A new proximal gradient algorithm for solving mixed variational inequality problems with a novel explicit stepsize and applications

Pham Thi Hoai[a]

[a]*Faculty of Mathematics and Informatics, Hanoi University of Science and Technology, 1 Dai Co Viet Road, Hanoi, Vietnam*

Abstract

In this paper, we propose a new algorithm for solving monotone mixed variational inequality problems in real Hilbert spaces based on proximal gradient method. Our new algorithm uses a novel explicit stepsize which is proved to be increasing to a positive limitation. This property plays an important role in improving the speed of the algorithm. To the best of our knowledge, it is the first time such a kind of stepsize has been proposed for the proximal gradient method solving mixed variational inequality problems. We prove the weak convergence and strong convergence with R-linear rate of our new algorithm under standard assumptions. The reported numerical simulations for applications in sparse logistic regression and image deblurring reveal the significant efficacy performance of our proposed method compared to the recent ones.

Keywords: proximal gradient method, mixed variational inequality, composite optimization model, sparse logistic regression, image deblurring problem
2010 MSC: 49J40, 47H04, 47H10

1. Introduction

The mixed variational inequality problems (MVIP) are known as generalizations of many optimization problems arisen from nonlinear programming and variational analysis such as minimization problems, linear complementary problems or variational inequalities. Its applications can be found in a variety of fields such as data science, image processing, mechanics, control, economics, structural engineering, social sciences, etc. See, e.g., [1, 4, 5, 6, 7, 9, 11, 13, 14, 15, 19, 20, 21, 28] and the references therein. The formulation of considered MVIP can be described as follows: Let $\mathcal{H}$ be a real Hilbert space endowed with the inner product $\langle ., . \rangle$ and the associated norm $\|.\|$; $A : \mathcal{H} \to \mathcal{H}$ be a single-valued mapping; $g : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ a proper, convex and lower semi-continuous function with $\mathrm{dom}(g) = \{x \in \mathcal{H} : g(x) < +\infty\}$. MVIP aims to find a point $x^* \in \mathrm{dom}(g)$ such that

$$\langle Ax^*, y - x^* \rangle + g(y) - g(x^*) \geq 0, \quad \forall y \in \mathrm{dom}(g). \tag{1.1}$$

By $\mathrm{Sol}(g, A)$ we denote the solution set of Problem (1.1), i.e.,

$$\mathrm{Sol}(g, A) = \{x \in \mathrm{dom}(g) : \langle Ax, y - x \rangle + g(y) - g(x) \geq 0, \quad \forall y \in \mathrm{dom}(g)\}.$$

Email address: hoai.phamthi@hust.edu.vn (Pham Thi Hoai)

It is known that MVIP has been extensively studied by many authors for theoretical as well as algorithmic aspects. One of the traditional approaches for finding a member of $\mathrm{Sol}(g, A)$ is the projection-type method. The advantage of this one is that these methods are specifically well-adapted to the situation as A is not Lipschitz continuous or no estimation of the Lipschitz constant is available. However, projection methods usually require line-search procedures. One has to compute a resolvent of ∂g together with one projection onto some half-space at each iteration. This costs expensive processing time in general. Moreover, the line-search stepsize often decreases to zeros that can makes the speed of the algorithm slow at large iterations. One can see [12, 17, 22, 23, 24, 25] and the references therein for more details.

Recently, gradient proximal-type methods have been actively studied in [19, 20, 29] with efficient computations. In Malitsky [19], the author proposed a proximal extrapolated gradient method (PEGM) with line-search procedure. It is worth pointing out that, line-search procedure over iterations probably increases the computation time of the method. However, as mentioned in [19] this kind of stepsize has a special property that "possibility at least theoretically to increase the stepsize from iteration to iteration" [19]. This advantage point therefore may overcome the main drawback of line-search method (usually giving the decreasing sequence of stepsize). After that, in 2019, the author continued to improve the algorithm to avoid the expensive calculation of the line-search process of PEGM. In particular, Malitsky [20] proposed another strategy of stepsize selection that can be easily computed by a simple closed formulation while still keeping the ability to increase the sequence of stepsize. This algorithm is called an adaptive golden ratio algorithm (aGRAAL). It is noticeable that the convergence of both PEGM and aGRAAL are obtained under the local Lipschitzness of A and the MVIP is considered in a finite dimensional space. For MVIP in general Hilbert spaces, very recently, Zhou et al. [29] introduced an explicit algorithm for solving (1.1). The self-adaptive stepsizes designed in [29] are decreasing and converging to a positive lower bound. To obtain the convergence of their proposed algorithm, the operator A needs to be globally Lipschitz. However, one know that the descent of the stepsize may take a disadvantage because of the possibility of slowing down the algorithm at the last steps. Additionally, the explicit stepsize by Zhou et al. [29] may need the effort to compute ρ_n that involves three norms in $\mathcal{H}$ and it costs not cheap if $\mathcal{H}$ is just finite but large dimensional.

In this paper, we introduce an efficient gradient proximal algorithm which uses our new adaptive stepsizes for solving the MVIP (1.1). Our stepsize can be computed easily by a simple formula and no need any line-search computation. Moreover, the most prominent feature of our new method is in the increasing property of the sequence of stepsize after a finite number of iterations. To the best of our knowledge, it is the first time such a kind of stepsize has been proposed. This provides a promising method which is able to overcome the main drawbacks of the other ones. The weak convergence of our algorithm is obtained for the monotone and globally Lipschitz A. If A satisfies the strong monotone condition we obtain the R-linear convergence result. The numerical experiments for the two applications in sparse logistic regression and image deblurring show the efficiency of our new method.

The rest of the paper is organized as follows. After collecting some definitions and fundamental results in Section 2, we propose the new strategy of stepsize selection and the convergence results in Section 3. Section 4 includes several numerical results to illustrate the performance of our algorithm in comparison to the other ones. The paper is closed with some conclusions in Section 5.

2. Preliminaries

For a sequence $\{x^k\} \subset \mathcal{H}$, we denote strong convergence of $\{x^k\}$ to $x \in \mathcal{H}$ by $x^k \to x$ and weak convergence by $x^k \rightharpoonup x$. With a proper, convex and lower semicontinuous function $g : \mathcal{H} \to (-\infty, +\infty]$ we see that for each $x \in \mathcal{H}$, the function

$$y \mapsto g(y) + \frac{1}{2}\|y - x\|^2$$

is proper, strongly convex and lower semicontinuous, thus the infimum is attained. The unique minimum of

$$y \mapsto g(y) + \frac{1}{2}\|y - x\|^2 \tag{2.1}$$

is called proximal point of g at x and it is denoted by $\text{prox}_g(x)$. Therefore the operator

$$\begin{aligned} \text{prox}_g(x) &: \mathcal{H} \to \mathcal{H} \\ x &\mapsto \underset{y\in\mathcal{H}}{\text{argmin}}\left\{ g(y) + \frac{1}{2}\|y - x\|^2 \right\} \end{aligned}$$

is well-defined and is said to be the proximity operator of g. When $g = \iota_C$ (the indicator function of the convex set $C \subset \mathcal{H}$), one has

$$\text{prox}_{\iota_C}(x) = P_C(x) = \underset{y\in C}{\text{argmin}}\left\{ \frac{1}{2}\|y - x\|^2 \right\}, \ \forall x \in \mathcal{H}.$$

The next lemmas are essential for our analysis in the sequel.

Lemma 2.1. ([2, Proposition 12.26]) *Let $g : \mathcal{H} \to (-\infty, +\infty]$ be a convex function, $x \in \mathcal{H}$. Then $p = \text{prox}_g(x)$ if and only if*

$$\langle p - x, y - p\rangle \geq g(p) - g(y) \ \ \forall y \in \text{dom}(g).$$

The result in the below lemma is easy to obtain from Lemma 2.1.

Lemma 2.2. *Let $\lambda > 0$ and g be proper l.s.c. convex function. Then x^* is a solution of (1.1) if and only if*

$$x^* = \text{prox}_{\lambda g}(x^* - \lambda A x^*).$$

Lemma 2.3. ([20, Lemma 2.1]) *Let $\{a_k\}$, $\{b_k\}$ be two sequences of nonnegative numbers fulfilling:*

$$a_{k+1} \leq a_k - b_k, \ \forall k \in \mathbb{N}. \tag{2.2}$$

Then $\{a_k\}$ is convergent and $\sum_{k=0}^{+\infty} b_k < +\infty$.

Lemma 2.4. ([2, Lemma 2.39]) *Let $\{x^k\}$ be a sequence in a real Hilbert space and let S be a nonempty subset of $\mathcal{H}$. Suppose that*

(1) *For every $z \in S$, $\lim_{k\to\infty} \|x^k - z\|$ exists;*

(2) *Any weak cluster point of $\{x^k\}$ belongs to S.*

Then, there exists $\bar{x} \in S$ such that $\{x^k\}$ converges weakly to $\bar{x}$.

3. A novel explicit stepsize for proximal gradient method solving MVIP

From now on, we assume that the following conditions hold:

(A1) $\mathrm{Sol}(g, A) \neq \emptyset$;

(A2) The mapping $A : \mathrm{dom}\, g \to \mathcal{H}$ is globally Lipschitz with Lipschitz constant $L > 0$ and A is monotone;

(A3) $g : \mathcal{H} \to (-\infty, +\infty]$ is proper, convex and lower semicontinuous function.

We now introduce our new proximal gradient algorithm (NPROX) solving MVIP under the three conditions (A1), (A2) and (A3).

