COAP manuscript No.
(will be inserted by the editor)

Proximal extrapolated gradient methods with prediction and correction for monotone variational inequalities

Xiaokai Chang[1,2] · **Sanyang Liu**[1] · **Jianchao Bai**[3] · **Jun Yang**[4]

Received: date / Accepted: date

Abstract An efficient proximal-gradient-based method, called proximal extrapolated gradient method, is designed for solving monotone variational inequality in Hilbert space. The proposed method extends the acceptable range of parameters to obtain larger step sizes. The step size is predicted based a local information of the operator and corrected by linesearch procedures to satisfy a very weak condition, which is even weaker than the boundedness of sequence generated and always holds when the operator is the gradient of a convex function. We establish its convergence and ergodic convergence rate in theory under the larger range of parameters. Furthermore, we improve numerical efficiency by employing the proposed method with non-monotonic step size, and obtain the upper bound of the parameter relating to step size by an extremely simple example. Related numerical experiments illustrate the improvements in efficiency from the larger step size.

Keywords Variational inequalities · proximal gradient method · convex optimization · nonmonotonic step size

Mathematics Subject Classification (2000) 47J20 · 65C10 · 65C15 · 90C33

1 Introduction

Let $\mathcal{H}$ be a real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. We consider the variational inequality problem:

$$\text{find } \ x^* \in \mathcal{H} \ \text{ s.t. } \langle F(x^*), y - x^* \rangle + g(y) - g(x^*) \geq 0, \ \forall y \in \mathcal{H}, \tag{1}$$

where $F : \mathcal{H} \to \mathcal{H}$ is an operator and $g : \mathcal{H} \to]-\infty, +\infty]$ is a proper lower semicontinuous convex function. We use $\operatorname{dom} g$ to represent the domain of g, defined by $\operatorname{dom} g := \{x \in \mathcal{H} : g(x) < +\infty\}$. For a continuously differentiable and convex function $f : \mathcal{H} \to]-\infty, +\infty[$ with its gradient denoted by $\nabla f = F$, then problem (1) is equivalent to

$$\min_{x \in \mathcal{H}} f(x) + g(x). \tag{2}$$

✉ Xiaokai Chang
xkchang@lut.cn
✉ Jianchao Bai
bjc1987@163.com

1 School of Science, Lanzhou University of Technology, Lanzhou, P. R. China.
2 School of Mathematics and Statistics, Xidian University, Xi'an, P. R. China.
3 Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, P. R. China.
4 School of Mathematics and Information Science, Xianyang Normal University, Xianyang, P. R. China.

Let C be a closed and convex subset of $\mathcal{H}$. Let l_C be the indicator function of the set C, that is, $l_C(x) = 0$ if $x \in C$ and ∞ otherwise. When $g(x) = l_C(x)$, variational inequality (1) reduces to

$$\text{find} \quad x^* \in C \quad \text{s.t.} \ \langle F(x^*), y - x^* \rangle \geq 0, \ \forall y \in \mathcal{H}. \tag{3}$$

Problem (1) and its special cases (2) and (3) have wide applications in disciplines including mechanics, signal and image processing, and economics [3–5, 14, 15, 32], to cite a few. Throughout the paper, the solution set $\mathcal{S}$ of problem (1) is assumed to be nonempty, and the following assumptions hold:

(A1) F is monotone, i.e.,

$$\langle F(x) - F(y), x - y \rangle \geq 0, \quad \forall x, y \in \mathcal{H};$$

(A2) F is L-Lipschitz continuous ($L > 0$), that is,

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{H};$$

(A3) $g|\operatorname{dom} g$ is a continuous function.

Many efficient methods have been proposed for solving the problem (1) and its special cases, for instance, alternating direction method of multipliers (ADMM) [5, 8, 9, 17], extragradient method [1, 20, 21, 27], proximal (projected) gradient method [11, 18, 23, 26, 38, 41] and its accelerated version [25, 33]. Here, we would concentrate on the most simple case of these approaches: forward-backward splitting (FBS) method. Under the assumption that F is L-Lipschitz continuous, the iterative scheme of the classical FBS method for problem (1) reads

$$x_{n+1} = \operatorname{prox}_{\lambda g}(x_n - \lambda F(x_n)), \tag{4}$$

where λ is some positive number and can be viewed as a step size of the forward step, and the proximal operator $\operatorname{prox}_{\lambda g} : \mathcal{H} \to \mathcal{H}$ is defined in Section 2.

To establish convergence of the iteration (4), it often requires the restrictive assumptions that F is L-Lipschitz continuous, strongly (or inverse strongly) monotone with $\lambda \in]0, \frac{2}{L}[$. To overcome this drawback, Korpelevich [21] and Antipin [1] proposed the following extragradient method for (3) with two-step projection procedures

$$y_n = P_C(x_n - \lambda_n F(x_n)), \qquad x_{n+1} = P_C(x_n - \lambda_n F(y_n)),$$

where $P_C : \mathcal{H} \to C$ denotes the (metric) projection onto C, λ_n is any positive sequence verifying $\lambda_n \in [l, u]$ for some values $l, u \in]0, \frac{1}{L}[$. The extragradient method has received great attentions and has been improved in various ways [10, 13, 16, 24, 28], including linesearch procedures or/and avoiding Lipschitz-continuity assumption, decreasing a number of metric projections, etc. For instance, Censor, Gibali and Reich [10] introduced

$$\left.\begin{array}{l} y_n = P_C(x_n - \lambda F(x_n)), \\ T_n = \{w \in \mathcal{H} | \langle x_n - \lambda F(x_n) - y_n, \ w - y_n \rangle \leq 0\}, \\ x_{n+1} = P_{T_n}(x_n - \lambda F(y_n)), \end{array}\right\} \tag{5}$$

where the step size satisfies $\lambda \in]0, \frac{1}{L}[$. Since the second projection P_{T_n} in (5) can be found in a closed form, this method is more applicable when a projection onto the closed convex set C is a nontrivial problem. For a more general problem (1), Tseng [41] modified the iteration (4) and proposed the following forward-backward-forward (FBF) method involving one proximal operator and two values of F per iteration:

$$y_n = \operatorname{prox}_{\lambda g}(x_n - \lambda F(x_n)), \qquad x_{n+1} = y_n + \lambda(F(x_n) - F(y_n)),$$

where $\lambda \in]0, \frac{1}{L}[$. Since then, Tseng's method has attracted a lot of interests due to its simplicity and generality, see [6, 7, 31] for more details.

In the literature, the inertial extrapolation has been conducted to accelerated proximal gradient methods in the spirit of Nesterov's extrapolation techniques [33,34], whose basic idea is to make full use of historical information at each iteration. A typical scheme of the proximal gradient method with extrapolation for solving (1) is

$$x_{n+1} = \mathrm{Prox}_{\lambda g}(x_n - \lambda F(y_n)), \quad y_{n+1} = x_{n+1} + \delta_n(x_{n+1} - x_n), \tag{6}$$

where $\delta_n > 0$. Recently, using a fixed parameter $\delta = 1$ in (6), Malitsky [28] introduced the iteration

$$x_{n+1} = P_C(x_n - \lambda F(2x_n - x_{n-1})), \quad \lambda \in]0, (\sqrt{2}-1)/L[,$$

for solving (3). However, the step size (λ or λ_n) requires the information of the Lipschitz constant L, which is a main drawback of the algorithms introduced above. In fact, these algorithms with a large value of L can lead to very small step size, which may give rise to a slow convergent algorithm [30]. To obtain a proper step size, Armijo-type line search and outer approximation techniques were involved in [19,20,22,40]. Due to the extra proximal operator as well as the evaluations of F, these algorithms will be computationally expensive when proximal operator or F is hard to compute and somewhat expensive.

For getting a proper step without using the Lipschitz constant L, Malitsky [28] introduced an efficient method whose main updates are

$$\begin{cases} \text{Choose} \quad x_0 = y_0 \in \mathcal{H}, \ \lambda_0 > 0, \ \alpha \in]0, \sqrt{2}-1[\\ \text{Choose } \lambda_n \quad s.t. \ \lambda_n \parallel Fy_n - Fy_{n-1} \parallel \le \alpha \parallel y_n - y_{n-1} \parallel \\ x_{n+1} = P_C(x_n - \lambda_n F(y_n)) \\ y_{n+1} = 2x_{n+1} - x_n. \end{cases} \tag{7}$$

By updating the step size λ_n given by a specific procedure according the progress of algorithm, a weak convergence result was proved, but this process involves the computation of additional projections onto C. Later, Mainge and Gobinddass [30] introduced a more general framework:

$$\theta_n = \frac{\lambda_n}{\delta \lambda_{n-1}}, \quad y_n = x_n + \theta_n(x_n - x_{n-1}), \quad x_{n+1} = P_C(x_n - \lambda_n F(y_n)),$$

where the step size λ_n needs to satisfy many inequality constraints and can be obtained by line-search procedure, see [30, Section 3.1 and Section 3.2.2]. Based on the scheme (7), local information of the operator and some linesearch procedures, Malitsky [29] proposed simpler schemes which do not require Lipschitz continuity of the operator. Furthermore, the involved linesearch procedure doesn't need extra prox or projection and it can be applied to a more general problem (1). By overcoming the estimation of L and linesearch procedure for the scheme (7), Yang and Liu [42] proposed an extragradient method with lower computational complexity but nonincreasing step sizes. The important parameter α relating to the step size λ_n was restricted on $\alpha \in]0, \frac{\sqrt{2}-1}{\delta}[$ with $\delta \in]1, +\infty[$ in [42] and $\alpha \in]0, \sqrt{2}-1[$ with variable δ_n from linesearch in [28] for guaranteeing the convergence.

The aim of this paper is to propose a proximal gradient algorithm with larger step size, extend the range of δ to that is less than or equal to 1, and then improve the range of α. Our proposed methods do not require Lipschitz constant, and its step size is predicted by using two previous iterates, and corrected by linesearch to satisfy a very weak condition, which always holds when $F = \nabla f$ for a convex function f. Specifically, by the aid of the vital inequalities in convergence's proof we first introduce a function $\kappa(\delta)$ defined as

$$\kappa(\delta) := \max_{\varepsilon_1>0,\varepsilon_2>0} \min \left\{ \frac{\varepsilon_1}{\delta(\varepsilon_1^2 + \varepsilon_2 + 1)}, \frac{(\delta^2+\delta-1)\varepsilon_1\varepsilon_2}{\delta^3(1+\varepsilon_2)} \right\} \tag{8}$$

for any $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$ to ensure some convergence properties. Then we get $\max_{\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[} \kappa(\delta) = \kappa(\sqrt{3}-1) = \frac{1}{2}$, and use $\alpha \in]0, \kappa(\delta)[$ to control the step size. Our range of α is larger than that presented in [28, 42], see Lemma 2 for more explanations. Secondly, the region of δ is partitioned as

$$\delta \in \left](\sqrt{5}-1)/2, +\infty\right[= \left](\sqrt{5}-1)/2, 1\right[\ \cup \ [1, +\infty[$$

to explore convergence of the proposed method, and the $\mathcal{O}(1/n)$ ergodic convergence rate is established. Finally, we obtain the upper bound of α by an extremely simple example, and improve numerical efficiency by introducing nonmonotonic step size λ_n but $\frac{\lambda_n}{\lambda_{n-1}} \to 1$. In fact, the proposed nonmonotonic step size can break away from overdependence on the initial point, but it would have to be monotonic in the end for getting convergence.

