langle g(x^*), x^* - z^* \rangle = 0. \tag{4.18}$$

From the boundedness of $\{z^k\}, \{y^k\}, \{\mu_k\}, \{G_k\}$ and Lemma 3.5, we have

$$\lim_{k\to\infty} \langle \hat{p}^k, x^k - y^k \rangle = 0. \tag{4.19}$$

If (3.3) holds, then $y^k = y_C^k$, $\hat{p}^k = g(y^k) - \varphi_k(z^k) + e^k$, hence we have

$$\langle g(y^k), x^k - y^k \rangle \geq \langle \hat{p}^k, x^k - y^k \rangle. \tag{4.20}$$

Thus, from (4.16), (4.18), (4.19) and (4.20) and taking into the monotonicity of $g(\cdot)$ account, we have

$$\langle g(x^*) - g(z^*), x^* - z^* \rangle = 0,$$

which, together with the co-coercivity of $g(\cdot)$, implies that

$$g(x^*) = g(z^*). \tag{4.21}$$

Choosing any $x \in C$, from (4.18) and (4.21), we have

$$
\begin{aligned}
\langle g(z^*), x - z^* \rangle &= \langle g(x^*), x - x^* \rangle + \langle g(x^*), x^* - z^* \rangle \\
&= \langle g(x^*), x - x^* \rangle \\
&\geq 0.
\end{aligned}
$$

It follows that $z^* = x^*$.

If $y^k = x^k + \lambda_k(z^k - x^k)$, where, λ_k is generated by linesearch (3.4). Then, from (4.15) and (4.17), we have

$$\lim_{k \in K, k \to \infty} y^k = y^* = x^* + \lambda^*(z^* - x^*).$$

Following the same line as in the above proof, we can get

$$y^* = x^* + \lambda^*(z^* - x^*) = x^*.$$

From the assumption, we know that $\lambda^* > 0$. It follows immediately from the above equality that $z^* = x^*$. $\qquad\square$

From Lemma 4.8 and Theorem 4.5, we can get the following corollary.

Corollary 4.9 *Suppose that* $\arg\min\limits_{x \in C} f(x) = \{x^*\}, x^*$ *is a regular solution and assumptions* (A_1'), *(A2), (A3) and (A4) hold. Then,* $\{x^k\}$ *superlinearly converges to* x^*.

5 Numerical results

To give some insight into the behavior of the algorithms presented in this paper, we implemented them in MATLAB to solve the following three examples. We use $\|z^k - x^k\| < \epsilon$ as the stopping criteria. Throughout the computational experiments, the parameter ϵ used in Algorithm 3.8 were set as $\epsilon = 10^{-6}$. In the results reported below, the approximate solution is referred to the last iteration and the number of iterations indicated for each test problem and each specified initial point is the number of main iterations; the number of inner iterations in the subroutine is not included.

Table 1 Number of iterations for Algorithm 3.8 being applied to Example 1

Starting points	Number of iterations	Approximate solution
$(0, 0, 0)^T$	14	$(0.3333, 1.3333, -0.6667)^T$
$(1, 1, 1)^T$	14	$(0.3333, 1.3333, -0.6667)^T$
$(1, 2, 3)^T$	14	$(-0.4574, 1.7287, -0.2713)^T$
$(100, -10, 30)^T$	21	$(1.3874, 0.8063, -1.1937)^T$

Table 2 Number of iterations for Algorithm 3.8 being applied to Example 2

Starting points	Number of iterations	Approximate solution
$(0, 0, 0)^T$	0	$(0, 0, 0)^T$
$(1, 1, 1)^T$	16	$(1.0000, 0.8500, 0.5500)^T$
$(1, 2, 3)^T$	18	$(0.0282, 0.5669, 0.6444)^T$
$(10, -5, 10)^T$	19	$(6.1111, -9.5834, 4.0278)^T$

Example 1 (An $\ell_2 -$ norm problem of linear equations[37])

Find a solution which satisfies that $\|x\|_2 \leq 2$ of the following linear equations:

$$\begin{cases} 2x_1 + 3x_2 + x_3 = 4 \\ x_1 - 2x_2 + 4x_3 = -5 \\ 3x_1 + 8x_2 - 2x_3 = 13 \\ 4x_1 - x_2 + 9x_3 = -6 \end{cases}$$

This problem is an SFP for $A = \begin{pmatrix} 2 & 3 & 1 \\ 1 & -2 & 4 \\ 3 & 8 & -2 \\ 4 & -1 & 9 \end{pmatrix}$, $C = \{x \,|\, \|x\|_2 \leq 2\}$ and

$Q = \{y \,|\, y = (4, -5, 13, -6)^T\}$. The test results for Algorithm 3.8 being applied to Example 1 are listed in Table 1 using different starting points.