Algorithm 3.1 (NPROX)

Step 0 (Initialization). Let $r \in (1, 2), \rho = \frac{1+\sqrt{1+4r}}{2r}$; $\lambda_0 > 0$, $0 < \eta_1 < \eta_0 < \sigma < \overline{\sigma} < \frac{\rho}{2}$ and a positive real sequence $\{\xi_k\}$ satisfying

$$\sum_{k=0}^{+\infty} \xi_k < +\infty. \tag{3.1}$$

Choose $x^0 = y^0 \in \mathcal{H}, x^1 \in \mathcal{H}$ and set $k = 1$.

Step 1. If

$$\|Ax^k - Ax^{k-1}\| > \frac{\eta_0}{\lambda_{k-1}} \|x^k - x^{k-1}\| \tag{3.2}$$

then

$$\lambda_k = \eta_1 \frac{\|x^k - x^{k-1}\|}{\|Ax^k - Ax^{k-1}\|} \tag{3.3}$$

else

$$\lambda_k = (1 + \xi_{k-1})\lambda_{k-1}. \tag{3.4}$$

Step 2. compute

$$y^k = \frac{(\rho - 1)x^k + y^{k-1}}{\rho}, \tag{3.5}$$

$$x^{k+1} = \operatorname{prox}_{\lambda_k g}(y^k - \lambda_k Ax^k). \tag{3.6}$$

Set $k := k + 1$, and return to **Step 1**.

Thanks to the idea of [16, 9] our proposed stepsize in Algorithm 3.1 is proved to have interesting properties as in Lemma 3.1 below.

Lemma 3.1. *Let $\{\lambda_k\}$ be the stepsizes sequence generated by Algorithm 3.1. Then*

(i) *for all* $k \geq 1$ *we have* $\lambda_k \geq \lambda_{min} := \min\left\{\frac{\eta_1}{L}, \lambda_0\right\} > 0$;

(ii) $\{\lambda_k\}$ *is convergent;*

(iii) *there exists a positive integer* $\overline{k}$ *such that* $\lambda_{k+1} \geq \lambda_k$ *for all* $k \geq \overline{k}$.

Proof. (i) Since A is $L-$ globally Lipschitz continuous then

$$\|Ax^k - Ax^{k-1}\| \leq L\|x^k - x^{k-1}\| \;\; \forall k \geq 0,$$

therefore $\lambda_1 \geq \min\left\{\frac{\eta_1}{L}, \lambda_0\right\}$. By induction, we obtain that the sequence $\{\lambda_k\}$ is lower bounded by $\min\left\{\frac{\eta_1}{L}, \lambda_0\right\}$.

(ii) Let $u_k = \ln \lambda_{k+1} - \ln \lambda_k \quad \forall k \geq 0$, we have $u_k = u_k^+ - u_k^-$, where

$$u_k^+ = \max\{0, u_k\}, u_k^- = -\min\{0, u_k\}.$$

Then $u_k^+ \geq 0 \quad \text{and} \quad u_k^- \geq 0 \quad \forall k \geq 0.$
From the definition of λ_k in Algorithm 3.1, we derive that

$$u_k = \ln \frac{\lambda_{k+1}}{\lambda_k} \leq \ln(1+\xi_k) \leq \xi_k \quad \forall k \geq 0,$$

which implies $u_k^+ \leq \xi_k$. Since $\sum\limits_{k=0}^{+\infty} \xi_k$ is convergent, we obtain $\sum\limits_{k=0}^{+\infty} u_k^+ < +\infty$. Observing that $\sum\limits_{k=0}^{+\infty} u_k^-$ is a nonnegative series and using the following relation

$$\ln \lambda_{k+1} - \ln \lambda_0 = \sum_{i=0}^{k} u_i = \sum_{i=0}^{k}(u_i^+ - u_i^-) = \sum_{i=0}^{k} u_i^+ - \sum_{i=0}^{k} u_i^-, \tag{3.7}$$

we see that if $\lim\limits_{k\to+\infty} \sum\limits_{i=0}^{k} u_i^- = +\infty$ then

$$\lim_{k\to+\infty} (\ln \lambda_{k+1}) = -\infty \iff \lim_{k\to+\infty} \lambda_k = 0.$$

This contradicts (i) and we have the convergence of $\sum\limits_{k=0}^{+\infty} u_k^-$. Finally, from (3.7) we get the desired conclusion that $\lim\limits_{k\to+\infty} \lambda_k = \lambda^* < +\infty$.

(iii) The third assertion is equivalent to showing that there exists $\overline{k}$ such that

$$\|Ax^k - Ax^{k-1}\| \leq \frac{\eta_0}{\lambda_{k-1}}\|x^k - x^{k-1}\| \;\; \forall k \geq \overline{k}.$$

Suppose by contradiction that there exists $\{k_j\}, k_j \to +\infty$ such that

$$||Ax^{k_j} - Ax^{k_j-1}|| > \frac{\eta_0}{\lambda_{k_j-1}}||x^{k_j} - x^{k_j-1}||.$$

For this case

$$\lambda_{k_j} = \eta_1 \frac{||x^{k_j} - x^{k_j-1}||}{\|Ax^{k_j} - Ax^{k_j-1}\|}$$

Therefore

$$\frac{\eta_1||x^{k_j} - x^{k_j-1}||}{\lambda_{k_j}} = \|Ax^{k_j} - Ax^{k_j-1}\| > \frac{\eta_0}{\lambda_{k_j-1}}||x^{k_j} - x^{k_j-1}||,$$

i.e.,

$$\frac{\lambda_{k_j}}{\lambda_{k_j-1}} < \frac{\eta_1}{\eta_0} \quad \forall k_j.$$

From (ii), we have

$$\lim_{k_j\to+\infty} \lambda_{k_j} = \lim_{k_j\to+\infty} \lambda_{k_j-1} = \lim_{k\to+\infty} \lambda_k = \lambda^* \tag{3.8}$$

hence we deduce that

$$\frac{\lambda^*}{\lambda^*} \le \frac{\eta_1}{\eta_0} < 1.$$

It is a contradiction and we finish the proof.

□

Remark 3.2. From the definition of ρ in Algorithm 3.1 we easily obtain the following relation

(i) $\rho > 1$;

(ii) $1 + \frac{1}{\rho} - r\rho = 0$.

Next, we study the convergence of Algorithm 3.1. Firstly, we claim the weak convergence of $\{x^k\}$ generated by Algorithm 3.1 in the following theorem.

Theorem 3.3. *Under conditions (A1)-(A3), the sequence $\{x^k\}$ generated by Algorithm 3.1 converges weakly to a solution of the problem (1.1).*

Proof. Let z be a member of $\mathrm{Sol}(g, A)$. From (3.6) we have

$$x^{k+1} = \mathrm{prox}_g\left(y^k - \lambda_k Ax^k\right).$$

Applying Lemma 2.1 derives

$$\left\langle x^{k+1} - y^k + \lambda_k Ax^k, y - x^{k+1}\right\rangle \ge \lambda_k\left(g\left(x^{k+1}\right) - g(y)\right), \quad \forall y \in \mathrm{dom}(g). \tag{3.9}$$

Replace y by z we have that

$$\left\langle x^{k+1} - y^k + \lambda_k Ax^k, z - x^{k+1}\right\rangle \ge \lambda_k\left(g\left(x^{k+1}\right) - g\left(z\right)\right).$$

Hence, we get

$$2\left\langle x^{k+1} - y^k, z - x^{k+1}\right\rangle + 2\lambda_k\left\langle Ax^k, z - x^{k+1}\right\rangle \ge 2\lambda_k\left(g\left(x^{k+1}\right) - g\left(z\right)\right).$$

It follows that

$$\begin{aligned}\left\|x^{k+1}-z\right\|^2 &\leq \left\|y^k-z\right\|^2-\left\|x^{k+1}-y^k\right\|^2+2\lambda_k\left\langle Ax^k, z-x^{k+1}\right\rangle+2\lambda_k\left(g\left(z\right)-g\left(x^{k+1}\right)\right)\\ &=\left\|y^k-z\right\|^2-\left\|x^{k+1}-y^k\right\|^2+2\lambda_k\left\langle Ax^k, z-x^{k}\right\rangle+2\lambda_k\left\langle Ax^k, x^k-x^{k+1}\right\rangle\\ &\quad+2\lambda_k\left(g\left(z\right)-g\left(x^{k+1}\right)\right)\\ &\leq \left\|y^k-z\right\|^2-\left\|x^{k+1}-y^k\right\|^2+2\lambda_k\left\langle Ax^k, x^k-x^{k+1}\right\rangle\\ &\quad-2\lambda_k\left(\left\langle Ax^k, x^k-z\right\rangle+g\left(x^{k+1}\right)-g(z)\right).\end{aligned} \tag{3.10}$$

Now, in view of (3.9) we subtitue k by $k-1$ to have

$$\left\langle x^k-y^{k-1}+\lambda_{k-1}Ax^{k-1}, y-x^k\right\rangle \geq \lambda_{k-1}\left(g\left(x^k\right)-g(y)\right),\quad \forall y\in \operatorname{dom}(g). \tag{3.11}$$

Replacing y by x^{k+1} in (3.11), we get the evaluation

$$\left\langle x^k-y^{k-1}+\lambda_{k-1}Ax^{k-1}, x^{k+1}-x^k\right\rangle \geq \lambda_{k-1}\left(g\left(x^k\right)-g\left(x^{k+1}\right)\right). \tag{3.12}$$

Multiplying both sides of (3.12) by $\frac{2\lambda_k}{\lambda_{k-1}} > 0$ and taking into account that $x^k-y^{k-1} = \rho\left(x^k-y^k\right)$, we arrive at

$$2\frac{\rho\lambda_k}{\lambda_{k-1}}\left\langle x^k-y^k, x^{k+1}-x^k\right\rangle+2\lambda_k\left\langle Ax^{k-1}, x^{k+1}-x^k\right\rangle \geq 2\lambda_k\left(g\left(x^k\right)-g\left(x^{k+1}\right)\right).$$