The paper is organized as follows. In Section 2, we provide some useful facts and notations. In Section 3, we introduce our algorithm and explore the properties of the function $\kappa(\delta)$. A weak convergence theorem of our method is proved in Section 3.1. In Section 3.2, we establish the ergodic convergence rate of the proposed algorithms, and we improve the algorithms in Section 3.3 to avoid the adverse effects of the nonincreasing step size. In Section 4, we show by an example that any value of $\alpha \in]\frac{2}{2\delta+1}, +\infty[$ with $\delta \in]0, +\infty[$ does not guarantee convergence of our algorithm. Numerical experiments on solving some problems tested in the literatures are provided and analyzed in Section 5. We finally conclude our paper in Section 6.

2 Preliminaries

In this section, we introduce some notations and facts on the well-known properties of the proximal operator, Opial condition and Young's inequality, which are used for the sequel convergence analyses.

The proximal operator $\text{prox}_{\lambda g} : \mathcal{H} \to \mathcal{H}$ with $\text{prox}_{\lambda g}(x) = (I + \lambda\partial g)^{-1}(x), \lambda > 0, x \in \mathcal{H}$, is defined by

$$\text{prox}_{\lambda g}(x) := \underset{y\in\mathcal{H}}{\text{argmin}} \left\{ g(y) + \frac{1}{2\lambda}\|x-y\|^2 \right\}, \quad \forall x \in \mathcal{H}, \lambda > 0.$$

Setting

$$\Phi(x, y) := \langle F(x), y - x\rangle + g(y) - g(x), \tag{9}$$

it is clear that problem (1) is equivalent to finding $x^* \in \mathcal{H}$ such that $\Phi(x^*, y) \geq 0$ for all $y \in \mathcal{H}$.

Fact 1 *[2] Let $g : \mathcal{H} \to (-\infty, +\infty]$ be a convex function, $\lambda > 0$ and $x \in \mathcal{H}$. Then $p = prox_{\lambda g}(x)$ if and only if*

$$\langle p - x, y - p\rangle \geq \lambda[g(p) - g(y)], \quad \forall y \in \mathcal{H}.$$

Fact 2 *[36] (Opial 1967) Let $\mathcal{S}$ be a nonempty set of $\mathcal{H}$ and $\{x_n\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that the following two conditions hold:*
(1) for every $x^* \in \mathcal{S}$, $\lim\limits_{n\to+\infty} \|x_n - x^*\|$ exists;
(2) every sequential weak cluster point of $\{x_n\}_{k\in\mathbb{N}}$ is in $\mathcal{S}$.
Then $\{x_n\}_{k\in\mathbb{N}}$ converges weakly to a point in $\mathcal{S}$.

Fact 3 *Let $\{a_n\}$, $\{b_n\}$ be two nonnegative real sequences and $\exists N > 0$ such that*

$$a_{n+1} \leq a_n - b_n, \quad \forall n > N.$$

Then $\{a_n\}$ is convergent and $\lim\limits_{n\to\infty} b_n = 0$.

Fact 4 *(Young's inequality) For all $a, b \geq 0$ and $\varepsilon > 0$, we have*

$$ab \leq \frac{a^2}{2\varepsilon} + \frac{\varepsilon b^2}{2}.$$

The following identity (cosine rule) appears in many times and we will use it for simplicity of convergence analyses. For all $x, y, z \in \mathcal{H}$,

$$\langle x - y, x - z\rangle = \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|x - z\|^2 - \frac{1}{2}\|y - z\|^2. \tag{10}$$

3 Proximal Extrapolated Gradient Method with Prediction and Correction

In this section, we state our proximal extrapolated gradient method with prediction and correction (PEG), by using the step size function $\kappa(\delta)$ defined in (8).

Algorithm 1 (PEG for solving (1))

Step 0. Take $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$, choose $x_0 \in \mathcal{H}$, $\lambda_0 > 0$, $\gamma \in (0, 1)$, $\alpha \in]0, \kappa(\delta)[$ and a bounded sequence $\{\zeta_n > 0\}$. Set $y_0 = x_0$, $x_1 = \mathrm{prox}_{\lambda_0 g}(x_0 - \lambda_0 F(x_0))$ and $n = 1$.

Step 1. **Prediction:**

1.a. Compute

$$y_n = x_n + \delta(x_n - x_{n-1}), \tag{11}$$

$$\lambda_n = \min\left\{\lambda_{n-1}, \ \frac{\alpha\|y_n - y_{n-1}\|}{\|F(y_n) - F(y_{n-1})\|}\right\}. \tag{12}$$

1.b. Compute

$$x_{n+1} = \mathrm{prox}_{\lambda_n g}(x_n - \lambda_n F(y_n)),$$

if $x_{n+1} = x_n = y_n$, then stop: x_{n+1} is a solution.

Step 2. **Correction when $\delta < 1$:**

Check

$$\|x_{n+1} - x_n\| \le \zeta_n,$$

if not hold, set $\lambda_n \leftarrow \gamma\lambda_n$ and return to Step 1.b.

Step 3. Set $n \leftarrow n + 1$ and return to Step 1.

The aim of Correction step is to bound $\{\|x_n - x_{n-1}\|\}$ by the given sequence $\{\zeta_n\}$ when $\delta < 1$, as convergence analysis requires $\|x_{n+1} - x_n\| < +\infty$. In practice, we don't need to give the sequence $\{\zeta_n\}$, but generate adaptively by

$$\zeta_n = \max\{\zeta_{\min}, \quad \min\{\mu\|x_n - x_{n-1}\|, \quad \nu\|x_1 - x_0\|\}\}, \tag{13}$$

for given $1 < \mu \le \nu$ and small $\zeta_{\min}$ (e.g., $\zeta_{\min} = 10^{-6}$), then $\zeta_n \le \nu\|x_1 - x_0\|$ for all $n \ge 1$ and $\zeta_n \ge \zeta_{\min}$. Moreover, we observe $\|x_{n+1} - x_n\| \le \mu\|x_n - x_{n-1}\|$ for bounding more tightly due to $\|x_{n+1} - x_n\| \to 0$.

For a convex function f, if $F = \nabla f$ we observe $\|x_{n+1} - x_n\| < +\infty$, see (28), so Correction step is not necessary. However for other cases, one needs to apply linesearch to ensure $\|x_{n+1} - x_n\| < +\infty$. Interestingly, for all the tested problems shown in Section 5, the linesearch in Correction step does not start to arrive termination conditions, when using (13) with $\mu = \nu = 10$. Namely, the predicted step is good enough for obtaining a convergent sequence for the tested problems, though the convergence without prediction is unknown in general.

The following lemma shows that the correction procedure described in Algorithm 1 is well-defined.

Lemma 1 *The correction procedure always terminates. i.e., $\{\lambda_n\}$ is well defined when $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$.*

Proof. Denote

$$A := \partial g \quad \text{and} \quad x_{n+1}(\lambda) := \text{prox}_{\lambda g}(x_n - \lambda F(y_n)).$$

From [2, Theorem 23.47], we have that $\text{prox}_{\lambda g}[x_{n+1}(0)] \to P_{\overline{\text{dom}\,A}}[x_{n+1}(0)]$ as $\lambda \to 0$ ($\overline{\text{dom}\,A}$ denotes the closures of dom A), which together with the nonexpansivity of $\text{prox}_{\lambda g}$ yields

$$\begin{aligned} &\|x_{n+1}(\lambda) - P_{\overline{\text{dom}\,A}}[x_{n+1}(0)]\| \\ \leq\ &\|x_{n+1}(\lambda) - \text{prox}_{\lambda g}[x_{n+1}(0)]\| + \|\text{prox}_{\lambda g}[x_{n+1}(0)] - P_{\overline{\text{dom}\,A}}[x_{n+1}(0)]\| \\ \leq\ &\lambda\|F(y_n)\| + \|\text{prox}_{\lambda g}[x_{n+1}(0)] - P_{\overline{\text{dom}\,A}}[x_{n+1}(0)]\|. \end{aligned}$$

By taking the limit as $\lambda \to 0$, we deduce that $x_{n+1}(\lambda) \to P_{\overline{\text{dom}\,A}}[x_{n+1}(0)]$. Notice that $x_{n+1}(0) = x_n$, we observe $P_{\overline{\text{dom}\,A}}[x_{n+1}(0)] = x_n$.

By a contradiction, suppose that the correction procedure in Algorithm 1 fails to terminate at the n-th iteration. Then, for all $\lambda = \gamma^i\lambda_n$ with $i = 0, 1, \cdots$, we have $\|x_{n+1}(\lambda) - x_n\| > \zeta_n$. Since $\gamma^i \to 0$ as $i \to \infty$, so $\lambda \to 0$, this gives a contradiction $0 \geq \zeta_n$, which completes the proof. □

Remark 1 *Note that the sequence $\{\lambda_n\}$ is monotonically decreasing. Since F is a L-Lipschitz continuous mapping ($L > 0$), we have*

$$\frac{\alpha\|y_n - y_{n-1}\|}{\|F(y_n) - F(y_{n-1})\|} \geq \frac{\alpha\|y_n - y_{n-1}\|}{L\|y_n - y_{n-1}\|} = \frac{\alpha}{L}$$

for $F(y_n) \neq F(y_{n-1})$. Thus the predicted step sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ has a lower bound $\tau := \min\{\frac{\alpha}{L}, \lambda_0\}$, then when $\delta \geq 1$ its limit exists and $\lim_{n\to\infty} \lambda_n \geq \tau > 0$. If $\delta < 1$, $\{\lambda_n\}$ is well defined from Lemma 1, and has a lower bound $\tau := \min\{\frac{\gamma^{i_0}\alpha}{L}, \lambda_0\}$ for some $i_0 \geq 0$, which implies $\lim_{n\to\infty} \lambda_n > 0$ as well.

Below, we derive the analytical expression of $\kappa(\delta)$.

