J. Nonlinear Funct. Anal. 2023 (2023) 16 https://doi.org/10.23952/jnfa.2023.16

MATHRES

Journal of Nonlinear Functional Analysis
Available online at http://jnfa.mathres.org

SELF-ADAPTIVE ALGORITHMS FOR SOLVING CONVEX BILEVEL OPTIMIZATION PROBLEMS

BOWEN ZHAO, PEICHAO DUAN*

College of Science, Civil Aviation University of China, Tianjin 300300, China

Abstract. In this paper, we present a self-adaptive algorithm and an inertial version for solving convex bilevel optimization problems. We establish the strong convergence of our proposed algorithms. The step-sizes in our algorithms for the inner level optimization problem are selected without prior knowledge of operator norms. A numerical experiment is included to illustrate the performances of our algorithms and some comparisons are present with related algorithms.

Keywords. Convex bilevel optimization; Inertial acceleration; Moreau-Yosida approximate; Self-adaptive algorithm; Split proximal algorithm.

1. INTRODUCTION

Bilevel optimization theory is more and more widely used in many disciplines, including aircraft conflicts [2], railway transport hub planning [11], strategic pricing in competitive electricity markets [8] and so on. The research on bilevel optimization has both irreplaceable significance and extensive prospect; see, e.g., [7, 16] and references therein. Convex bilevel optimization is a kind of important bilevel optimization problem, and a number of authors investigated various methods for solving convex bilevel optimization problems, which can be found in [5, 6, 20, 23].

Convex bilevel optimization problems consist of two convex optimization problems: the inner and the outer levels. In [20], Sabach and Shtern investigated a type of convex bilevel optimization problems. The outer level is given by the following constrained minimization problem

$$\min_{x\in\Omega} \omega(x), \tag{1.1}$$

where ω is a strongly convex and differentiable function defined on $\Omega \neq \emptyset$ and Ω is the set of minimizers of the unconstrained convex optimization problem

$$\min_{x\in H} g(x) + f(x), \tag{1.2}$$

*Corresponding author.
E-mail address: pcduancauc@126.com (P. Duan).
Received October 25, 2022; Accepted April 16, 2023.

where $g, f \in \Gamma_0(H)$ and f is differentiable. The notation $\Gamma_0(H)$ denotes the space of functions that are proper, convex, and lower semicontinuous in a real Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$.

However, function f sometimes does not have differentiability. In this paper, we study the following general form of inner problem

$$\min_{x \in H_1} g(x) + f(Ax), \tag{1.3}$$

where $g \in \Gamma_0(H_1)$, $f \in \Gamma_0(H_2)$, and $A : H_1 \to H_2$ is a bounded operator. It is worth noting that neither f nor g may be differentiable. As far as we know, Moreau-Yosida approximate is a useful tool that makes them obtain differentiability. For this reason, we transform (1.3) to the following problem

$$\min_{x \in H_1} g(x) + f_\lambda(Ax), \tag{1.4}$$

where $f_\lambda(y) = \min_{u \in H_2} \{ f(u) + \frac{1}{2\lambda} \|u - y\|^2 \}$ stands for the Moreau-Yosida approximate of f with parameter λ.

In recent years, numerous researchers have studied iterative methods for the inner level optimization problem. In their studies, various methods were proposed to obtain the solutions of inner level optimization problem. The corresponding weak or strong convergence results were discussed; see, e.g., [14, 19, 21] and the references therein. Especially, if A is an identity operator and f is differential, a number of researchers proposed the proximal gradient algorithm to solve problem (1.2); see, e.g., [1, 4, 25] and the references therein.

Actually, the inner problem can be transformed into the zero point problem of two monotone operators from first order optimality condition. In order to make the iterative algorithms easy to implement, it is transformed into the fixed point form of the operator. Note that the optimality condition of (1.4) is as follows:

$$0 \in \partial (g(x^*) + f_\lambda(Ax^*)) = \partial g(x^*) + A^* (\frac{I - prox_{\lambda f}}{\lambda}) Ax^*.$$

An equivalent fixed point formulation of (1.4) was deduced in [14]

$$x^* = prox_{\lambda \mu g}(I - \mu A^*(I - prox_{\lambda f})A)x^*$$

for $\lambda > 0$ and $\mu > 0$.

In 2017, Sabach and Shtern [20] proposed an iterative algorithm for solving convex bilevel optimization problems (1.1) and (1.2). They established a first-order method based on the existing fixed-point algorithm.

$$\begin{cases} s_n = prox_{\mu g}(I - \mu \nabla f)x_n, \\ z_n = x_n - \gamma \nabla \omega(x_n), \\ x_{n+1} = \beta_n z_n + (1 - \beta_n)s_n, n \geq 1, \end{cases}$$

where $\mu \in (0, \frac{2}{L_f})$, $\gamma \in (0, \frac{2}{L_\omega + \varpi}]$, and L_f, L_ω are the Lipschitz constants for the gradients of f, ω, respectively. The parameter sequence $\{\beta_n\} \subset (0,1)$ satisfies

$$\lim_{n \to \infty} \beta_n = 0, \ \sum_{n=1}^{\infty} \beta_n = \infty.$$

In 2019, Shehu and Vuong [22] devised the following iterative scheme, which generates a sequence via

$$\begin{cases} y_n = x_n + \delta_n(x_n - x_{n-1}), \\ s_n = prox_{\lambda g}(I - \lambda \nabla f)y_n, \\ z_n = y_n - \gamma \nabla \omega(y_n), \\ x_{n+1} = \beta_n z_n + (1-\beta_n)s_n, n \geq 1. \end{cases}$$

They gave convergence analysis of the inertial algorithm for convex bilevel optimization problems under appropriate conditions.