Therefore we continue getting the following estimate

$$\begin{aligned}0 &\leq \frac{\rho\lambda_k}{\lambda_{k-1}}\left(||x^{k+1}-y^k\|^2-\|x^k-y^k\|^2-\|x^{k+1}-x^k\|^2\right)\\ &\quad+2\lambda_k\left\langle Ax^{k-1}, x^{k+1}-x^k\right\rangle+2\lambda_k\left(g\left(x^{k+1}\right)-g\left(x^k\right)\right).\end{aligned} \tag{3.13}$$

Adding both sides of relations (3.10) and (3.13), we see that

$$\begin{aligned}\left\|x^{k+1}-z\right\|^2 &\leq \left\|y^k-z\right\|^2-\left(1-\frac{\rho\lambda_k}{\lambda_{k-1}}\right)\left\|x^{k+1}-y^k\right\|^2-\frac{\rho\lambda_k}{\lambda_{k-1}}\left(\left\|x^k-y^k\right\|^2+\left\|x^{k+1}-x^k\right\|^2\right)\\ &\quad+2\lambda_k\left\langle Ax^k-Ax^{k-1}, x^k-x^{k+1}\right\rangle-2\lambda_k\left(\left\langle Ax^k, x^k-z\right\rangle+g\left(x^k\right)-g\left(z\right)\right).\end{aligned} \tag{3.14}$$

By the monotone of A and $z\in \operatorname{Sol}(g, A)$, we obtain

$$\left\langle Ax^k, x^k-z\right\rangle+g\left(x^k\right)-g\left(z\right) \geq \left\langle Az, x^k-z\right\rangle+g\left(x^k\right)-g\left(z\right) \geq 0. \tag{3.15}$$

Thus, we arrive at

$$\begin{aligned}\left\|x^{k+1}-z\right\|^2 &\leq \left\|y^k-z\right\|^2-\left(1-\frac{\rho\lambda_k}{\lambda_{k-1}}\right)\left\|x^{k+1}-y^k\right\|^2\\ &\quad-\frac{\rho\lambda_k}{\lambda_{k-1}}\left(\left\|x^k-y^k\right\|^2+\left\|x^{k+1}-x^k\right\|^2\right)+2\lambda_k\left\langle Ax^k-Ax^{k-1}, x^k-x^{k+1}\right\rangle.\end{aligned} \tag{3.16}$$

Now, since the positive series $\sum_{k=0}^{+\infty} \xi_k$ converges and $0 < \eta_0 < \sigma, 1 < r < 2$ then one can take $k_1 \in \mathbb{N}$ such that

$$\xi_{k-1} < \min\left\{r-1, \frac{\sigma}{\eta_0} - 1\right\} \ \forall k \geq k_1. \tag{3.17}$$

The way choosing k_1 like that helps us to show the correctness of the following inequality

$$2\lambda_k \left\langle Ax^k - Ax^{k-1}, x^k - x^{k+1}\right\rangle \leq \sigma(\|x^k - x^{k-1}\|^2 + \|x^{k+1} - x^k\|^2), \forall k \geq k_1. \tag{3.18}$$

Indeed, from Algorithm 3.1, if condition (3.2) is true then $\lambda_k = \eta_1 \frac{\|x^k - x^{k-1}\|}{\|Ax^k - Ax^{k-1}\|}$ and we have (3.18) easily by using Cauchy–Schwarz inequality as follows

$$\begin{aligned} 2\lambda_k \left\langle Ax^k - Ax^{k-1}, x^k - x^{k+1}\right\rangle &\leq 2\lambda_k \|Ax^k - Ax^{k-1}\|\|x^k - x^{k+1}\| \\ &= 2\eta_1 \|x^k - x^{k-1}\|\|x^k - x^{k+1}\| \\ &< \sigma(\|x^k - x^{k-1}\|^2 + \|x^{k+1} - x^k\|^2). \end{aligned} \tag{3.19}$$

Otherwise, we have $\|Ax^k - Ax^{k-1}\| \leq \frac{\eta_0}{\lambda_{k-1}}\|x^k - x^{k-1}\|$ and

$$\begin{aligned} 2\lambda_k \left\langle Ax^k - Ax^{k-1}, x^k - x^{k+1}\right\rangle &\leq 2(1+\xi_{k-1})\lambda_{k-1}\|Ax^k - Ax^{k-1}\|\|x^k - x^{k+1}\| \\ &\leq 2(1+\xi_{k-1})\eta_0\|x^k - x^{k-1}\|\|x^k - x^{k+1}\| \\ &\leq (1+\xi_{k-1})\eta_0(\|x^k - x^{k-1}\|^2 + \|x^{k+1} - x^k\|^2). \end{aligned} \tag{3.20}$$

From (3.17) and (3.20) we obtain (3.18).
We continue to confirm that

$$1 + \frac{1}{\rho} - \frac{\rho\lambda_k}{\lambda_{k-1}} \geq 0 \ \forall k \geq k_1. \tag{3.21}$$

Indeed, by Remark 3.2 we have $1 + \frac{1}{\rho} - r\rho = 0$. If condition (3.2) is right then it is obvious to see that $\frac{\lambda_k}{\lambda_{k-1}} < \frac{\eta_1}{\eta_0} < 1 < r$ and

$$1 + \frac{1}{\rho} - \frac{\rho\lambda_k}{\lambda_{k-1}} > 1 + \frac{1}{\rho} - r\rho = 0.$$

The remaining case $\lambda_k = (1+\xi_{k-1})\lambda_{k-1}$, from (3.17) and (3.4) we deduce that

$$1 + \frac{1}{\rho} - \frac{\rho\lambda_k}{\lambda_{k-1}} = 1 + \frac{1}{\rho} - (1+\xi_{k-1})\rho > 1 + \frac{1}{\rho} - r\rho = 0, \ \forall k \geq k_1.$$

Now, from Lemma 3.1, $\{\lambda_k\}$ converges to λ^* then we derive that

$$\lim_{k\to+\infty} \frac{\rho\lambda_k}{\lambda_{k-1}} = \rho > 2\overline{\sigma} \left(\text{because } \overline{\sigma} < \frac{\rho}{2}\right).$$

Therefore there exists $k_2 \geq 1$ such that

$$\frac{\rho\lambda_k}{\lambda_{k-1}} > 2\overline{\sigma}, \ \forall k \geq k_2. \tag{3.22}$$

Taking into account that, in a real Hilbert space $\mathcal{H}$, the equality

$$\|ax + by\|^2 = a\|x\|^2 + b\|y\|^2 - ab\|x - y\|^2 \tag{3.23}$$

keeps for all $x, y \in \mathcal{H}$ and $a + b = 1$, $a, b \in \mathbb{R}$. On the other hand,it can be easily seen from (3.5) that $\rho y^{k+1} = (\rho - 1)x^{k+1} + y^k$. Hence, applying (3.23) we obtain

$$\left\|x^{k+1} - z\right\|^2 = \frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 - \frac{1}{\rho - 1}\left\|y^k - z\right\|^2 + \frac{1}{\rho}\left\|x^{k+1} - y^k\right\|^2. \tag{3.24}$$

Combining (3.16), (3.18) with (3.24) and $\rho > 1$ (in Remark 3.2), we get

$$\begin{aligned}\frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 \leq \frac{\rho}{\rho - 1}\left\|y^k - z\right\|^2 - \left(1 + \frac{1}{\rho} - \frac{\rho\lambda_k}{\lambda_{k-1}}\right)\left\|x^{k+1} - y^k\right\|^2 - \frac{\rho\lambda_k}{\lambda_{k-1}}\left\|x^k - y^k\right\|^2 \\ + \sigma\left\|x^k - x^{k-1}\right\|^2 - \left(\frac{\rho\lambda_k}{\lambda_{k-1}} - \sigma\right)\left\|x^{k+1} - x^k\right\|^2, \ \forall k \geq k_1.\end{aligned} \tag{3.25}$$

From (3.21), (3.22) and (3.25) we obtain that

$$\begin{aligned}\frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 + (2\overline{\sigma} - \sigma)\left\|x^{k+1} - x^k\right\|^2 \leq \frac{\rho}{\rho - 1}\left\|y^k - z\right\|^2 + \\ + \sigma\left\|x^k - x^{k-1}\right\|^2 - 2\overline{\sigma}\left\|x^k - y^k\right\|^2, \ \forall k \geq k_0 = \max\{k_1, k_2\}.\end{aligned} \tag{3.26}$$

Since $\sigma < \overline{\sigma}$, then $2\overline{\sigma} - \sigma > \sigma$ and we get from (3.26) that

$$\begin{aligned}\frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 + \sigma\left\|x^{k+1} - x^k\right\|^2 \leq \frac{\rho}{\rho - 1}\left\|y^k - z\right\|^2 + \\ + \sigma\left\|x^k - x^{k-1}\right\|^2 - 2\overline{\sigma}\left\|x^k - y^k\right\|^2, \ \forall k \geq k_0 = \max\{k_1, k_2\}\end{aligned} \tag{3.27}$$