Lemma 2 *For the function $\kappa(\delta)$ defined in (8), we have $\kappa(\delta) = \frac{\sqrt{a+1}}{\delta(a+1+\sqrt{a+1})}$ with $a = \frac{\delta^2}{\delta^2+\delta-1}$ for $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$.*

Proof. Fix $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$, then $\delta^2 + \delta - 1 > 0$. Noting that the structure of (8) and $\kappa(\delta)$ is a maximum value, so $\frac{\varepsilon_1}{\delta(\varepsilon_1^2+\varepsilon_2+1)} = \frac{(\delta^2+\delta-1)\varepsilon_1\varepsilon_2}{\delta^3(1+\varepsilon_2)}$, which together with $a = \frac{\delta^2}{\delta^2+\delta-1}$ and $\varepsilon_1 = \sqrt{a+1}$ shows

$$\kappa(\delta) = \max_{\varepsilon_2>0} \frac{\sqrt{a\varepsilon_2 + (a-1)\varepsilon_2^2 - \varepsilon_2^3}}{\delta(a + a\varepsilon_2)}. \tag{14}$$

By the first-order optimality condition of the optimization problem (14), we have $\varepsilon_2 = \sqrt{a+1} - 1$. Substituting it into (14), the result can be deduced. □

By Lemma 2 and Fig. 1, the maximum value of $\kappa(\delta)$ is $\frac{1}{2}$ when $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$, and in fact,

$$\max_{\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[} \kappa(\delta) = \kappa(\sqrt{3} - 1) = \frac{1}{2}.$$

In this case, we have $a = 2$, $\varepsilon_1 = \sqrt{3}$ and $\varepsilon_2 = \sqrt{3} - 1$.

Remark 2 *It can be noticed that the method proposed in [42] is a special case of Algorithm 1, when $g(x) = l_C(x)$ and $\delta \in]1, +\infty[$, but $\kappa(\delta) > \frac{\sqrt{2}-1}{\delta}$ from Lemma 2. Namely, we extend the range of δ and then improve the upper bound of α when the operator is the gradient of a convex function or using linesearch, see Fig. 1, which causes larger step size λ_n that will be more efficient for numerical experiments.*

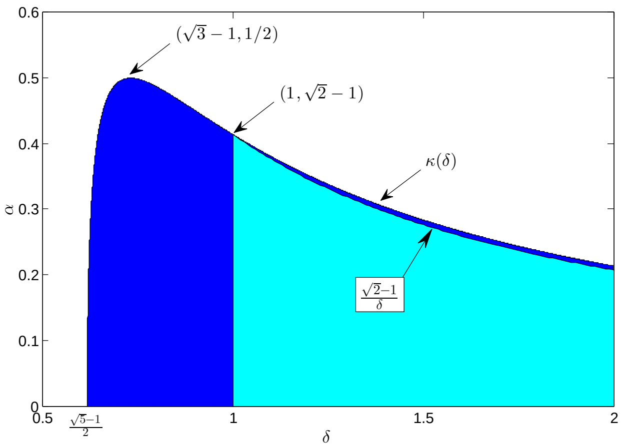


Fig. 1 The improved region (blue) of the parameters α and δ.

3.1 Convergence Analysis

This section devotes to studying convergence properties of Algorithm 1. For $\delta \in [1, +\infty[$, its convergence and convergence rate can be obtained by combining the methods in [29, 42] with the basic theory of limit. However, it is a completely different situation for $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$, since the desired properties (such as monotonicity and nonnegativity) are no longer valid in the case of $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$ although we can adopte a larger value of α.

We next give a basic lemma about the iterations generated by Algorithm 1 for any $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$, which play a crucial role in proving the main convergence results.

Lemma 3 *Let $\{x_n\}$ and $\{y_n\}$ be two sequences generated by Algorithm 1. For any $x \in \mathcal{H}$, we have*

$$\begin{aligned} \|x_{n+1} - x\|^2 \le \|x_n - x\|^2 - \frac{\lambda_n}{\delta\lambda_{n-1}}[\|y_n - x_n\|^2 + \|x_{n+1} - y_n\|^2] + \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - x_n\|^2 \\ +2\alpha\|y_n - y_{n-1}\|\|x_{n+1} - y_n\| - 2\lambda_n[(1+\delta)g(x_n) - \delta g(x_{n-1}) - g(x)] \\ -2\lambda_n\langle F(y_n), y_n - x\rangle. \end{aligned} \tag{15}$$

Proof. Followed by $x_{n+1} = \mathrm{prox}_{\lambda_n g}(x_n - \lambda_n F(y_n))$ and Fact 1, we have

$$\langle x_{n+1} - x_n + \lambda_n F(y_n), x - x_{n+1}\rangle \ge \lambda_n(g(x_{n+1}) - g(x)), \quad \forall x \in \mathcal{H}, \tag{16}$$

which shows

$$\langle x_n - x_{n-1} + \lambda_{n-1}F(y_{n-1}), x - x_n\rangle \ge \lambda_{n-1}(g(x_n) - g(x)), \quad \forall x \in \mathcal{H}.$$

Substituting $x := x_{n+1}$ and $x := x_{n-1}$ into the above inequality respectively, we obtain

$$\langle x_n - x_{n-1} + \lambda_{n-1}F(y_{n-1}), x_{n+1} - x_n\rangle \ge \lambda_{n-1}(g(x_n) - g(x_{n+1})), \tag{17}$$

$$\langle x_n - x_{n-1} + \lambda_{n-1}F(y_{n-1}), x_{n-1} - x_n\rangle \ge \lambda_{n-1}(g(x_n) - g(x_{n-1})). \tag{18}$$

Multiplying (18) by δ and then adding it to (17), which by $y_n = x_n + \delta(x_n - x_{n-1})$ yields

$$\langle x_n - x_{n-1} + \lambda_{n-1}F(y_{n-1}), x_{n+1} - y_n\rangle \ge \lambda_{n-1}[(1+\delta)g(x_n) - g(x_{n+1}) - \delta g(x_{n-1})]. \tag{19}$$

Multiplying (19) by $\frac{\lambda_n}{\lambda_{n-1}}$ and using $y_n = x_n + \delta(x_n - x_{n-1})$ again, we get

$$\left\langle \frac{\lambda_n(y_n - x_n)}{\delta\lambda_{n-1}} + \lambda_n F(y_{n-1}), x_{n+1} - y_n \right\rangle \geq \lambda_n[(1+\delta)g(x_n) - g(x_{n+1}) - \delta g(x_{n-1})]. \tag{20}$$

Finally, adding (16) to (20) gives us

$$\begin{aligned}
&\langle x_{n+1} - x_n, x - x_{n+1}\rangle + \left\langle \frac{\lambda_n(y_n - x_n)}{\delta\lambda_{n-1}}, x_{n+1} - y_n \right\rangle + \lambda_n\langle F(y_n) - F(y_{n-1}), y_n - x_{n+1}\rangle \\
&\geq \lambda_n[(1+\delta)g(x_n) - g(x) - \delta g(x_{n-1})] + \lambda_n\langle F(y_n), y_n - x\rangle.
\end{aligned}$$

Then, using (10), the updating of λ_n and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\|x_{n+1} - x\|^2 \leq\ & \|x_n - x\|^2 - \frac{\lambda_n}{\delta\lambda_{n-1}}[\|y_n - x_n\|^2 + \|x_{n+1} - y_n\|^2] + \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - x_n\|^2 \\
&+ 2\lambda_n\|F(y_n) - F(y_{n-1})\|\|x_{n+1} - y_n\| - 2\lambda_n[(1+\delta)g(x_n) - \delta g(x_{n-1}) - g(x)] \\
&- 2\lambda_n\langle F(y_n), y_n - x\rangle \\
\leq\ & \|x_n - x\|^2 - \frac{\lambda_n}{\delta\lambda_{n-1}}[\|y_n - x_n\|^2 + \|x_{n+1} - y_n\|^2] + \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - x_n\|^2 \\
&+ 2\alpha\|y_n - y_{n-1}\|\|x_{n+1} - y_n\| - 2\lambda_n[(1+\delta)g(x_n) - \delta g(x_{n-1}) - g(x)] \\
&- 2\lambda_n\langle F(y_n), y_n - x\rangle.
\end{aligned}$$

The proof is completed. □

Lemma 4 *Let $\{x_n\}$, $\{y_n\}$ be two sequences generated by Algorithm 1 and $\bar{x} \in \mathcal{S}$ (the solution set of problem (1)). Then, for any $\varepsilon_1, \varepsilon_2 > 0$, we have*

$$\begin{aligned}
\|x_{n+1} - \bar{x}\|^2 + 2\lambda_n(1+\delta)\Phi(\bar{x}, x_n) \leq\ & \|x_n - \bar{x}\|^2 + 2\lambda_{n-1}(1+\delta)\Phi(\bar{x}, x_{n-1}) \\
&+ \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right]\|x_n - y_n\|^2 \\
&+ \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - x_n\|^2 \\
&+ \left(\varepsilon_1\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2,
\end{aligned}$$

where Φ is defined as in (9).

Proof. Using Fact 4, for any $\varepsilon_1 > 0$ we have

$$2\alpha\|y_n - y_{n-1}\|\|y_n - x_{n+1}\| \leq \alpha(\frac{1}{\varepsilon_1}\|y_n - y_{n-1}\|^2 + \varepsilon_1\|x_{n+1} - y_n\|^2).$$

Meanwhile, for any $\varepsilon_2 > 0$ we deduce

$$\begin{aligned}
\|y_n - y_{n-1}\|^2 &= \|y_n - x_n\|^2 + \|x_n - y_{n-1}\|^2 + 2\langle y_n - x_n, x_n - y_{n-1}\rangle \\
&\leq \|y_n - x_n\|^2 + \|x_n - y_{n-1}\|^2 + 2\|y_n - x_n\|\|x_n - y_{n-1}\| \\
&\leq (1 + \frac{1}{\varepsilon_2})\|y_n - x_n\|^2 + (1+\varepsilon_2)\|x_n - y_{n-1}\|^2.
\end{aligned}$$

Combining the above inequalities we have

$$\begin{aligned}
&2\alpha\|y_n - y_{n-1}\|\|y_n - x_{n+1}\| \\
&\leq \alpha\left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\|y_n - x_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\|x_n - y_{n-1}\|^2 + \varepsilon_1\|x_{n+1} - y_n\|^2\right].
\end{aligned} \tag{21}$$

In addition, the monotonicity of F implies for any $x \in \mathcal{H}$

$$\begin{aligned}\lambda_n\langle F(y_n), y_n - x\rangle &\geq \lambda_n\langle F(x), y_n - x\rangle \\ &= \lambda_n[(1+\delta)\langle F(x), x_n - x\rangle - \delta\langle F(x), x_{n-1} - x\rangle]. \end{aligned} \tag{22}$$

Substituting (21) and (22) into (15), we deduce by the aids of $\varPhi(x, y)$ in (9) that

$$\begin{aligned}\|x_{n+1} - x\|^2 \leq{}& \|x_n - x\|^2 - \|x_{n+1} - x_n\|^2 + \frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha\|y_n - x_n\|^2 \\ &+ \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 + \varepsilon_1\alpha\|x_{n+1} - y_n\|^2 \\ &+ \frac{\lambda_n}{\delta\lambda_{n-1}}(\|x_{n+1} - x_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2) \\ &- 2\lambda_n[(1+\delta)\varPhi(x, x_n) - \delta\varPhi(x, x_{n-1})]. \end{aligned} \tag{23}$$

Since $\delta > 0$ and $\{\lambda_n\}_{n\in\mathbb{N}}$ is a monotone decreasing sequence, we have $\lambda_n\delta \leq \lambda_{n-1}\delta \leq (1+\delta)\lambda_{n-1}$. Note that $\varPhi(\bar{x}, x_{n-1}) \geq 0$ for any $\bar{x} \in \mathcal{S}$, then

$$\begin{aligned}\|x_{n+1} - \bar{x}\|^2 \leq{}& \|x_n - \bar{x}\|^2 + \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right]\|x_n - y_n\|^2 \\ &+ \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - x_n\|^2 \\ &+ \left(\varepsilon_1\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 \\ &- 2\lambda_n(1+\delta)\varPhi(\bar{x}, x_n) + 2\lambda_{n-1}(1+\delta)\varPhi(\bar{x}, x_{n-1}). \end{aligned}$$

This completes the proof. □

By Lemma 4 and some transpositions, we have the following results directly.