Example 2 (A SFP)

Let $C = \{x \in \Re^3 \,|\, x_1 + x_2 + x_3 \leq 3\}$, $Q = \{y \in \Re^4 \,|\, y_1 - 2y_2 + 3y_3 + 7y_4 \leq 5\}$.
$A = \begin{pmatrix} 2 & -1 & 3 \\ 4 & 2 & 5 \\ 2 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$. Find some point $x \in C$ with $Ax \in Q$.

The test results for Algorithm 3.8 being applied to Example 2 are listed in Table 2 using different starting points.

Example 3 (A SFP generated randomly[38])

The SFP to find $x \in C = \{x \in \Re^n \,|\, \|x\|_2 \leq 0.25\}$, such that $Ax \in Q = \{y \in \Re^m \,|\, 0.6 \leq y_j \leq 1, j = 1, 2, \cdots, m\}$, where, the matrix $A = (a_{ij})_{m \times n}$ and $a_{ij} \in (0, 50)$ are generated randomly. In what follows, we denote $e_1 = (1, 1, \cdots, 1)^T$.

Table 3 Number of iterations for Algorithm 3.8 being applied to Example 3

m	n	Starting points	Number of iterations
10	9	$(0, 0, 0)^T$	8
		$100e_1$	49
		rand(N,1)	95
50	55	$(0, 0, 0)^T$	11
		$100e_1$	36
		rand(N,1)	100

Table 4 Number of iterations for CQ Algorithm being applied to Example 1

Starting points	Number of iterations	Approximate solution
$(0, 0, 0)^T$	55	$(0.3333, 1.3333, -0.6667)^T$
$(1, 1, 1)^T$	56	$(0.3333, 1.3333, -0.6667)^T$
$(1, 2, 3)^T$	56	$(-0.1922, 1.5961, -0.4039)^T$
$(100, -10, 30)^T$	41	$(1.3846, 0.8077, -1.1923)^T$

Table 5 Number of iterations for CQ Algorithm being applied to Example 2

Starting points	Number of iterations	Approximate solution
$(0, 0, 0)^T$	0	$(0, 0, 0)^T$
$(1, 1, 1)^T$	503	$(1.0000, 0.8500, 0.5501)^T$
$(1, 2, 3)^T$	565	$(0.0202, 0.5621, 0.6460)^T$
$(10, -5, 10)^T$	584	$(6.0535, -9.6179, 4.0394)^T$

Table 6 Number of iterations for CQ Algorithm being applied to Example 3

m	n	Starting points	Number of iterations
10	9	$(0, 0, 0)^T$	204
		$100e_1$	692
		rand(N,1)	1591
50	55	$(0, 0, 0)^T$	114
		$100e_1$	2371
		rand(N,1)	7647

The test results for Algorithm 3.8 being applied to Example 3 are listed in Table 3 using different starting points.

Finally, we also compared the implementation of our algorithm with CQ algorithm of [2]. For a benchmark, we also tested the CQ algorithm on Examples 1, 2 and 3. The test results are listed in Tables 4, 5 and 6.

The results given in Tables 1, 2 and 3 are quite promising as these test problems were solved using just a small amount of iterations. They showed that the algorithms of

this paper is effective. Tables 1, 2, 3, 4, 5 and 6 tells us that the iteration numbers of our algorithm are relatively small. The numerical experiments demonstrate the viability of the new method proposed in this paper. We know that choice of parameter values affects performance of a method in practice. As this is not the main objective of the paper, we would not give detailed test in this aspect.

6 Concluding remarks

In this paper, we give two Newton projection methods, that is, Algorithms 3.1 and 3.8 for solving the split feasibility problem. As long as the solution set of the SFP is nonempty, the two algorithms will converges globally to a solution of the SFP proved by Theorem 3.7. While Theorem 4.5 and its corollaries prove that Algorithm 3.1 superlinearly converges to a regular solution. It is worth noting that if we do some feasible modification of Step 3 in Algorithm 3.1, the modified algorithm can solve a class of convex optimization problem, which is general than the SFP, in which the gradient of the objective function is nonsmooth, but locally Lipschitzian. Or, it can also be used to solve a class of nonsmooth and monotone $VIP(C, F)$, where, F is locally Lipschitzian. For the extended algorithm, we can also prove that the conclusion corresponding to Theorems 3.7 and 4.5 also hold.

References

1. Censor, Y., Elfving, T.: A multiprojection algorithm using Bregman projections in a product space. Numer. Algorithms **8**, 221–239 (1994)
2. Byrne, C.L.: Iterative oblique projection onto convex sets and the split feasibility problem. Inverse Probl. **18**, 441–453 (2002)
3. Yang, Q.Z.: The relaxed CQ algorithm solving the split feasibility problem. Inverse Probl. **20**, 1261–1266 (2004)
4. Qu, B., Xiu, N.H.: A new halfspace-relaxation projection method for the split feasibility problem. Linear Algebra Appl. **428**(5–6), 1218–1229 (2008)
5. Byrne, C.L.: Bregman–Legendre mulidistance projection algorithms for convex feasibility and optimization. In: Butnairu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, pp. 87–100. Elsevier, Amsterdam (2001)
6. Liu, B.H., Qu, B., Zheng, N.: A successive projection algorithm for solving the multiple-sets split feasibility problem. Numer. Funct. Anal. Optim. **35**, 1459–1466 (2014)
7. Qu, B., Xiu, N.H.: A note on the CQ algorithm for the split feasibility problem. Inverse Probl. **21**, 1655–1665 (2005)
8. Wang, C.Y., Xiu, N.H.: Convergence of the gradient projection method for generalized convex minimization. Comput. Optim. Appl. **16**, 111–120 (2000)
9. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
10. Byrne, C.L.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Probl. **20**, 103–120 (2004)
11. Xiu, N.H., Zhang, J.Z.: Some recent advances in projection-type methods for variational inequalities. J. Comput. Appl. Math. **152**, 559–585 (2003)