Let

$$\begin{aligned}a_k &= \frac{\rho}{\rho - 1}\left\|y^k - z\right\|^2 + \sigma\left\|x^k - x^{k-1}\right\|^2, \\ b_k &= 2\overline{\sigma}\left\|x^k - y^k\right\|^2\end{aligned}$$

then

$$a_{k+1} \leq a_k - b_k, \ \forall k \geq k_0,$$

Thus, by Lemma 2.3, the limit $\lim_{k \to +\infty} a_k$ exists and is finite. In addition, $\lim_{k \to +\infty} b_k = 0$ follows the sequences $\{y^k\}$ and $\{x^k\}$ are bounded and

$$\lim_{k \to +\infty}\left\|x^k - y^k\right\|^2 = 0. \tag{3.28}$$

From the equality $x^k - y^{k-1} = \rho\left(x^k - y^k\right)$ one easily obtains $\lim_{k \to +\infty}\left\|x^k - y^{k-1}\right\|^2 = 0$. Therefore $\left\|x^{k+1} - y^k\right\|^2 \to 0$ as $k \to +\infty$. This together with (3.28) imply that

$$\lim_{k \to +\infty}\left\|x^{k+1} - x^k\right\|^2 = 0. \tag{3.29}$$

Let us show that any weakly cluster point of the sequence $\{y^k\}$ belongs to the solution set $\mathrm{Sol}(g, A)$. Indeed, let x^* be an arbitrary weakly cluster point of $\{y^k\}$. Since $\{y^k\}$ is bounded, there exists a subsequence $\{y^{k_l}\}$ of $\{y^k\}$ such that $y^{k_l} \rightharpoonup x^*$. By (3.28), we also have $x^{k_l} \rightharpoonup x^*$. Fix an arbitrary $y \in \mathrm{dom}(g)$, we have from (3.11) that

$$\left\langle x^{k_l} - y^{k_l-1} + \lambda_{k_l-1} A x^{k_l-1}, y - x^{k_l} \right\rangle \geq \lambda_{k_l-1} \left(g\left(x^{k_l}\right) - g(y) \right). \tag{3.30}$$

By using (3.30) we deduce that

$$\begin{aligned} \lambda_{k_l-1} \left(g\left(x^{k_l}\right) - g(y) \right) &\leq \left\langle x^{k_l} - y^{k_l-1}, y - x^{k_l} \right\rangle + \lambda_{k_l-1} \Big\langle A x^{k_l-1}, y - x^{k_l} \Big\rangle \\ &= \left\langle x^{k_l} - y^{k_l-1}, y - x^{k_l} \right\rangle + \lambda_{k_l-1} \Big\langle A x^{k_l-1}, y - x^{k_l-1} \Big\rangle \\ &\qquad + \lambda_{k_l-1} \Big\langle A x^{k_l-1}, x^{k_l-1} - x^{k_l} \Big\rangle \\ &\leq \left\langle x^{k_l} - y^{k_l-1}, y - x^{k_l} \right\rangle + \lambda_{k_l-1} \Big\langle A y, y - x^{k_l-1} \Big\rangle \\ &\qquad + \lambda_{k_l-1} \Big\langle A x^{k_l-1}, x^{k_l-1} - x^{k_l} \Big\rangle, \end{aligned} \tag{3.31}$$

where last inequality is implied by the monotone of A. Let $l \to +\infty$ in (3.31), using the fact $\lim_{l \to +\infty} \|x^{k_l} - x^{k_l-1}\| = \lim_{l \to +\infty} \|y^{k_l-1} - x^{k_l}\| = 0$, $\lim_{l \to +\infty} \lambda_{k_l-1} = \lambda^* > 0$, the bounded of $\{x^k\}$, A globally Lipschitz and g lower semicontinuous; we arrive at

$$\langle Ay, y - x^* \rangle + g(y) - g(x^*) \geq 0. \tag{3.32}$$

For any arbitrary $x \in \mathrm{dom}(g)$, let $y_t = tx + (1-t)x^*$, $t \in (0,1)$. Since $\mathrm{dom}(g)$ is convex, $y_t \in \mathrm{dom}(g)$. Putting $y = y_t$ in (3.32) we get

$$\langle Ay_t, y_t - x^* \rangle + g(y_t) - g(x^*) \geq 0.$$

Thus, by the convexity of g, we have

$$\langle Ay_t, x - x^* \rangle + g(x) - g(x^*) \geq 0. \tag{3.33}$$

Since A is Lipschitz continuous, taking limit as $t \to 0$ in the (3.33), we obtain

$$\langle Ax^*, x - x^* \rangle + g(x) - g(x^*) \geq 0.$$

On the other hand, $x \in \mathrm{dom}(g)$ is arbitrary then we have $x^* \in \mathrm{Sol}(g, A)$. Since x^* is an arbitrary weakly cluster point we can conclude that the set of all weakly cluster points of $\{y^k\}$ belongs to the solution set $\mathrm{Sol}(g, A)$. Taking into account the convergence of the sequence $\left\{ \frac{\rho}{\rho-1} \|y^k - z\|^2 + \sigma \|x^k - x^{k-1}\|^2 \right\}$ and (3.29), we infer that the sequence $\left\{ \|y^k - z\| \right\}$ is convergent. Hence, it follows from Lemma 2.4 that the sequence $\{y^k\}$ weakly converges to a solution of the problem (1.1). Again by (3.28), we also obtain x^k weakly converges to a solution of the problem (1.1). This finishes the proof of the theorem.

□

Remark 3.4. It is observed that if we set $r = 1$ in Algorithm 3.1 (NPROX) then $\rho = \frac{1+\sqrt{5}}{2}$ which is known as golden ratio. However, in our algorithm we cannot choose such a r to ensure the the convergence of NPROX. In the proof of Theorem 3.3, the argument related to formulas (3.17), we need the condition $r > 1$ to confirm the existence of k_1. In contrast, this golden ratio is used adaptively in aGRAAL ([20] Algorithm 1) which is described as $\phi \in (1, \varphi]$, where $\varphi = \frac{1+\sqrt{5}}{2}$. Therefore our algorithm is not a kind of golden ratio algorithm and it is different from aGRAAL.

We next deals with the convergence rate of Algorithm 3.1. In order to establish the R-linear convergence, we assume that the operator A is strongly monotone on $\operatorname{dom} g$ with modulus $\gamma > 0$, i.e., for any $x, y \in \operatorname{dom} g$

$$\langle Ax - Ay, x - y\rangle \geq \gamma \|x - y\|^2. \tag{3.34}$$

Theorem 3.5. *Assume that Problem (1.1) satifies the conditions (A1) - (A3) and A is strongly monotone on $\operatorname{dom} g$ with modulus $\gamma > 0$. Then the sequence $\{x^k\}$ generated by Algorithm 3.1 converges R- linearly to $x^* \in \operatorname{Sol}(g, A)$, i.e., there exists an integer number $\hat{k}$, and $D > 0$, $\delta \in (0, 1)$ such that $\|x^k - x^*\| \leq D\delta^k,\ \forall k \geq \hat{k}$.*

Proof. By using the similar argument of the proof of Theorem 3.3 we update the inequality (3.15) by

$$\left\langle Ax^k, x^k - z\right\rangle + g\left(x^k\right) - g\left(z\right) \geq \left\langle Az, x^k - z\right\rangle + g\left(x^k\right) - g\left(z\right) + \gamma\|x^k - z\|^2 \geq \gamma\|x^k - z\|^2 \tag{3.35}$$

and therefore the inequality (3.16) becomes

$$\begin{aligned}\left\|x^{k+1} - z\right\|^2 \leq \left\|y^k - z\right\|^2 - \left(1 - \frac{\rho\lambda_k}{\lambda_{k-1}}\right)\left\|x^{k+1} - y^k\right\|^2 - \frac{\rho\lambda_k}{\lambda_{k-1}}\left(\left\|x^k - y^k\right\|^2 + \left\|x^{k+1} - x^k\right\|^2\right) + \\ + 2\lambda_k\left\langle Ax^k - Ax^{k-1}, x^k - x^{k+1}\right\rangle - 2\lambda_k\gamma\|x^k - z\|^2.\end{aligned} \tag{3.36}$$

Consequently we get the following inequality from (3.26)

$$\begin{aligned}\frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 + (2\overline{\sigma} - \sigma)\left\|x^{k+1} - x^k\right\|^2 \leq \frac{\rho}{\rho - 1}\left\|y^k - z\right\|^2 + \sigma\left\|x^k - x^{k-1}\right\|^2 \\ - 2\overline{\sigma}\left\|x^k - y^k\right\|^2 - 2\lambda_k\gamma\|x^k - z\|^2,\ \forall k \geq k_0.\end{aligned} \tag{3.37}$$

Next, because of the global Lipschitzness and the strong monotone of A we derive that $0 < \gamma \leq L$. Therefore

$$0 < 2\gamma\lambda_{min} \leq 2L\lambda_{min} \leq L\frac{\eta_1}{L} = 2\eta_1 < 2\overline{\sigma}\ \forall k \geq 0.$$

As a result we obtain that

$$2\overline{\sigma}\left\|x^k - y^k\right\|^2 + 2\lambda_k\gamma\|x^k - z\|^2 \geq 2\lambda_{min}\gamma\left(\|x^k - y^k\|^2 + \|x^k - z\|^2\right) \geq \lambda_{min}\gamma\|y^k - z\|^2. \tag{3.38}$$

Thus, from (3.37) we arrive at

$$\begin{aligned}\frac{\rho}{\rho - 1}\left\|y^{k+1} - z\right\|^2 + (2\overline{\sigma} - \sigma)\left\|x^{k+1} - x^k\right\|^2 \leq \left(\frac{\rho}{\rho - 1} - \gamma\lambda_{min}\right)\left\|y^k - z\right\|^2 \\ + \sigma\left\|x^k - x^{k-1}\right\|^2,\ \forall k \geq k_0,\end{aligned} \tag{3.39}$$

that is equivalent to

$$\left\|y^{k+1}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k+1}-x^k\right\|^2\leq\left(1-\gamma\lambda_{min}\frac{\rho-1}{\rho}\right)\left\|y^k-z\right\|^2+ + \sigma\frac{\rho-1}{\rho}\left\|x^k-x^{k-1}\right\|^2,\ \forall k\geq k_0. \tag{3.40}$$