Lemma 5 *Let $\{x_n\}$, $\{y_n\}$ be two sequences generated by Algorithm 1 and $\bar{x} \in \mathcal{S}$. Then, for any $\varepsilon_1, \varepsilon_2 > 0$, we have*

$$a_{n+1} \leq a_n - b_n, \tag{24}$$

where

$$\begin{cases} a_n = \|x_n - \bar{x}\|^2 + 2\lambda_{n-1}(1+\delta)\varPhi(\bar{x}, x_{n-1}) + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 \\ \qquad + \left(1 - \frac{\lambda_{n-1}}{\delta\lambda_{n-2}}\right)\|x_n - x_{n-1}\|^2, \quad n \geq 2, \\ b_n = \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \left(\varepsilon_1 + \frac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha\right]\|x_{n+1} - y_n\|^2 \\ \qquad + \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha + \frac{1}{\delta^2}\left(1 - \frac{\lambda_{n-1}}{\delta\lambda_{n-2}}\right)\right]\|x_n - y_n\|^2, \end{cases} \tag{25}$$

or

$$\begin{cases} a_n = \|x_n - \bar{x}\|^2 + 2\lambda_{n-1}(1+\delta)\varPhi(\bar{x}, x_{n-1}) + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2, \quad n \geq 2, \\ b_n = \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \left(\varepsilon_1 + \frac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha\right]\|x_{n+1} - y_n\|^2 \\ \qquad + \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha\right]\|x_n - y_n\|^2 + \left(1 - \frac{\lambda_n}{\delta\lambda_{n-1}}\right)\|x_{n+1} - x_n\|^2. \end{cases} \tag{26}$$

Because the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ is monotonically decreasing, we have $1 - \frac{\lambda_n}{\delta\lambda_{n-1}} \geq 0$ for any $\delta \geq 1$. But for $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$, we have $\lim_{n\to+\infty}\left(1 - \frac{\lambda_n}{\delta\lambda_{n-1}}\right) = 1 - \frac{1}{\delta} < 0$. So, convergence of Algorithm 1 with $\delta < 1$ is different from that with $\delta \geq 1$, and hence cannot be established by the similar methods as in [29, 42].

Notice that $a_n \geq 0$ in (25) when $\delta \geq 1$, we take (25) to study the convergence of Algorithm 1 with $\delta \geq 1$. Consequently, a larger upper bound $\kappa(\delta)$ of α is obtained than that in [42]. While for the case of $\delta < 1$, we take (26) as $a_n \geq 0$ for all $n \geq 1$, and further investigate the properties of b_n to ensure convergence of Algorithm 1.

Below we state and prove our main convergence result of Algorithm 1 for above two different regions: $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$ and $\delta \in [1, +\infty[$.

Theorem 1 *Let $\{x_n\}$ be the sequence generated by Algorithm 1 with $\delta \in [1, +\infty[$. Then, $\{x_n\}$ converges weakly to a solution of problem (1).*

Proof. From Remark 1, we have $\lim_{n\to\infty} \lambda_n = \lambda > 0$. Then for any $\delta \in [1, +\infty[$ and $\alpha < \kappa(\delta)$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \left(\varepsilon_1 + \frac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha\right] = \frac{1}{\delta} - \left(\frac{\varepsilon_1^2+\varepsilon_2+1}{\varepsilon_1}\right)\alpha > 0,\\
&\lim_{n\to\infty} \left[\frac{\lambda_n}{\delta\lambda_{n-1}} - \frac{1}{\varepsilon_1}\left(1+\frac{1}{\varepsilon_2}\right)\alpha + \frac{1}{\delta^2}\left(1 - \frac{\lambda_{n-1}}{\delta\lambda_{n-2}}\right)\right]\\
&= \frac{1}{\delta} - \frac{1}{\varepsilon_1}\left(1+\frac{1}{\varepsilon_2}\right)\alpha + \frac{1}{\delta^2}\left(1-\frac{1}{\delta}\right) > 0.
\end{aligned}$$

Thus, there exists an integer $N > 2$, such that for any $n > N$,

$$\left.\begin{aligned}
\tfrac{\lambda_n}{\delta\lambda_{n-1}} - \left(\tfrac{\varepsilon_1^2+\varepsilon_2+1}{\varepsilon_1}\right)\alpha > 0,\\
\tfrac{\lambda_n}{\delta\lambda_{n-1}} - \tfrac{1}{\varepsilon_1}\left(1+\tfrac{1}{\varepsilon_2}\right)\alpha + \tfrac{1}{\delta^2}\left(1-\tfrac{\lambda_{n-1}}{\delta\lambda_{n-2}}\right) > 0,
\end{aligned}\right\}$$

which implies that $b_n \geq 0$ in (25) when $n > N$. Recall $1 - \frac{\lambda_{n-1}}{\delta\lambda_{n-2}} \geq 0$ for any $\delta \geq 1$, we deduce $a_n \geq 0$ in (25). Hence, by Lemma 5 and Fact 3, $\{a_n\}_{n\in\mathbb{N}}$ is convergent and $\lim_{n\to\infty} b_n = 0$. This means that $\{\|x_n - \bar{x}\|^2\}$ is bounded and so does $\{x_n\}_{n\in\mathbb{N}}$. Also, we have $\lim_{n\to\infty} \|x_{n+1} - y_n\| = 0$ and $\lim_{n\to\infty} \|x_n - y_n\| = 0$. By $\|x_{n+1} - x_n\| = \frac{1}{\delta}\|x_{n+1} - y_{n+1}\|$, we also have that $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$ and $\{y_n\}_{n\in\mathbb{N}}$ is bounded.

In what follows, we prove the sequence $\{x_n\}$ converges weakly to a solution of problem (1). For any cluster $x^* \in \mathcal{H}$ of $\{x_n\}$, there exists a subsequence $\{x_{n_k}\}$ that converges weakly to x^*, namely $x_{n_k} \rightharpoonup x^*$. It is obvious that $\{y_{n_k}\}$ also converges weakly to x^*. Next we verify that $x^* \in \mathcal{S}$. Applying Fact 1, we deduce

$$\left\langle \frac{x_{n_k+1} - x_{n_k}}{\lambda_{n_k}} + F(y_{n_k}), x - x_{n_k+1} \right\rangle \geq g(x_{n_k+1}) - g(x), \quad \forall x \in \mathcal{H}. \tag{27}$$

Letting $k \to \infty$ in (27) and using the facts $\lim_{k\to\infty} \|x_{n_k+1} - x_{n_k}\| = 0$, $g(x)$ is lower semicontinuous and $\lim_{n\to\infty} \lambda_n = \lambda > 0$, we obtain

$$\langle F(x^*), x - x^* \rangle \geq \liminf_{k\to\infty} g(x_{n_k+1}) - g(x) \geq g(x^*) - g(x), \quad \forall x \in \mathcal{H},$$

which confirms $x^* \in \mathcal{S}$.

Finally, we prove that $x_n \rightharpoonup x^*$. We take $\bar{x} = x^*$ in the definition (25) of a_n and label as a_n^*. Notice that $\{\lambda_n\}$ is bounded and $\Phi(x^*, \cdot)$ is continuous from (**A3**), we observe

$$\begin{aligned}
&\lim_{n\to\infty} a_n^* = \lim_{k\to\infty} a_{n_k+1}^*\\
&= \lim_{k\to\infty} \begin{pmatrix} \|x_{n_k+1} - x^*\|^2 + 2\lambda_{n_k}(1+\delta)\Phi(x^*, x_{n_k}) + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_k+1} - y_{n_k}\|^2 \\ + \left(1 - \frac{\lambda_{n_k}}{\delta\lambda_{n_k-1}}\right)\|x_{n_k+1} - x_{n_k}\|^2 \end{pmatrix}\\
&= 0.
\end{aligned}$$

Therefore, $\lim_{k\to\infty} \|x_n - x^*\| = 0$, which by Fact 2 shows $x_n \rightharpoonup x^*$. □

Now, we focus on convergence analysis of Algorithm 1 with $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$ and use (26). For this case, we can not establish the nonnegativity of $\{b_n\}$ and the monotonic decreasing of $\{a_n\}$ because $\frac{1}{\delta} - 1 > 0$. Consequently, convergence of $\{a_n\}_{n\in\mathbb{N}}$ can not be obtained from (24). We thus need to further investigate the sequence $\{a_n\}$ for getting a clear convergence, by using the boundedness of $\{\|x_n - x_{n-1}\|\}$ from Correction step.

First, we show that $\|x_{n+1} - x_n\| < +\infty$ when $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$ and the operator F is the gradient of a convex function $f : \mathcal{H} \to \mathbb{R}$, i.e., $F = \nabla f$. From (23) with $x = x_n$ and $\Phi(x_n, x_n) = 0$, we deduce

$$\begin{aligned}
&\left[2 - \frac{\lambda_n}{\delta\lambda_{n-1}}\right] \|x_{n+1} - x_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n+1} - y_n\|^2 \\
&\le \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right] \|y_n - x_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 \\
&\quad + \left[\left(\varepsilon_1 + \frac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right] \|x_{n+1} - y_n\|^2 + 2\lambda_n\delta\Phi(x_n, x_{n-1}).
\end{aligned}$$

Using $F = \nabla f$ and the convexity of f yields

$$\begin{aligned}
\Phi(x_n, x_{n-1}) &= \langle \nabla f(x_n), x_{n-1} - x_n\rangle + g(x_{n-1}) - g(x_n) \\
&\le \phi(x_{n-1}) - \phi(x_n),
\end{aligned}$$

where $\phi = f + g$. This together with $\lim_{n\to\infty} \lambda_n = \lambda > 0$, $\alpha < \kappa(\delta)$ and $\lambda_n \le \lambda_{n-1}$ gives us

$$\begin{aligned}
&\left(2 - \frac{1}{\delta}\right) \|x_{n+1} - x_n\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n+1} - y_n\|^2 + 2\lambda_n\delta(\phi(x_n) - \phi(\bar{x})) \\
&\le \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 + 2\lambda_{n-1}\delta(\phi(x_{n-1}) - \phi(\bar{x})), ,
\end{aligned}$$

where $\bar{x} \in \mathcal{S}$, which implies from $\phi(x_n) - \phi(\bar{x}) \ge 0$ that

$$\left(2 - \frac{1}{\delta}\right) \|x_{n+1} - x_n\|^2 \le \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{N+1} - y_N\|^2 + 2\lambda_N\delta(\phi(x_N) - \phi(\bar{x})) < +\infty, \quad \forall n > N. \tag{28}$$

That is to say, Correction step is not necessary when $F = \nabla f$, for a convex function f.