12. Ito, K.F., Kunisch, K.: On a semi-smooth Newton method and its globalization. Math. Program. **118**(2), 347–370 (2009)
13. Ito, K.F., Kunisch, K.: Applications of semi-smooth Newton methods to variational inequalities. Int. Ser. Numer. Math. **155**, 175–192 (2007)
14. Kanzow, C., Klug, A.: An interior-point affine scaling trust region method for semismooth equations with box constraints. Comput. Optim. Appl. **37**, 329–353 (2007)
15. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
16. Qi, H.D.: A regularized smoothing Newton method for box constrained variational inequality problems with P$_0$-functions. SIAM J. Optim. **10**, 315–330 (1999)
17. Sun, D., Womersley, R.S., Qi, H.D.: A feasible semismooth asymptocially Newton method for mixed complementarity problems. Math. Program. **94**, 167–187 (2002)
18. Huyer, W., Neumaier, A.: A new exact penalty function. SIAM J. Optim. **13**(4), 1141–1158 (2003)
19. Li, D.H., Fukushima, M.: Globally convergent Broyden-like methods for semismooth equations and applications to VIP, NCP and MCP. Ann. Oper. Res. **103**, 71–79 (2001)
20. Sun, D., Qi, L.: Solving variational inequality problems via smoothing-nonsmooth reformulations. J. Comput. Appl. Math. **129**, 37–62 (2001)
21. Wang, C.Y., Liu, Q., Ma, C.: Smoothing SQP algorithm for semismooth equations with box constraints. Comput. Optim. Appl. **55**, 399–425 (2013)
22. Zhou, Z.G., Yu, B.: A smoothing homotopy method for variational inequality problems on polyhedral convex sets. J. Glob. Optim. **58**, 151–168 (2014)
23. Armand, P., Benoist, J., Omheni, R., Pateloup, V.: Study of a primal-dual algorithm for equality constrained minimization. Comput. Optim. Appl. **59**, 405–433 (2014)
24. Kanzow, C., Fukushima, M.: Theoretical and numerical investgation of the D-gap function for box constrained variational inequalities. Math. Program. **83**, 55–87 (1998)
25. Liu, J.M.: Linear stability of generalized equation part I: basic theory. Math. Oper. Res. **3**, 706–720 (1994)
26. Liu, J.M.: Linear stability of generalized equation part II: applications to nonlinear programming. Math. Oper. Res. **3**, 721–742 (1994)
27. Liu, J.M.: Strong stability in variational inequalities. SIAM J. Control Optim. **33**, 725–749 (1995)
28. Peng, J.M., Fukushima, M.: A hybrid Newton method for solving the variational inequality problem via the D-gap function. Math. Program. **86**, 367–386 (1999)
29. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
30. Solodov, M.V., Svaiter, B.F.: A trully globally convergent Newton-type method for the monotone nonlinear complementarity problem. SIAM J. Optim. **10**(2), 605–625 (2000)
31. Sun, D., Fukushima, M., Qi, L.: A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems. In: Ferris, M.C., Pang, J.S. (eds.) Complementarity and Variational Problems: State of the Art, pp. 452–472. SIAM, Philadelphia (1997)
32. Taji, K., Fukushima, M., Ibaraki, T.: A globally convergent Newton method for solving strongly monotone variational inequalities. Math. Program. **58**, 369–383 (1993)
33. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
34. Hiriart-Urruty, J.B., Strodiot, J.J., Hien Nguyen, V.: Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. **11**, 43–56 (1984)
35. Mifflin, M.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 957–972 (1977)
36. Harker, P., Pang, J.S.: Finite-demensional variational inequalities and complementarity problems: a survey of theory, algorithms and applications. Math. Program. **48**(1), 161–220 (1990)
37. Qu, B., Liu, B.H., Zheng, N.: On the computation of the step-size for the CQ-like algorithms for the split feasibility problem. Appl. Math. Comput. **262**, 218–223 (2015)
38. Zhao, J.L., Zhang, Y.J., Yang, Q.Z.: Modified projection methods for the split feasibility problem and the multiple-sets split feasibility problem. Appl. Math. Comput. **219**, 1644–1653 (2012)