Now, taking $\Gamma=\max\left\{1-\gamma\lambda_{min}\frac{\rho-1}{\rho},\frac{\sigma}{2\overline{\sigma}-\sigma}\right\}$ then $\Gamma\in(0,1)$ and (3.40) follows

$$\left\|y^{k+1}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k+1}-x^k\right\|^2\leq\Gamma\left(\left\|y^k-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^k-x^{k-1}\right\|^2\right),\ \forall k\geq k_0. \tag{3.41}$$

It follows that

$$\left\|y^{k+1}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k+1}-x^k\right\|^2\leq \leq\Gamma^{k-k_0+1}\left(\left\|y^{k_0}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k_0}-x^{k_0-1}\right\|^2\right),\ \forall k\geq k_0. \tag{3.42}$$

Setting

$$M=\frac{\left\|y^{k_0}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k_0}-x^{k_0-1}\right\|^2}{\Gamma^{k_0-1}},$$

then (3.42) can be rewrite as

$$\left\|y^{k+1}-z\right\|^2+(2\overline{\sigma}-\sigma)\frac{\rho-1}{\rho}\left\|x^{k+1}-x^k\right\|^2\leq M\Gamma^k,\ \forall k\geq k_0. \tag{3.43}$$

Thus

$$\left\|x^{k+1}-x^k\right\|\leq K\delta^k,\ \forall k\geq k_0, \tag{3.44}$$

where

$$K=\frac{M\rho}{(2\overline{\sigma}-\sigma)(\rho-1)},\quad \delta=\sqrt{\Gamma}\in(0,1).$$

Now taking $n_2>n_1\geq k_0$ arbitrary and using (3.44) we derive that

$$\|x^{n_2}-x^{n_1}\|\leq\sum_{k=n_1}^{n_2-1}\left\|x^{k+1}-x^k\right\|\leq\frac{K\delta^{n_1}}{1-\delta}. \tag{3.45}$$

Therefore $\{x^k\}$ is a Cauchy sequence and then has a limitation x^*. By tending n_2 to infinity in (3.45) we have

$$\|x^{n_1}-x^*\|\leq\frac{K\delta^{n_1}}{1-\delta},\ \ \forall n_1\geq k_0. \tag{3.46}$$

Combining the result of Theorem 1 we conclude that $\{x^k\}$ R-linearly converges to a unique solution x^* of problem (1.1) with $D=\frac{K}{1-\delta}$ and $\hat{k}=k_0$. □

4. Numerical results

In this section, we consider the following problem that has a lot of applications in many areas

$$\min_{x\in\mathbb{R}^n} f(x)+g(x), \tag{4.1}$$

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a smooth convex function, $\mathrm{dom}(f)$ is convex, f is globally Lipschitz gradient over $\mathrm{int}(\mathrm{dom}(f))$; $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is proper convex, lower semincontinuous and nonsmooth in general, $\mathrm{dom}(g)\subset\mathrm{int}(\mathrm{dom}(f))$. It is worth noting that (4.1) is called a composite optimization model that can be rewritten as the following inclusion problem:

$$\text{finding } x^*\in\mathbb{R}^n \text{ such that } 0\in(\nabla f+\partial g)(x^*). \tag{4.2}$$

Problem (4.2) is indeed (2.1) with $Ax=\nabla f(x)$ and the function $g(x)$ keeps the same. To investigate the performance of our new proposed algorithm, we implement NPROX (Algorithm 3.1) and the recent algorithms proposed for solving MVIP. Which include Algorithm 3 in [19] (PEGM), Algorithm 1 in [20] (aGRAAL) and Algorithm 4 in [29] (Zhou et al. Algo.). The two concrete applications of (4.1) consisting of sparse logistic regression and image deblurring are used for testing. We run all the programs on a PC Core i7 1360P 2.2GHz, Ram 16Gb (onboard) LPDDR5x 7500.

To execute NPROX we use the positive convergent series $\sum_{k=0}^{+\infty}\xi_k$ as follows

$$\xi_{k-1}=\frac{\alpha(\log(t.k))^\beta}{k^s},\ \ \alpha,\beta,t>0; s>1,\forall k\geq 1.$$

4.1. Sparse logistic regression

In this section, by using Python 3.9 we coded the mentioned algorithms for sparse logistic regression - one of popular problems in machine learning. We aim to find $x\in\mathbb{R}^n$ such that minimizing the loss function given by

$$h(x)=\sum_{i=1}^N \log\left(1+\exp\left(-b_i\left\langle a_i,x\right\rangle\right)\right)+\overline{\gamma}\|x\|_1,$$

where $a^i\in\mathbb{R}^n, b_i\in\{-1,1\}, i=1,...,N$ are observations; and $\overline{\gamma}>0$ is a regularization parameter. It is clear that problem $\min_{x\in\mathbb{R}^n} h(x)$ is a kind of problem (4.1) with

$$f(x)=\sum_{i=1}^N \log\left(1+\exp\left(-b_i\left\langle a_i,x\right\rangle\right)\right); \quad g(x)=\overline{\gamma}\|x\|_1.$$

The tested data are taken from LIBSVM[1] library including: $a9a$; $kdda.t.bz2$; $real-sim.bz2$; $ijcnn1.bz2$. We choose $x^0=\mathrm{zeros}(n)$, $x^1=x^0+10^{-9}\mathrm{random}(n)$, $\overline{\gamma}=0.005\|B^Tb\|_\infty$ (B composed by $a^i, i=1,...,N$, $b=(b_i)_{i=1}^N$) for all algorithms. The parameters[2] for each algorithm as follows:

[1] http://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets

[2] The parameters for PEGM, aGRAAL, Zhou et al. Algo. are chosen based on the corresponding reference [19, 20, 29].

PEGM: $a = 0.41\left(2 - \frac{1}{\theta}\right)$, $\theta = 2$, $\sigma = 0.7$, $\lambda_0 = a\frac{\|x^1 - x^0\|}{\|Ax^1 - Ax^0\|}$, $\lambda_{\max} = 10^7$; Err $= \|x_{n+1} - y_n\| + \|x_n - y_n\|$;

aGRAAL: $\varphi = 1.5$, $\lambda_0 = \frac{\varphi}{2}\frac{\|x^1 - x^0\|}{\|Ax^1 - Ax^0\|}$, $\rho = \frac{1}{\varphi} + \frac{1}{\varphi^2}$, $\bar{\lambda} = 10^7$, Err $= \|z^{k+1} - \bar{z}^k\| + \|\bar{z}^k - z^k\|$;

Zhou et al. Algo.: $\delta = 0.6$, $\mu_1 = 0.001$, $\epsilon = 5$, $\eta = 0.005$, Err $= \|u_{n+1} - u_n\| + \|u_n - v_n\|$;

NPROX: $\lambda_0 = 0.001$, $r = \frac{10}{9}$, $\rho = \frac{1+\sqrt{1+4r}}{2r}$, $\eta_0 = 0.2$, $\eta_1 = 0.15$, $\xi_{k-1} = \frac{0.9(\log k)^5}{k^{1.1}}$, $k \geq 1$, Err $= \|x^{k+1} - y^k\| + \|x^k - y^k\|$.

Each algorithm terminates either Err$< 10^{-10}$ or the number of iterations larger than N_{max}, where N_{max} is 10000 for $a9a$ and $real - sim.bz2$; 3000 for $kdda.t.bz2$ and 7000 for $ijcnn1.bz2$.

Denoting h_* as the mininmum of $h(x^k)$ obtained by all algorithms. The computational results are shown in Fig 1, 2, 3, 4 for the data sets $a9a$, $kdda.t.bz2$, $real - sim.bz2$, $ijcnn1.bz2$, respectively. The provided line graphs reveals the deviation of $h(x^k)$ and h_* with respect to the number of iterations and the running time. It is noticeable that almost figures experienced the best performance of NPROX for both of iterations and execution time in comparison with the remaining ones. It is worth noting that for $a9a$, $real - sim.bz2$, $ijcnn1.bz2$ in Fig 1, 3, 4 the number of iterations of PEGM is less than aGRAAL but PEGM takes larger running time. The reason is due to the line-search computation in each iteration of PEGM that makes it slower. Conversely, during the execution both of aGRAAL and NPROX use adaptive stepsize hence take little time so that the bigger iterations still give the faster running time compared to PEGM. This phenomenon is also stated in Fig. 2 for $kdda.t.bz2$ data set. Regarding the iteration, PEGM is better than NPROX but for the running time NPROX is clearly faster.

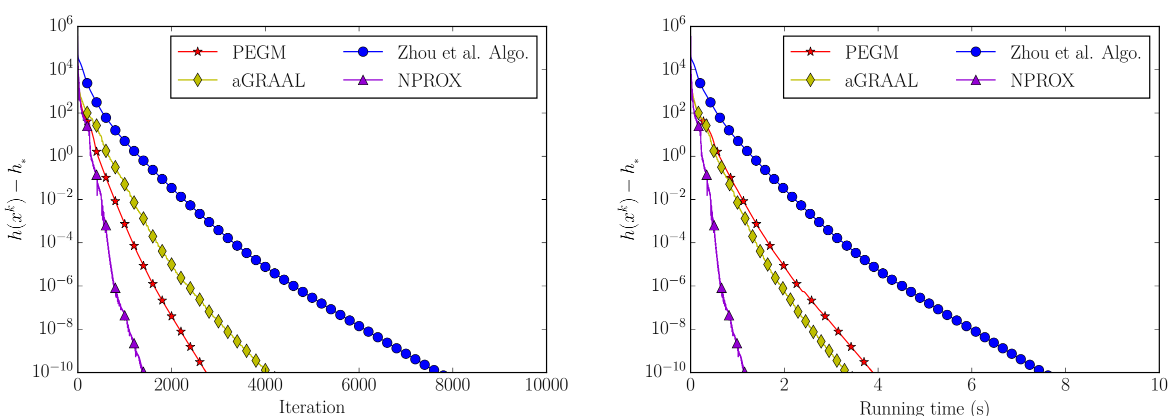

Figure 1: The numerical result of the data $a9a$ ($N = 32561, n = 123$).