Theorem 2 *Let $\{x_n\}$ be the sequence generated by Algorithm 1 with $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$. Then, $\{x_n\}$ converges weakly to a solution of problem (1).*

Proof. Firstly, $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$ gives $\frac{1}{\delta} > \frac{\delta^2+\delta-1}{\delta^3} > 0$. Note that $\lim_{n\to\infty} \lambda_n = \lambda > 0$, by taking the limit and from $\alpha < \kappa(\delta)$, we have

$$\left.\begin{aligned}
\lim_{n\to\infty} \left[\left(\varepsilon_1 + \tfrac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}}\right] &= \left(\tfrac{\varepsilon_1^2+\varepsilon_2+1}{\varepsilon_1}\right)\alpha - \tfrac{1}{\delta} < 0, \\
\lim_{n\to\infty} \left[\tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}}\right] &= \tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{1}{\delta} < 0, \\
\lim_{n\to\infty} \left[\tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}} + \tfrac{1}{\delta^2}\left(\tfrac{\lambda_{n-1}}{\delta\lambda_{n-2}} - 1\right)\right] &= \tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{\delta^2+\delta-1}{\delta^3} < 0,
\end{aligned}\right\} \tag{29}$$

for any $\delta \in]\frac{\sqrt{5}-1}{2}, 1[$. Thus, there exists an integer $N > 2$, such that for any $n > N$,

$$\left.\begin{aligned}
\left(\tfrac{\varepsilon_1^2+\varepsilon_2+1}{\varepsilon_1}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}} &< 0, \\
\tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}} &< 0, \\
\tfrac{1}{\varepsilon_1}\left(1 + \tfrac{1}{\varepsilon_2}\right)\alpha - \tfrac{\lambda_n}{\delta\lambda_{n-1}} + \tfrac{1}{\delta^2}\left(\tfrac{1}{\delta} - 1\right) &< 0.
\end{aligned}\right\} \tag{30}$$

By $x_{n+1}-x_n=\frac{y_{n+1}-x_{n+1}}{\delta}$, Remark 1 and Lemma 5, for any $\varepsilon_1,\varepsilon_2>0$ and $M>N+1$, we have

$$\begin{aligned}
a_{M+1}-a_{N+1} &= \sum_{n=N+1}^{M}(a_{n+1}-a_n) \qquad (31)\\
&\le \sum_{n=N+1}^{M}\left[\left(\varepsilon_1+\frac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha-\frac{\lambda_n}{\delta\lambda_{n-1}}\right]\|x_{n+1}-y_n\|^2\\
&\quad+\sum_{n=N+2}^{M}\left[\frac{1}{\varepsilon_1}\left(1+\frac{1}{\varepsilon_2}\right)\alpha-\frac{\lambda_n}{\delta\lambda_{n-1}}+\frac{1}{\delta^2}\left(\frac{\lambda_{n-1}}{\delta\lambda_{n-2}}-1\right)\right]\|x_n-y_n\|^2\\
&\quad+\left[\frac{1}{\varepsilon_1}\left(1+\frac{1}{\varepsilon_2}\right)\alpha-\frac{\lambda_{N+1}}{\delta\lambda_N}\right]\|x_{N+1}-y_{N+1}\|^2+\xi_M\\
&\le \xi_M. \qquad (32)
\end{aligned}$$

where $\xi_M=\frac{1}{\delta^2}\left(\frac{\lambda_M}{\delta\lambda_{M-1}}-1\right)\|x_{M+1}-y_{M+1}\|^2<+\infty$ from Remark 1 and Correction step. This together with $a_n\ge 0$ in (26) implies that $\{a_n\}_{n\in\mathbb{N}}$ is bounded and

$$\left.\begin{aligned}
0\le -\sum_{n=N+1}^{\infty}\left[\left(\varepsilon_1+\tfrac{1+\varepsilon_2}{\varepsilon_1}\right)\alpha-\tfrac{\lambda_n}{\delta\lambda_{n-1}}\right]\|x_{n+1}-y_n\|^2<+\infty\\
0\le -\sum_{n=N+2}^{\infty}\left[\tfrac{1}{\varepsilon_1}\left(1+\tfrac{1}{\varepsilon_2}\right)\alpha-\tfrac{\lambda_n}{\delta\lambda_{n-1}}+\tfrac{1}{\delta^2}\left(\tfrac{\lambda_{n-1}}{\delta\lambda_{n-2}}-1\right)\right]\|x_n-y_n\|^2<+\infty,
\end{aligned}\right\}$$

so $\lim_{n\to\infty}\|x_{n+1}-y_n\|=0$ and $\lim_{n\to\infty}\|x_n-y_n\|=0$. By the fact $\|x_{n+1}-x_n\|=\frac{1}{\delta}\|x_{n+1}-y_{n+1}\|$, we have $\lim_{n\to\infty}\|x_{n+1}-x_n\|=0$.

Due to $\|x_n-\bar{x}\|^2\le a_n$, then $\{x_n\}_{n\in\mathbb{N}}$ is bounded. We can complete the proof by Remark 1 and the similar methods as in the proof of Theorem 1. □

Remark 3 *By the above analysis, it seems that convergence of the proposed algorithm could be still ensured without the assumption (A3), but it is not clear how to prove this as far as we known. Actually, the assumption (A3) is not restrictive, g is continuous on $\operatorname{dom} g$ when $\operatorname{dom} g$ is an open set (this includes all finite-valued functions) or $g=\delta_C$ for any closed convex set C. Moreover, (A3) holds for any separable lower semicontinuous convex function from [2, Corollary 9.15].*

3.2 Ergodic Convergence Rate for $\delta\in]\frac{\sqrt{5}-1}{2},1]$

Since there are many researches about the convergence rate when $\delta\ge 1$, we just focus on the case when $\delta\in]\frac{\sqrt{5}-1}{2},1]$. Actually, the optimal rate of convergence is $\mathcal{O}(1/n)$ for the extragradient method [35]. In this subsection, we investigate the ergodic convergence rate of the sequence $\{y_n\}_{n\in\mathbb{N}}$ for the general case (1).

From [15] and [29, Lemma 2.12], $x^*\in\mathcal{S}$ if and only if $x^*\in\operatorname{dom} g$ and

$$\max_{x\in\operatorname{dom} g}\Phi(x,x^*):=\langle F(x),x^*-x\rangle+g(x^*)-g(x)=0.$$

The following theorem shows that the above criteria can be used to find x^* under a desired accuracy.

Theorem 3 *Let $\{x_n\}$ and $\{y_n\}$ be generated by Algorithm 1. For any $n_1>N$ and a sufficiently large $J\in\mathbb{N}$ related to n_1, we define*

$$\hat{\lambda}_j=\sum_{l=n_1}^{j}\lambda_l+\delta\lambda_{n_1}\quad\text{and}\quad \hat{x}_j=\frac{1}{\hat{\lambda}_j}\left(\sum_{l=n_1+1}^{j}\lambda_l y_l+(1+\delta)\lambda_{n_1}x_{n_1}\right)$$

for any $j > J$, then $\hat{x}_j \in \operatorname{dom} g$ and

$$\Phi(x, \hat{x}_j) \leq \frac{\|x_{n_1} - \bar{x}\|^2 + \delta\lambda_{n_1}\Phi(\bar{x}, x_{n_1-1}) + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_1} - y_{n_1-1}\|^2}{2\hat{\lambda}_j}, \quad \forall x \in \mathcal{H}.$$

Proof. First of all, we have by (23) that

$$\begin{aligned}
&2\lambda_n(1+\delta)\Phi(\bar{x}, x_n) - 2\lambda_n\delta\Phi(\bar{x}, x_{n-1}) \\
&\leq \|x_n - \bar{x}\|^2 - \|x_{n+1} - \bar{x}\|^2 \\
&\quad + \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_n}{\delta\lambda_{n-1}}\right]\|x_n - y_n\|^2 - \left(1 - \frac{\lambda_n}{\delta\lambda_{n-1}}\right)\frac{1}{\delta^2}\|x_{n+1} - y_{n+1}\|^2 \\
&\quad + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_n - y_{n-1}\|^2 - \left(\frac{\lambda_n}{\delta\lambda_{n-1}} - \varepsilon_1\alpha\right)\|x_{n+1} - y_n\|^2.
\end{aligned}$$

Since $\|x_n - y_n\| \to 0$ as $n \to +\infty$, there exists a sufficiently large J such that for any $j > J$, it holds $\|x_j - y_j\| \leq \|x_{n_1} - y_{n_1}\| \neq 0$ (If $\|x_{n_1} - y_{n_1}\| = 0$, then $\|x_{n_1+1} - y_{n_1+1}\| \neq 0$, else x_{n_1+1} is a solution). So, we let $\|x_{n_1} - y_{n_1}\| \neq 0$ with $n_1 > N$. Recalling (30) we deduce for any $j > J$ that

$$\begin{aligned}
&2\left(\lambda_j(1+\delta)\Phi(\bar{x}, x_j) + \sum_{l=n_1}^{j-1}[\lambda_l(1+\delta) - \lambda_{l+1}\delta]\Phi(\bar{x}, x_l)\right) \\
&\leq \|x_{n_1} - \bar{x}\|^2 + \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_{n_1}}{\delta\lambda_{n_1-1}}\right]\|x_{n_1} - y_{n_1}\|^2 + \left(\frac{\lambda_j}{\delta\lambda_{j-1}} - 1\right)\frac{1}{\delta^2}\|x_{j+1} - y_{j+1}\|^2 \\
&\quad + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_1} - y_{n_1-1}\|^2 + \delta\lambda_{n_1}\Phi(\bar{x}, x_{n_1-1}) \\
&\leq \|x_{n_1} - \bar{x}\|^2 + \left[\frac{1}{\varepsilon_1}\left(1 + \frac{1}{\varepsilon_2}\right)\alpha - \frac{\lambda_{n_1}}{\delta\lambda_{n_1-1}} + \left(\frac{1}{\delta} - 1\right)\frac{1}{\delta^2}\right]\|x_{n_1} - y_{n_1}\|^2 \\
&\quad + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_1} - y_{n_1-1}\|^2 + \delta\lambda_{n_1}\Phi(\bar{x}, x_{n_1-1}) \\
&\leq \|x_{n_1} - \bar{x}\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_1} - y_{n_1-1}\|^2 + \delta\lambda_{n_1}\Phi(\bar{x}, x_{n_1-1}).
\end{aligned}$$

Note that the function $\Phi(\bar{x}, \cdot)$ is convex. Now, applying the Jensen's inequality to the left-hand side of the above inequality and taking

$$\lambda_j(1+\delta) + \sum_{l=n_1}^{j-1}[\lambda_l(1+\delta) - \lambda_{l+1}\delta] = \sum_{l=n_1}^{j}\lambda_l + \delta\lambda_{n_1}$$

into account, we have

$$2\left(\sum_{l=n_1}^{j}\lambda_l + \delta\lambda_{n_1}\right)\Phi(\bar{x}, \hat{x}_j) \leq \|x_{n_1} - \bar{x}\|^2 + \frac{1+\varepsilon_2}{\varepsilon_1}\alpha\|x_{n_1} - y_{n_1-1}\|^2 + \delta\lambda_{n_1}\Phi(\bar{x}, x_{n_1-1}),$$

where

$$\hat{\lambda}_j\hat{x}_j = \lambda_j(1+\delta)x_j + \sum_{l=n_1}^{j-1}[\lambda_l(1+\delta) - \lambda_{l+1}\delta]x_l = \sum_{l=n_1+1}^{j}\lambda_l y_l + (1+\delta)\lambda_{n_1}x_{n_1}.$$

Evidently, $\hat{x}_j \in \operatorname{dom} g$ which ends the proof. □

Notice that $\{\lambda_n\}$ has a lower bound $\tau > 0$ from Remark 1. Fixing $n_1 > N$, then we get $\hat{\lambda}_j \to \infty$ as $j \to \infty$. This implies $\hat{\lambda}_j \geq (j - n_1)\tau$ and Algorithm 1 has the ergodic convergence rate $\mathcal{O}(1/j)$ when $j > J$.

3.3 Heuristics on Nonmonotonic Step Sizes

Generally speaking, the variable step is more beneficial than a fixed step for the proximal gradient methods. In Algorithm 1, the step size $\{\lambda_n\}_{n\in\mathbb{N}}$ is updated but in a nonincreasing way, which might be adverse if the algorithm starts in the region with a big curvature of F. Namely, the step size in Algorithm 1 is overdependent on the initial point. For the purpose of obtaining nonmonotonic step sizes, we present an improved algorithm as follows:

Algorithm 2 (Improved PEG with nonmonotonic step size.)

Step 0. Take $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$, choose $x_0 \in \mathcal{H}$, $\lambda_0 > 0$, $\gamma \in (0,1)$, $\alpha \in]0, \kappa(\delta)[$ and a bounded sequence $\{\zeta_n\}$. Set $y_0 = x_0$, $x_1 = \text{prox}_{\lambda_0 g}(x_0 - \lambda_0 F(x_0))$ and $n = 1$. Choose $\widehat{\lambda} > 0$ and a sequence $\{\phi_n\}$ with $\phi_n \in [1, \frac{1+\delta}{\delta}]$ and $\phi_n = 1$ when $n \geq n_0$ for given n_0.

Step 1. **Prediction:**

1.a. Compute

$$y_n = x_n + \delta(x_n - x_{n-1}), \tag{33}$$

$$\lambda_n = \min \left\{ \phi_{n-1}\lambda_{n-1}, \ \frac{\alpha\|y_n - y_{n-1}\|}{\|F(y_n) - F(y_{n-1})\|}, \ \widehat{\lambda} \right\}. \tag{34}$$

1.b. Compute

$$x_{n+1} = \text{prox}_{\lambda_n g}(x_n - \lambda_n F(y_n)),$$

if $x_{n+1} = x_n = y_n$, then stop: x_{n+1} is a solution.

Step 2. **Correction:**

Check

$$\|x_{n+1} - x_n\| \leq \zeta_n,$$

if not hold, set $\lambda_n \leftarrow \gamma\lambda_n$ and return to Step 1.b.

Step 3. Set $n \leftarrow n+1$ and return to Step 1.

Since the step size is no longer monotonically decreasing, $a_n \geq 0$ in (25) is not necessarily valid when $\delta \geq 1$, so Algorithm 2 implements Correction step for any $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$. By $\phi_n \in [1, \frac{1+\delta}{\delta}]$ and $\lambda_{n+1} \leq \phi_n\lambda_n$, we can deduce $\delta\lambda_{n+1} \leq (1+\delta)\lambda_n$. Then Lemmas 3, 4 and 5 with (26) are still valid for sequences $\{x_n\}$ and $\{y_n\}$ generated by Algorithm 2.

The constant $\widehat{\lambda}$ in Algorithm 2 is given only to ensure the upper boundedness of $\{\lambda_n\}$. Hence, it makes sense to choose $\widehat{\lambda}$ quite large. In this case, the step sizes generated are allowed to increase but be bounded from Remark 1. Consequently, it follows from $\phi_n = 1$ when $n \geq n_0$ for given n_0 that the sequence $\{\lambda_n\}_{n>n_0}$ generated by Algorithm 2 is monotonically decreasing and then convergent,

$$\lim_{n\to\infty} \lambda_n > 0, \quad \lim_{n\to\infty} \frac{\lambda_n}{\lambda_{n-1}} = 1,$$

and $\frac{1}{\delta^2}\left(\frac{\lambda_n}{\delta\lambda_{n-1}} - 1\right)\|x_{n+1} - y_{n+1}\|^2 < +\infty$. Under these conditions, it is not difficult to prove the following convergence theorem by using Lemma 5 with (26), though we do not know how to choose a proper n_0.

Theorem 4 *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 2 with $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$. Then, $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a solution of problem (1).*

4 Further Discussion

From the statement above, the condition $\alpha \in]0, \kappa(\delta)[$ for any $\delta \in]\frac{\sqrt{5}-1}{2}, +\infty[$ is sufficient to ensure convergence of the proposed method. In this section, we explain by an extremely simple example that Algorithm 1 is not convergent when $\alpha \in]\frac{2}{2\delta+1}, +\infty[$ for any $\delta \in]0, +\infty[$. That is to say, we would derive an upper bound of α to guarantee the convergence of Algorithm 1, but Algorithm 1 with (δ, α) in some regions remains to be further studied, see Fig. 2.

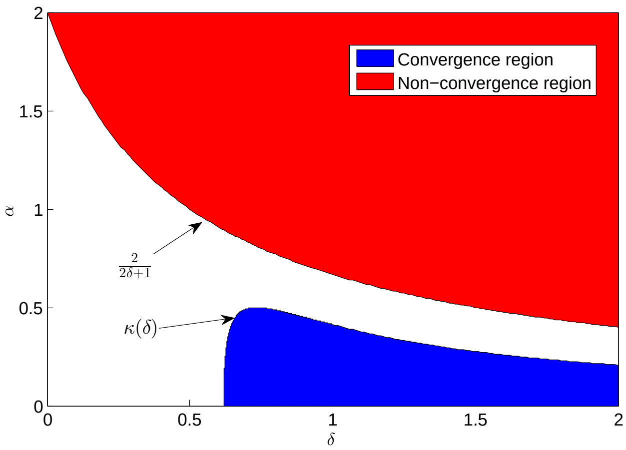


Fig. 2 The convergence and non-convergence region of the parameters δ and α.

Consider the simplest optimization problem

$$\min_{x \in \mathbb{R}^m} \frac{1}{2}\|x\|^2.$$

Obviously, it can be formulated as a special case of problem (3) with $F = I$ (the identity operator), $L = 1$ and $C = \mathbb{R}^m$. Followed by the updates of Algorithm 1, we have

$$x_{n+1} = (1 - \lambda_n - \delta\lambda_n)x_n + \delta\lambda_n x_{n-1}, \;\; \delta \in]0, +\infty[. \tag{35}$$

For any $\widetilde{d}_n$, $\widehat{d}_n \in \mathbb{R}$, if

$$\widetilde{d}_n + \widehat{d}_n = 1 - \lambda_n - \delta\lambda_n \;\; \text{and} \;\; \widetilde{d}_n\widehat{d}_n = -\delta\lambda_n, \tag{36}$$

then we can rewrite (35) as

$$x_{n+1} - \widetilde{d}_n x_n = \widehat{d}_n(x_n - \widetilde{d}_n x_{n-1}) \;\; \text{or} \;\; x_{n+1} - \widehat{d}_n x_n = \widetilde{d}_n(x_n - \widehat{d}_n x_{n-1}). \tag{37}$$

By (36) and Vieta's Theorem, we have

$$\widetilde{d}_n, \widehat{d}_n = \frac{1 - \lambda_n - \delta\lambda_n \pm \sqrt{(1 - \lambda_n - \delta\lambda_n)^2 + 4\delta\lambda_n}}{2}.$$

If $\max\left\{|\widetilde{d}_n|, \; |\widehat{d}_n|\right\} > 1$, then the iterative (37) is not convergent. As a result, (35) is not convergent either. Namely, if

$$\lambda_n > \frac{2}{2\delta + 1}, \quad \forall \delta > 0,$$

then the iterative (35) is not convergent. By Remark 1 and $L = 1$, the convergence of Algorithm 1 can not be guaranteed if $\lambda_0 > \frac{2}{2\delta+1}$ and $\alpha \in]\frac{2}{2\delta+1}, +\infty[$ for any $\delta \in]0, +\infty[$.

5 Numerical Experiments

In this section, we perform Algorithm 2 [1] (denoted by "IPEG") for solving some randomly generated minimization problems over difficult nonlinear constraints. The following state-of-the-art algorithms are compared to investigate the computational efficiency of IPEG:

– Tseng's forward-backward-forward splitting method used as in [29, Section 4] (denoted by "TFBF"), with $\beta = 0.7, \theta = 0.99$;
– Proximal extrapolated gradient methods [29, Algorithm 2] (denoted by "PEG"), with line search and $\alpha = 0.41, \sigma = 0.7$;
– Modified projected gradient method [42] (denoted by "MPG"), with $\alpha = 0.41, \delta = 1.01$.
– FISTA [33] with standard linesearch (denoted by "FISTA"), with $\beta = 0.7, \lambda_0 = 1$;

We denote the random number generator by *seed* for generating data again in Python 3.8. All experiments are performed on an Intel(R) Core(TM) i5-4590 CPU@ 3.30 GHz PC with 8GB of RAM running on 64-bit Windows operating system.

Since solutions of (1) coincide with zeros of the residual function

$$r(x, y) := \|y - \text{prox}_{\lambda_n g}(x - \lambda F(y))\| + \|x - y\|,$$

for some positive number λ, and $r_n := r(x_n, y_n) = \|x_{n+1} - y_n\| + \|x_n - y_n\| = 0$ implies $x_{n+1} = x_n = y_n$, thus we use $r_n < \epsilon$ with given $\epsilon = 10^{-6}$ to terminate our algorithms, and the same ϵ is used to terminate PEG, MPG, FB and FISTA. In particular for TFBF, we use

$$r_n := \|x_n - \text{prox}_{\lambda_n g}(x_n - \lambda F(x_n))\| \leq \epsilon$$

as in [29].