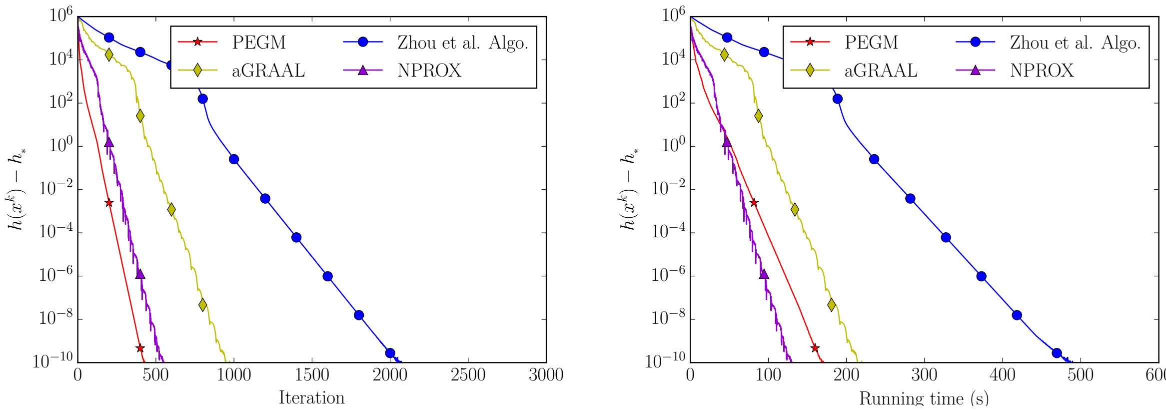

Figure 2: The numerical result of the data $kdda.t.bz2$ ($N = 510302, n = 2014669$).

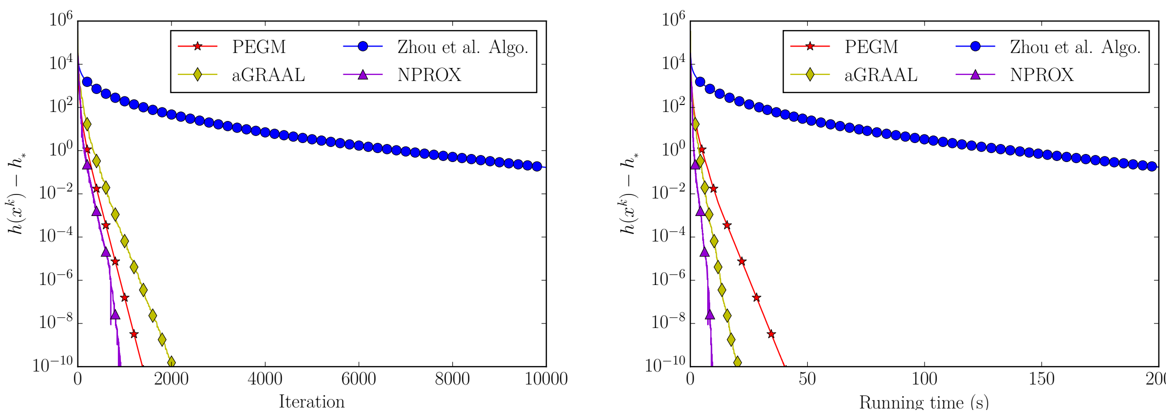

Figure 3: The numerical result of the data $real - sim.bz2$ ($N = 72309, n = 20958$).

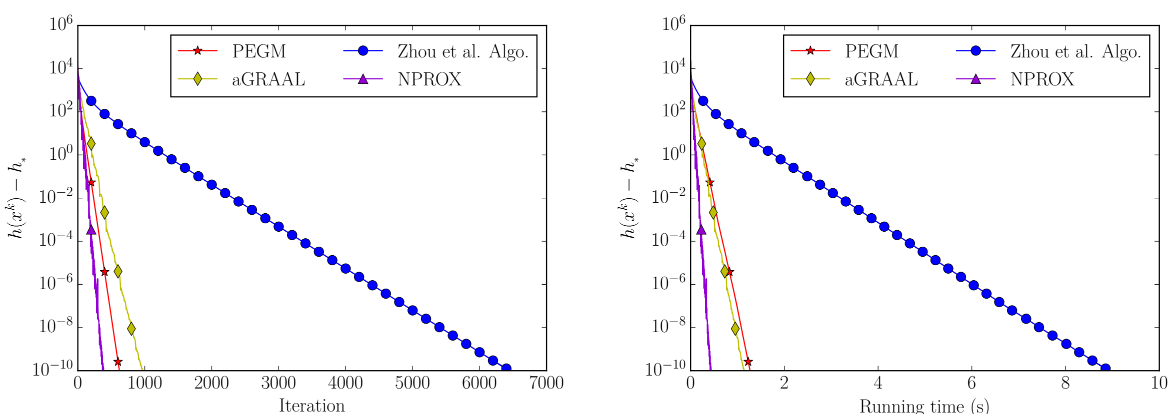

Figure 4: The numerical result of the data *ijcnn1.bz2* ($N = 49990, n = 22$).

4.2. Image deblurring

In this section, we present the experiments for a wavelet-based image deblurring problem as follows

$$\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Ux-c\|^2 + \zeta\|Wx\|_1 \right\}, \tag{4.3}$$

where c is a vectorized observed image, the matrix $U = MW$ with a blur operator M and an inverse of a three stage Haar wavelet transform W and the regularization parameter $\zeta > 0$. Obviously, (4.3) belongs to the class of problem (4.1) with

$$f(x) = \frac{1}{2}\|Ux-c\|^2 \text{ and } g(x) = \zeta\|Wx\|_1.$$

We do the experiment for the RGB image "Girl.tif"[3] size 256×256. And therefore $n = 256 \times 256 \times 3 = 196608$. Firstly, this image is blurred and corrupted by one of the following blur types:

1. Motion blur with motion length of 10 pixels and motion orientation $60°$.
2. Gaussian blur of filter size 7×7 with standard deviation 8.

After that, we add an additive zero-mean white Gaussian noise with standard deviation 10^{-3} to the obtained images. The pairs of original image final and noisy image are presented in Fig. 5, 6 and Fig. 8, 9 corresponding to the mentioned blur types. For each algorithm, the used parameters[4] are the following:

PEGM: $a = 0.41\left(2-\frac{1}{\theta}\right),\ \theta = 2,\ \sigma = 0.7$, $\lambda_0 = a\frac{\|x^1-x^0\|}{\|Ax^1-Ax^0\|}$, $\lambda_{\max} = 10^7$.

aGRAAL: $\varphi = 1.5$, $\lambda_0 = \frac{\varphi}{2}\frac{\|x^1-x^0\|}{\|Ax^1-Ax^0\|}$, $\rho = \frac{1}{\varphi} + \frac{1}{\varphi^2}$, $\bar{\lambda} = 10^7$.

Zhou et al. Algo.: $\delta = 0.6$, $\mu_1 = 0.3$, $\epsilon = 5$, $\eta = 0.005$.

NPROX: $\lambda_0 = 0.9$, $r = 1.01$, $\rho = \frac{1+\sqrt{1+4r}}{2r}$, $\eta_0 = 0.35\rho$, $\eta_1 = 0.01\eta_0$, and

$$\xi_{k-1} = \frac{0.05(\log(1.3k))^{6.7}}{k^{1.03}},\ k \geq 1.$$

The quality of the restoration is measured by PSNR (the peak signal-to-noise ratio) in decibel (dB):

$$\mathrm{PSNR}(x) = 10\log_{10}\left(\frac{255^2}{MSE}\right),$$

with

$$MSE = \frac{1}{n}\sqrt{\sum_{j=1}^{n}(\bar{x}(j)-x(j))^2},$$

where $\bar{x}$ is the original image.

[3] http://fantacci.wikidot.com/licap

[4] The parameters for PEGM, aGRAAL, Zhou et al. Algo. are chosen based on the corresponding references [19, 20, 29].

We set $\zeta = 2e - 5$ and stopping criterion is either PSNR ≥ 40 or the maximum iteration over 3000. For this experiment, code is written by Matlab 2019. The restored images obtained by the considered algorithms are shown in Fig. 7 and Fig. 11 for Gaussian and motion noise, respectively. We also report the evolution of the objective value and PSNR in Fig. 11 and Fig. 12 in both aspects of iterations and running time. The obtained results show that NPROX works better than all the remaining ones for both of considered blur types. In particular, it achieves the requirement PSNR in very fewer iterations and less computation time compared to the other methods.

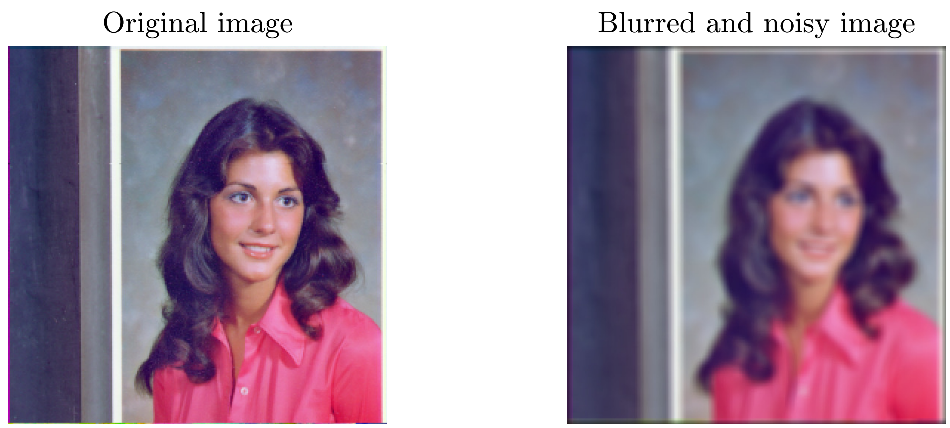

Figure 5: Original image (left); the blurred and noisy image by Gaussian (right).