We generate λ_0 as in [13], choose y_{-1} as a small perturbation of y_0 and take $\lambda_0 = \frac{\|y_{-1} - y_0\|}{\|F(y_{-1}) - F(y_0)\|}$. This gives us an approximation of the local inverse Lipschitz constant of F at y_0. There are many choices of the sequence $\{\phi_n\}_{n\in\mathbb{N}}$, but in the earlier iterations the large range of λ_n is benefit for selecting proper step size, we thus use

$$\phi_n = \begin{cases} \frac{1+\delta}{\delta}, & \text{if } n \leq \hat{n}; \\ \frac{1+\delta+n-\widehat{n}}{\delta+n-\widehat{n}}, & \text{if } n > \widehat{n}, \end{cases} \tag{38}$$

for a given $\hat{n} \in \mathbb{N}$. In this section, we fix $\hat{n} = 500$ and $n_0 = 1000$. For applying Correction step, we use $\gamma = 0.7$ and (13) with $\zeta_{\min} = 10^{-6}$ and $\mu = \nu = 10$.

We report the number of iterations (Iter), the number of proximal operators (# prox), the number of F ($\#F$) and the computing time (Time) measured in seconds. Note that the number of iterations equals that of proximal operators for PEG and IPEG, and is 2 smaller than that of F for IPEG, we thus report the number of iterations and the number of F for PEG and only the number of iterations for IPEG. The bold letter indicates the best results in the following tables.

Problem 1 *The first problem (called Sun's problem) was considered in [28,39,42], and the Lipschitz-continuous and monotone operator was generated by*

$$F(x) = G(x) + H(x),$$

where

$$G(x) = (g_1(x), g_2(x), \ldots, g_m(x)),$$
$$g_i(x) = x_{i-1}^2 + x_i^2 + x_{i-1}x_i + x_i x_{i+1}, \quad i = 1, 2, \ldots, m, \quad x_0 = x_{m+1} = 0.$$

[1] All codes are available at http://www.escience.cn/people/changxiaokai/Codes.html

and $H(x) = Ex + c$. Here E is a square matrix $m \times m$ defined by

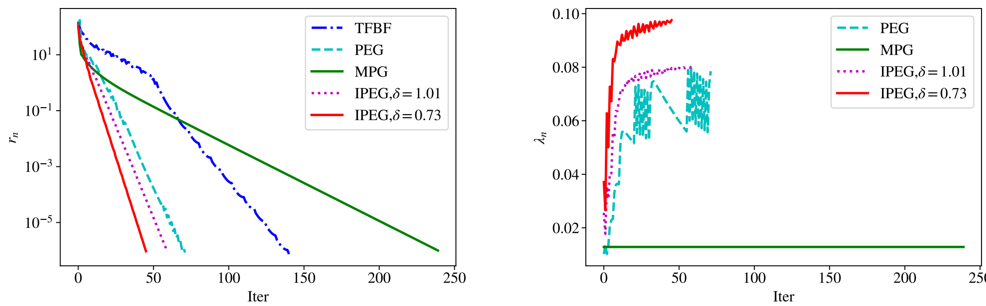
$$e_{ij} = \begin{cases} 4, & j = i \\ 1, & i - j = 1 \\ -2, & i - j = -1 \\ 0, & otherwise, \end{cases}$$

and $c = (-1, -1, \ldots, -1)$. We choose the feasible set C as $C_1 = \mathbb{R}^m_+$ and $C_2 = \{x \in \mathbb{R}^m_+ \mid \sum_{i=1}^m x_i = m\}$.

For Problem 1, the initial point x_0 is generated uniformly randomly from $[-10, 10]^d$. For every $d = 10^3, 10^4, 10^5$ and every C above, the test results are listed in Table 1. In addition, we show the evolutions of r_n and λ_n with respect to Iter for solving Problem 1 with $C = C_1$, $d = 10^3$ in Fig. 3.

Table 1 Results for Problem 1 with different d and C.

C	d	TFBF				PEG			MPG		IPEG ($\delta = 1.01$)		IPEG ($\delta = 0.73$)	
		Iter	# prox	# F	Time	Iter	# F	Time	Iter	Time	Iter	Time	Iter	Time
C_1	10^3	141	294	435	0.05	73	143	0.02	243	0.03	62	0.01	**48**	0.01
	10^4	163	341	504	0.1	76	149	0.04	262	0.03	66	0.02	**50**	0.01
	10^5	174	365	539	2.21	80	157	0.76	284	1.23	70	0.32	**53**	0.31
C_2	10^3	139	292	431	0.05	78	154	0.02	229	0.03	77	0.01	**63**	0.01
	10^4	145	305	450	0.31	83	164	0.09	249	0.24	83	0.08	**67**	0.07
	10^5	170	359	529	4.89	88	174	1.44	270	3.41	88	1.13	**71**	1.01



(a) r_n. (b) λ_n.

Fig. 3 Comparison of r_n and λ_n for solving Problem 1 with $C = C_1$, $d = 10^3$.

Problem 2 *The second test problem is the so-called Kojima-Shindo Nonlinear Complementarity Problem (NCP), considered in [30, 37], where $m = 4$ and the mapping F is defined by*

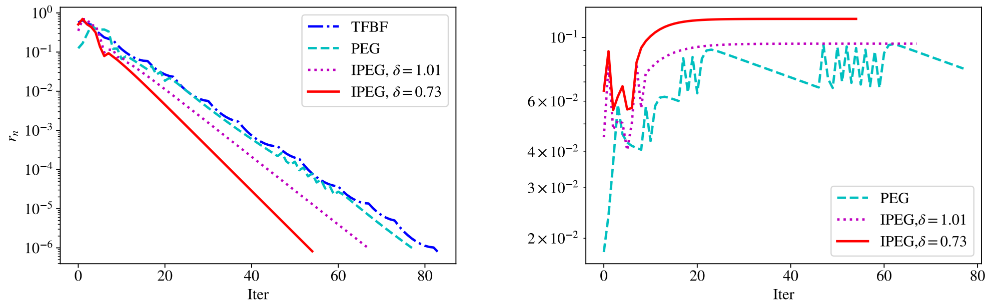
$$F(x_1, x_2, x_3, x_4) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

The feasible set is $C = \{x \in \mathbb{R}^4_+ \mid x_1 + x_2 + x_3 + x_4 = 4\}$ and $g(x) = l_C(x)$.

We choose three particular starting points: $(0, 0, 0, 0)$, $(1, 1, 1, 1)$ and $(0.5, 0.5, 2, 1)$. The numerical results are reported in Table 2 and the evolutions of r_n and λ_n with respect to Iter for solving Problem 1 with $x_0 = (1, 1, 1, 1)$ are shown in Fig. 4.

Table 2 Results for Problem 2 with different x_0.

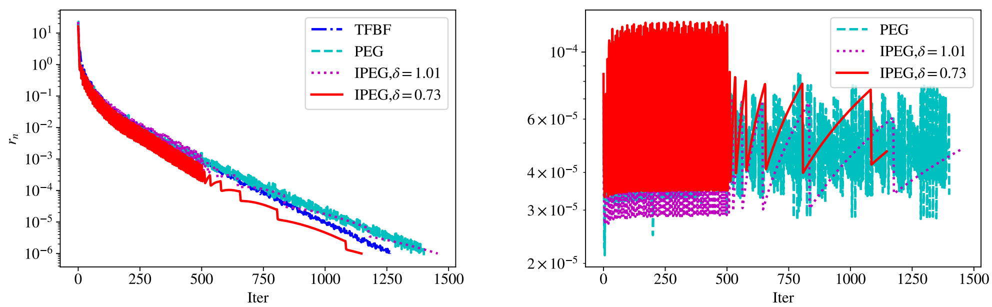
x_0	TFBF				PEG			IPEG($\delta = 1.01$)		IPEG($\delta = 0.73$)	
	Iter	# prox	# F	Time	Iter	# F	Time	Iter	Time	Iter	Time
$(0, 0, 0, 0)$	81	173	254	0.02	82	164	0.1	72	0.01	**58**	0.01
$(1, 1, 1, 1)$	84	177	261	0.02	79	156	0.1	70	0.01	**56**	0.01
$(0.5, 0.5, 2, 1)$	88	186	274	0.02	85	169	0.1	75	0.01	**59**	0.01



(a) r_n. (b) λ_n.

Fig. 4 Comparison of r_n and λ_n for solving Problem 2 with $x_0 = (1, 1, 1, 1)$.

Problem 3 *The third problem is HpHard problem, considered as in [29, 42]. Let* $F(x) = Mx + q$ *with* $M = NN^T + S + D$ *and* $q \in \mathbb{R}^m$*, where* N*,* D *and* $S \in \mathbb{R}^{m\times m}$*,* S *is a skew-symmetric matrix, every entry of* N *and* S *is uniformly generated from* $(-5, 5)$*. The matrix* D *is diagonal and its diagonal entry is uniformly generated from* $(0, 0.3)$*. Every entry of* q *is uniformly generated from* $(-500, 0)$*. The feasible set is* $C = \{x \in \mathbb{R}^m_+ \mid \sum_{i=1}^m x_i = m\}$ *and* $g(x) = l_C(x)$*.*



(a) r_n. (b) λ_n.

Fig. 5 Comparison of r_n and λ_n for solving Problem 3 with *seed* $= 1$ and $d = 500$.

Table 3 Results for Problem 3 with different cases.

seed	m	TFBF				PEG		IPEG($\delta = 1.01$)		IPEG($\delta = 0.73$)	
		Iter	# prox	# F	Time	Iter	Time	Iter	Time	Iter	Time
1	500	1066	2279	3345	0.25	1185	0.17	1185	0.13	**972**	0.09
	1000	1155	2469	3624	1.75	1323	0.93	1268	0.42	**1033**	0.39
	5000	1389	2969	4358	56.87	1575	27.89	1630	26.43	**1326**	22.86
2	500	1270	2715	3985	0.29	1447	0.19	1480	0.14	**1165**	0.12
	1000	1134	2424	3558	1.56	1274	0.86	1262	0.41	**1028**	0.40
	5000	1365	2918	4283	55.74	1554	33.91	1603	29.94	**1303**	25.64

For every m, as shown in Table 3, we have generated randomly two different M and q with *seed* $= 1$ and 2. For all tests, we take $x_0 = (1, 1, \cdots, 1)$. Since F is an affine operator, the number of iterations is 2 smaller than that of F for PEG, thus we just report the number of iterations.

Problem 4 *The fourth example is a sparse logistic regression problem for binary classification. Let* $(h_i, l_i) \in \mathbb{R}^n \times \{\pm 1\}, i = 1, \cdots, m$ *be the training set, where* $h_i \in \mathbb{R}^n$ *is the feature vector of each data sample, and* l_i *is the binary label. The formulation of sparse logistic regression reads*

$$\min_{x \in \mathbb{R}^n} \phi(x) := \mu \|x\|_1 + \frac{1}{m} \sum_{i=1}^{m} \log(1 + e^{l_i h_i^T x}), \tag{39}$$

where $\mu > 0$ *and is set to be* $0.005\|H^T l\|_\infty$ *in the numerical test.*

Let $K_{ij} = -l_i h_{ij}$ and set $\hat{f}(y) = \sum_{i=1}^{m} \log(1 + \exp(y_i))$. Then the objective in (39) is $\phi(x) = f(x) + g(x)$ with $g(x) = \mu\|x\|_1$ and $f(x) = \hat{f}(Kx)$. It is easy to derive that $L_{\nabla \hat{f}} = \frac{1}{4}$. Thus, $L_{\nabla f} = \frac{1}{4}\|K^T K\|$. We take three popular datasets from LIBSVM [2]: **w7a** with $m = 24692$, $n = 300$, **a9a** with $m = 32561$, $n = 123$ and **real-sim** with $m = 72309$, $n = 20958$.

Since f is convex and $F = \nabla f$, we apply IPEG to (39) without Correction step. We use $\epsilon = 10^{-10}$ to terminate all the algorithms for getting more accurate solution, and choose the smallest objective value among all methods and set it to $\phi(x^*)$. The results are shown in Table 4. To illustrate how does the value $\phi(x_n) - \phi(x^*)$ and r_n change over times, we give two convergence plots for data "a9a" in Fig. 6.

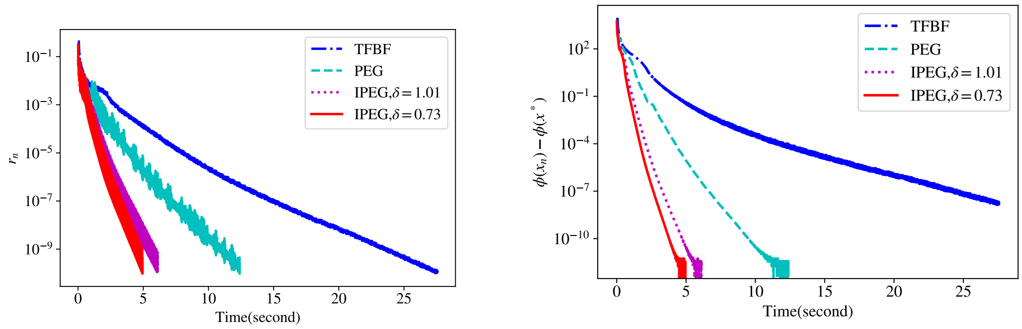
Table 4 Results for Problem 4.

data	TFBF				PEG			IPEG($\delta = 1.01$)		IPEG($\delta = 0.73$)	
	Iter	# prox	# F	Time	Iter	# F	Time	Iter	Time	Iter	Time
w7a	971	1950	2867	4.1	968	1933	2.9	827	1.6	**716**	1.4
a9a	6758	14439	21197	27.8	4241	8601	12.2	3498	6.1	**2844**	5.0
real-sim	3984	8510	12494	153.8	2651	5312	70.9	2230	35.1	**1796**	32.8

To summarize our numerical experiments on Problems 1-4, we want to make some observations. Firstly, the advantage of IPEG in comparison with other algorithms is a larger interval for possible step size λ_n, see Fig. 3(b), Fig. 4(b) and Fig. 5(b), which resulted from the proper choice of δ and the larger value of α.

Secondly, we observed that for the majority of the test problems, IPEG is more efficient than other algorithms in both the number of iterations and the CPU time. Furthermore, IPEG with $\delta = 0.73$ performs efficiently than that with $\delta = 1.01$ from the convergence plots of r_n shown in Fig. 3(a), Fig. 4(a) and Fig. 5(a), which is extremely due to the larger step size λ_n and the use of only one value of the mapping required per iteration. Although linesearch is involved in Correction step, the condition required is so weak that the linesearch is not started for many problems.

[2] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/



(a) r_n. (b) λ_n.

Fig. 6 Comparison of r_n and $\phi(x_n) - \phi(x^*)$ for solving Problem 4 with data "a9a".

In addition, since MPG [42] adopted nonincreasing step sizes, it is adverse when starting in the region with a big curvature of F, see Fig. 3(b) and the results of MPG for Problem 1. From Fig. 5, the step sizes generated by IEPG have fluctuated within a range at the first 500 iterations, after that the range decreases as we use (38) with $\widehat{n} = 500$ to control the increase of step sizes.

6 Conclusions

Without the knowledge of Lipschitz constant, we have proposed a proximal extrapolated gradient method using a prediction-correction procedure to determine stepsizes, and improved it numerically with non-monotonic step size. The method extended the range of parameters (considering the case of $\delta < 1$) and obtained a larger step size than the existing methods by using correction step. Finally, a number of experiments illustrate that the proposed method is efficient, and the improvement can be resulted from the larger step size.

In addition, we have shown by an extremely simple example that our method is not convergent if $\lambda_0,\ \alpha \in]\frac{2}{2\delta+1}, +\infty[$ for any $\delta > 0$. From Fig. 3, the convergence of the proposed method remains unknown for (δ, α) in some regions. Especially for $\delta \in]0, \frac{\sqrt{5}-1}{2}]$, it remains to be explored whether there are any (larger) $\alpha > 0$ such that Algorithms 1 and 2 are convergent. Perhaps our method without the correction step is convergent as well, and can be generalized to other methods that need to estimate the Lipschitz constant. We leave this as an interesting topic for our future research.

Acknowledgements The research of Xiaokai Chang was supported by the Hongliu Foundation of First-class Disciplines of Lanzhou University of Technology. The project was supported by the National Natural Science Foundation of China under Grant 61877046 and the Natural Science Basic Research Plan in Shaanxi Province of China (2017JM1014).

References

1. Antipin, A.S.: On a method for convex programs using a symmetrical modification of the Lagrange function. Ekonomika i Matematicheskie Metody, 12(6), 1164–1173 (1976)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer Berlin, New York (2011)
3. Bertsekas, D.P., Gafni, E.M.: Projection methods for variational inequalities with applications to the traffic assignment problem. Math. Program. Study, 17, 139–159 (1982)
4. Burachik, R.S., Lopes, J.O., Svaiter, B.F.: An outer approximation method for the variational inequality problem. SIAM J. Control Optim. 43(6), 2071–2088 (2005)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011)

6. Bot R.I., Csetnek, E.R.: Forward-backward and Tseng's type penalty schemes for monotone inclusion problems. Set-Valued Var. Anal. 22, 313–331 (2014)
7. Bot R.I., Csetnek, E.R.: An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. Numer. Algor. 71, 519–540 (2016)
8. Chang, X., Liu, S., Zhao, P., Li, X.: Convergent prediction-correction-based ADMM for multi-block separable convex programming. J. Comput. Appl. Math. 335, 270–288 (2018)
9. Chang, X., Liu, S., Zhao, P., Song, D.: A generalization of linearized alternating direction method of multipliers for solving two-block separable convex programming. J. Comput. Appl. Math. 357, 251–272 (2019)
10. Censor,Y., Gibali,A., Reich,S.: The subgradient extragradient method for solving variational inequalities in Hilbert space. J. Optim. Theory Appl. 148, 318–335 (2011)
11. Combettes, P.L., Băng, C.V.: Variable metric Forward-backward splitting with applications to monotone inclusions in duality. Optimization, 63(9), 1289–1318 (2014)
12. Cruz, J.B., Nghia, T.: On the convergence of the proximal forward-backward splitting method with linesearches. arXiv:1501.02501 (2015)
13. Denisov, S., Semenov, V., Chabak, L.: Convergence of the modified extragradient method for variational inequalities with non-Lipschitz operators. Cybern. Syst. Anal. 51, 757–765 (2015)
14. Ekeland, I., Temam R.: Convex Analysis and Variational Problems. North-Holland, Amsterdam, Holland (1976)
15. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problem. Springer-Verlag, New York (2003)
16. Gibali, A.: A new non-Lipschitzian projection method for solving variational inequalities in Euclidean spaces. J. Nonlinear Anal. Optim. 6, 41–51 (2015)
17. He, B., Yuan, X.: A class of ADMM-based algorithms for multi-block separable convex programming. Comput. Optim. Appl. 70(3), 791–826 (2018)
18. Huang, Y., Dong, Y.: New properties of forward-backward splitting and a practical proximal-descent algorithm. Appl. Math. Comput. 237, 60–68 (2014)
19. Iusem, A.N., Svaiter, B.F.: A variant of Korpelevich's method for variational inequalities with a new search strategy. Optimization, 42, 309–321 (1997)
20. Iusem, A.N., Pérez, L.R.: An extragradient-type algorithm for nonsmooth variational inequalities. Optimization, 48, 309–332 (2000)
21. Korpelevich, G.M.: The extragradient method for finding saddle points and other problem. Ekonomika i Matematicheskie Metody, 12, 747–756 (1976)
22. Khobotov, E.N.: Modification of the extragradient method for solving variational inequalities and certain optimization problems. USSR Comp. Math. Phys. 27, 120–127 (1987)
23. Lorenz, D., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vis. 51, 311–325 (2015)
24. Lyashko, S.I., Semenov, V.V., Voitova, T.A.: Low-cost modification of Korpelevich's method for monotone equilibrium problems. Cybern. Syst. Anal. 47, 631–639 (2011)
25. Liang, J., Fadili, J., Peyré, G.: Activity identification and local linear convergence of forward-backward-type methods. SIAM J. Optim. 27(1), 408–437 (2017)
26. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. 16(6), 964–979 (1979)
27. Malitsky, Y.V., Semenov, V.V.: An extragradient algorithm for monotone variational inequalities. Cybern. Syst. Anal. 50, 271–277 (2014)
28. Malitsky, Y.V.: Projected reflected gradient methods for variational inequalities. SIAM J. Optim. 25(1), 502–520 (2015)
29. Malitsky, Y.V.: Proximal extrapolated gradient methods for variational inequalities. Optim. Methods Soft. 33(1), 140–164 (2018)
30. Mainge, P.E., Gobinddass, M.L.: Convergence of one-step projected gradient methods for variational inequalities. J. Optim. Theory Appl. 171, 146–168 (2016)
31. Monteiro, R.D., Svaiter, B.F.: Complexity of variants of Tseng's modified FB splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J. Optim. 21, 1688–1720 (2011)
32. Noor, M.A.: Modified projection method for pseudomonotone variational inequalities. Appl. Math. Lett. 15, 315–320 (2002)
33. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. Soviet Mathematics Doklady, 27(2), 372–376 (1983)
34. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers, Boston (2004)
35. Nemirovski, A.: Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. 15, 229–251 (2004)
36. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. 73(4), 591–597 (1967)
37. Pang, J.S., Gabriel, S.A.: NE/SQP: A robust algorithm for the nonlinear complementarity problem. Math. Program. 60(1-3), 295–337 (1993)

38. Solodov, M.V., Svaiter, B.F.: A new projection method for monotone variational inequalities. SIAM J. Control Optim. 37, 765–776 (1999)
39. Sun, D.: A projection and contraction method for the nonlinear complementarity problems and its extensions. Math. Numer. Sinica, 16, 183–194 (1994)
40. Solodov, M.V., Svaiter, B.F.: A new projection method for variational inequality problems. SIAM J.Control Optim. 37, 765–776 (1999)
41. Tseng, P.: A modified forward-backward splitting method for maximal monotone mapping. SIAM J. Control Optim. 38, 431–446 (2000)
42. Yang J., Liu H.: A modified projected gradient method for monotone variational inequalities. J. Optim. Theory Appl. 179(1), 197–211 (2018)