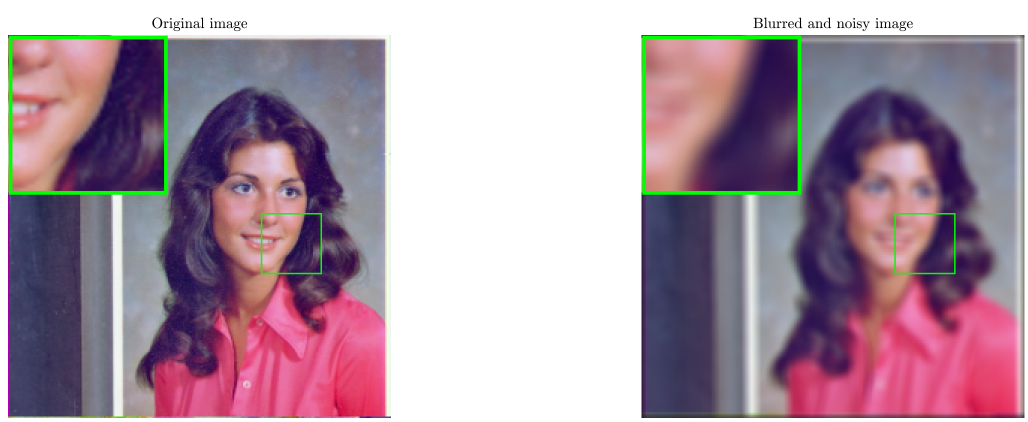

Figure 6: Original image (left); blurred and noisy image (right) with a zoom patch.

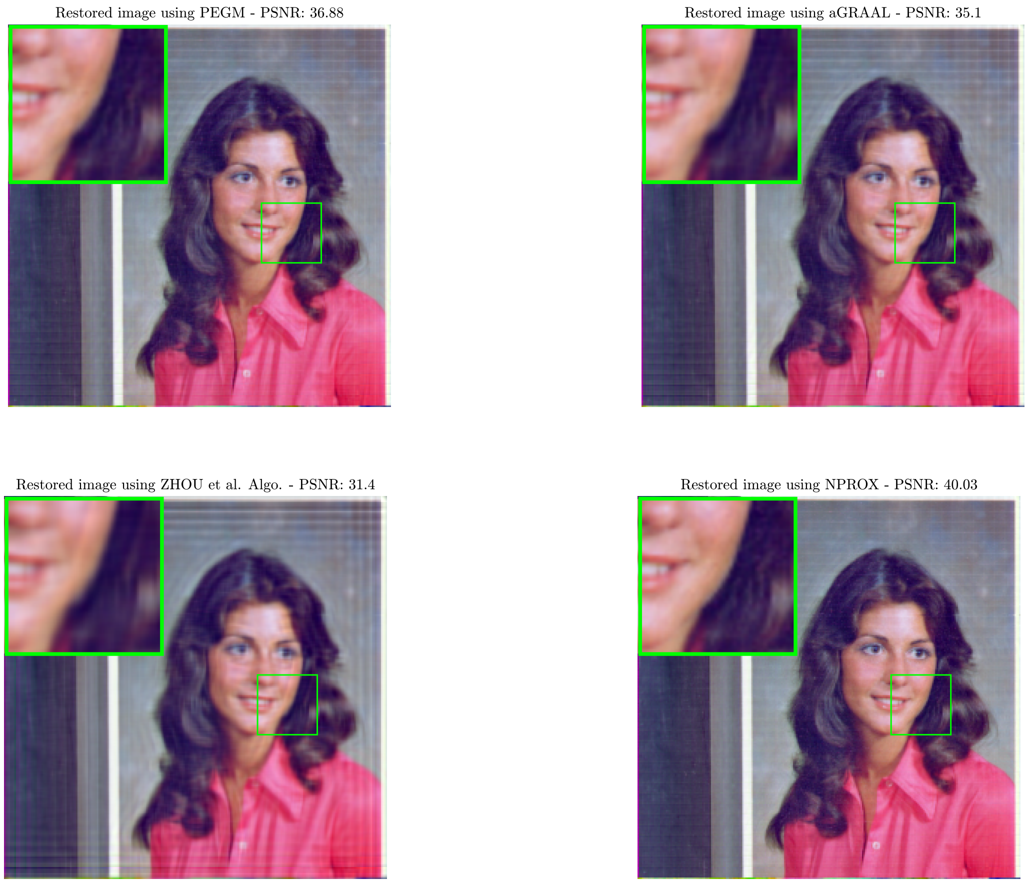

Figure 7: Restored images by considered algorithms for the case of Gaussian blur.

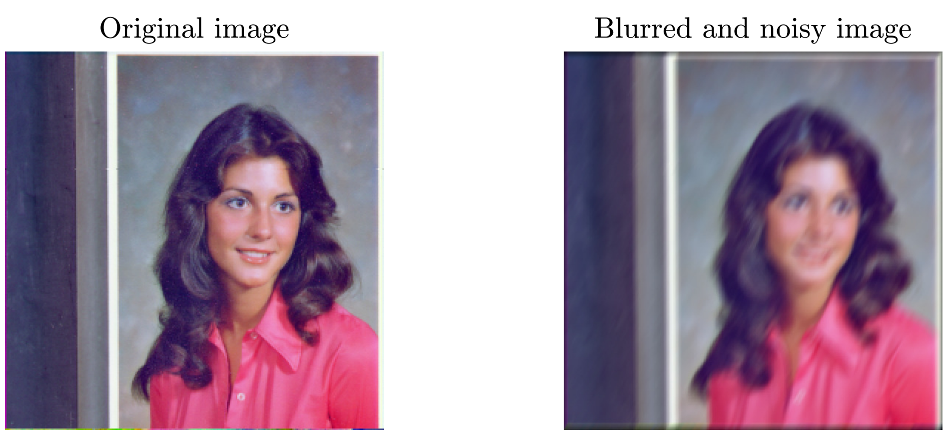

Figure 8: Original image (left); the blurred and noisy image by motion blur (right).

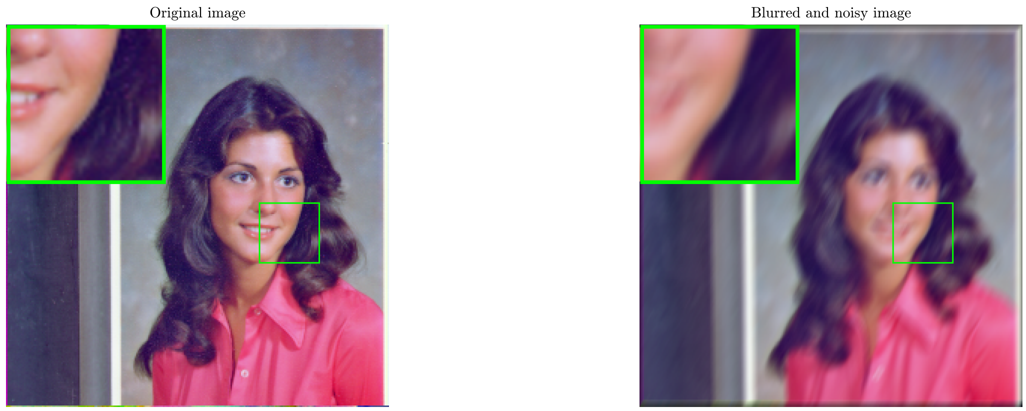

Figure 9: Original image; the motion blurred image (right) with a zoom patch.

Figure 10: Restored images by considered algorithms for the case of motion blur

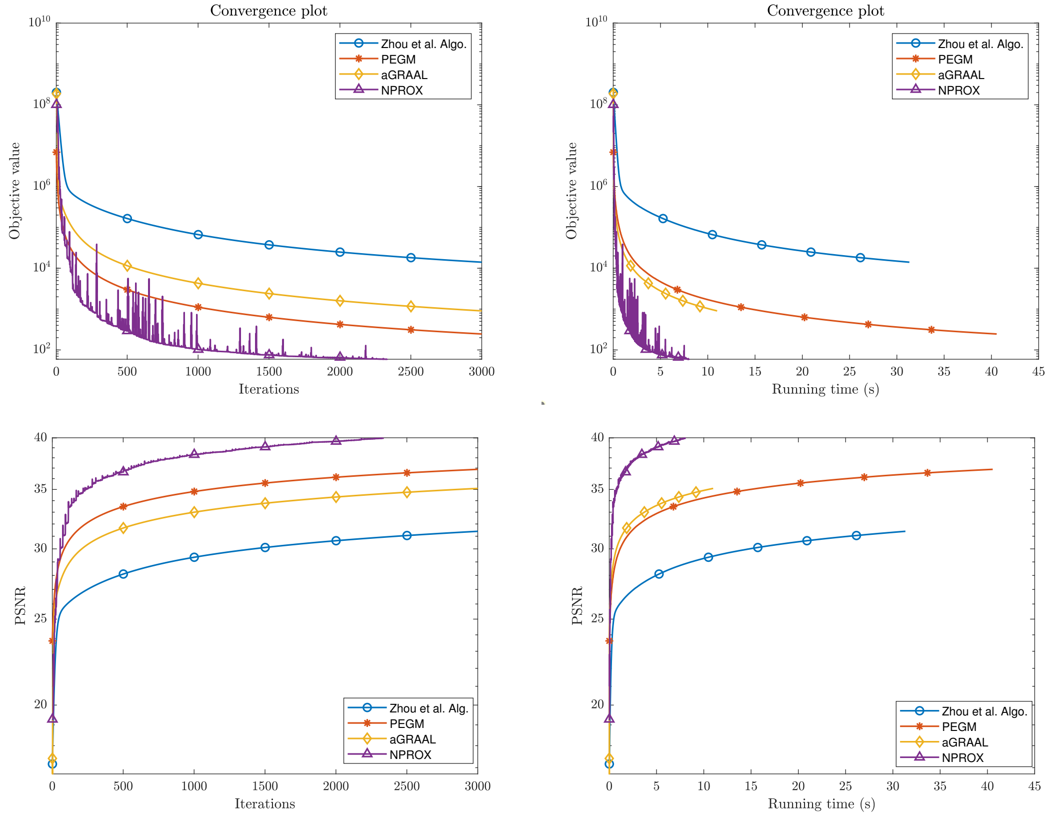

Figure 11: Evolution of the objective value and PSNR for the case of Gaussian blur.

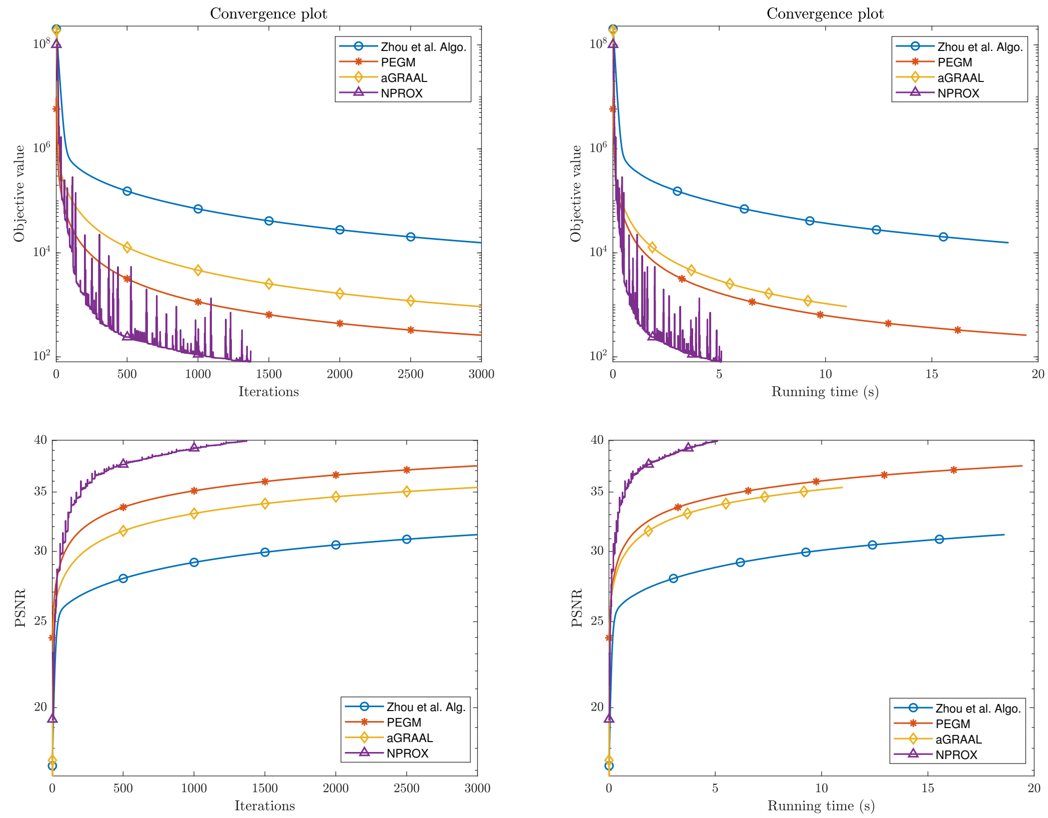

Figure 12: Evolution of the objective value and PSNR for the case of motion blur.

5. Conclusions

In this paper, we propose a novel stepsize selection for proximal gradient method solving mixed variational inequality problems (MVIP) in Hilbert spaces. We not only show the weak and R-linear convergence of the proposed algorithm under the standard assumptions of (MVIP) but also assert the increasing of the sequence of our stepsize from some fixed iteration. This interesting property can help to speed up the algorithm in general. The illustration for applications in sparse logistic regression and image deblurring problem show the efficiency of the proposed method. For the future research, one can consider the operator A under more general assumptions such as locally Lipschitz or nonmonotone.

Acknowledments

The author wishes to thank Professor Y. Malitsky for his sharing available Python codes on his page [5] that partly supports her in simulating the numerical results.

[5]https://github.com/ymalitsky

References

[1] Addi K., Goeleven D.: Complementarity and Variational Inequalities in Electronics. In: Daras, N., Rassias, T. (eds) Operations Research, Engineering, and Cyber Security. Springer Optimization and Its Applications, Vol. 113. Springer, Cham (2017).

[2] Bauschke H.H., Combettes P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, New York (2011).

[3] Beck A., Teboulle M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problem. SIAM J. Imaging Sci. 2, 183–202 (2009).

[4] Goeleven, D.: Complementarity and Variational Inequalities in Electronics, Elsevier (2017).

[5] Giannessi F., Maugeri A., Pardalos P.M.: Equilibrium problems and variational models, Kluwer, Boston (2001).

[6] Grad S. M., Lara F.: Solving mixed variational inequalities beyond convexity. Journal of Optimization Theory and Applications 190(2), 565-580 (2021).

[7] Han W.M., Reddy B.: On the finite element method for mixed variational inequalities arising in elastoplasticity. SIAM J. Numer. Anal. 32, 1778-1807 (1995).

[8] He Y.R.: A new projection algorithm for mixed variational inequalities (in Chinese), Acta Math. Sci. 27A, 215-220 (2007).

[9] Hoai P.T., Vinh N.T., Chung N.P.H.: A novel stepsize for gradient descent method, Operations Research Letters, available online 24 January 2024, 107072. https://doi.org/10.1016/j.orl.2024.107072

[10] Izuchukwu C., Shehu Y., Okeke C.C.: Extension of forward-reflected-backward method to non-convex mixed variational inequalities. J. Glob. Optim. 86, 123–140 (2023).

[11] Konnov I.V., Volotskaya, E.O.: Mixed variational inequalities and economics equilibrium problems. J. Appl. Math. 2, 289-314 (2002).

[12] Jolaoso L. O., Shehu Y., Yao J. C.: Inertial extragradient type method for mixed variational inequalities without monotonicity. Mathematics and Computers in Simulation, 192, 353-369 (2022).

[13] Ju X., Che H., Li C., He X.: Solving mixed variational inequalities via a proximal neurodynamic network with applications. Neural Processing Letters, 54(1), 207-226 (2022).

[14] Ju X., Li C., Dai Y. H., Chen J.: A new dynamical system with self-adaptive dynamical stepsize for pseudomonotone mixed variational inequalities. Optimization, 1-30 (2022). https://doi.org/10.1080/02331934.2022.2094795

[15] Ju X., Hu D., Li, C., He X., Feng G.: A novel fixed-time converging neurodynamic approach to mixed variational inequalities and applications. IEEE Transactions on Cybernetics, 52(12), 12942-12953 (2021).

[16] Liu, H., Yang, J.: Weak convergence of iterative methods for solving quasimonotone variational inequalities. Comput Optim Appl **77**, 491–508 (2020).

[17] Maingé P. E.: Projected subgradient techniques and viscosity methods for optimization with variational inequality constraints. European journal of operational research, 205(3), 501-506 (2010).

[18] Malitsky Y.: Projected reflected gradient methods for monotone variational inequalities. SIAM J. Optim. **25**, 502-520 (2015).

[19] Malitsky, Y.: Proximal extrapolated gradient methods for variational inequalities. Op-

timization Methods and Software, 33(1), 140-164 (2018).

[20] Malitsky Y.: Golden ratio algorithms for variational inequalities. Math. Program. **184**, 383–410 (2020).

[21] Peng Z. Y., Li D., Zhao Y., Liang R. L.: An accelerated subgradient extragradient algorithm for solving bilevel variational inequality problems involving non-Lipschitz operator. Commun. Nonlinear Sci. Numer. Simul. 127, 107549 (2023) .https://doi.org/10.1016/j.cnsns.2023.107549

[22] Tang G. J., Huang N. J.: Strong convergence of an inexact projected subgradient method for mixed variational inequalities. Optimization, 63(4), 601-615 (2014).

[23] Tang G. J., Zhu M., Liu H. W.: A new extragradient-type method for mixed variational inequalities. Operations Research Letters, 43(6), 567-572 (2015).

[24] Tang G. J., Wan Z., Huang N.J.: Strong convergence of a projection-type method for mixed variational inequalities in Hilbert spaces. Numerical Functional Analysis and Optimization, 39(11), 1103-1119 (2018).

[25] Xia F. Q., Li T., Zou Y. Z.: A projection subgradient method for solving optimization with variational inequality constraints. Optimization Letters, 8(1), 279-292 (2014).

[26] Xu Y., Huang Z.H.: Properties of the Solution Set of a Class of Mixed Variational Inequalities. Numerical Functional Analysis and Optimization 43(16), 1779-1800 (2022)

[27] Yang J., Liu H.: A modified projected gradient method for monotone variational inequalities. J Optim Theory Appl **179**, 197–211 (2018).

[28] Zhang H., Liu X.: Weak and strong convergence of a modified double inertial projection algorithm for solving variational inequality problems. Commun. Nonlinear Sci. Numer. Simul. 130, 107766 (2024) https://doi.org/10.1016/j.cnsns.2023.107766

[29] Zhou X., Cai G., Cholamjiak P., Kesornprom S.: A generalized proximal point algorithm with new step size update for solving monotone variational inequalities in real Hilbert spaces. J. Comput. Appl. Math. 438, 115518 (2024).