On the other hand, to avoid computing the norm of the operator involved since it is difficult to calculate or estimate, some self-adaptive algorithms were introduced; see e.g. [3, 12, 17, 18, 24]. In 2014, Moudafi and Thakur [14] proposed a self-adaptive split proximal algorithm for solving (1.3). The implementation of their algorithm does not need any prior information about the operator norm. They proved that the sequence $\{x_n\}$ weakly converges to a solution of (1.3). They computed $\{x_n\}$ via the rule

$$x_{n+1} = prox_{\lambda \mu_n g}(I - \mu_n A^*(I - prox_{\lambda f} A)x_n),$$

where the stepsize $\mu_n = \tau_n \frac{h(x_n)+l_{\mu_n}(x_n)}{\theta^2(x_n)}$ with $0 < \tau_n < 4$.

Mainly based on the research work [14, 20, 22], we investigate the adaptive selection of the step-size parameters and inertial acceleration of the algorithm for solving convex bilevel optimization problems. Under suitable conditions, we prove that our algorithms converge strongly to some solution of the inner problem, which is the unique solution of a variational inequality problem (corresponding the outer level problem). In addition, a numerical example is performed to illustrate performance and some comparisons with related algorithms are presented to demonstrate the efficiency of our algorithms.

2. Preliminaries

Throughout this paper, we denote by $Fix(U)$ the set of fixed points of U and by I the identity operator. $x_n \to x$ (resp., $x_n \rightharpoonup x$) indicates that $\{x_n\}$ is strongly (resp., weakly) convergent to x. Given a sequence $\{x_n\}$, we use $\omega_w(x_n)$ to denote the weak ω-limit set of $\{x_n\}$, that is, $\omega_w(x_n) := \{x \mid \exists \{x_{n_j}\} \subset \{x_n\} \textit{ such that } x_{n_j} \rightharpoonup x\}$. In this section, we collect some important definitions and lemmas which will be used in the next section.

Definition 2.1. A mapping $U : H \to H$ is said to be

(i) nonexpansive iff $\|Ux - Uy\| \leq \|x - y\|$ for all $x, y \in H$;

(ii) firmly nonexpansive iff $2U - I$ is nonexpansive, or equivalently,

$$\langle x - y, Ux - Uy \rangle \geq \|Ux - Uy\|^2, \ \forall x, y \in H.$$

As we know, projection operators are firmly nonexpansive.

(iii) ρ-contractive iff there exists $\rho \in [0, 1)$ such that

$$\|Ux - Uy\| \leq \rho \|x - y\|, \ \forall x, y \in H.$$

(iv) L-Lipschizian iff there exists a constant $L > 0$ such that

$$\|Ux - Uy\| \leq L\|x - y\|, \ \forall x, y \in H.$$

Lemma 2.2. *Let C be a nonempty, convex, and closed subset of a real Hilbert space H. Given $x \in H$ and $z \in C$, $y = P_C x$ if and only if $\langle x-y, y-z\rangle \geq 0$ for all $z \in C$.*

Lemma 2.3. *Let H be a real Hilbert space. Then the following statements hold.*

(i) $\|x+y\|^2 = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2, \ \forall x,y \in H.$
(ii) $\|x+y\|^2 \leq \|x\|^2 + 2\langle x+y,y\rangle, \ \forall x,y \in H.$
(iii) $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2$ *for all* $\alpha \in \mathbb{R}$ *and* $x,y \in H$.

Definition 2.4. (see [13]) The proximal operator of $g \in \Gamma_0(H)$ is defined by

$$prox_g(x) = arg\min_{u\in H}\{g(u) + \frac{1}{2}\|u-x\|^2\}, \quad x \in H.$$

Lemma 2.5. *The proximal identity*

$$prox_{\lambda g}(x) = prox_{\mu g}(\frac{\mu}{\lambda}x + (1-\frac{\mu}{\lambda})prox_{\lambda g}x) \tag{2.1}$$

holds for $g \in \Gamma_0(H)$, $\lambda > 0$ and $\mu > 0$.

Lemma 2.6. *(Demiclosedness principle, see [9]) Let H be a real Hilbert space, and let $T : H \to H$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. If $\{x_n\}$ is a sequence in H weakly converging to x and if $\{(I-T)x_n\}$ converges strongly to y, then $(I-T)x = y$. In particular, if $y = 0$, then $x \in Fix(T)$.*

Lemma 2.7. *(see [10]) Assume that $\{b_n\}$ is a sequence of nonnegative real numbers such that*

$$b_{n+1} \leq (1-\sigma_n)b_n + \sigma_n a_n, \ n \geq 1,$$
$$b_{n+1} \leq b_n - \eta_n + \varphi_n, \quad n \geq 1,$$

where $\{\sigma_n\}$ is a sequence in $(0,1)$, $\{\eta_n\}$ is a sequence of nonnegative real numbers and $\{a_n\}$ and $\{\varphi_n\}$ are two sequences in $\mathbb{R}$ such that

(i) $\sum_{n=0}^{\infty} \sigma_n = \infty$,
(ii) $\lim_{n\to\infty} \varphi_n = 0$,
(iii) $\lim_{k\to\infty} \eta_{n_k} = 0$ *implies* $\limsup_{k\to\infty} a_{n_k} \leq 0$ *for any subsequence* $\{n_k\} \subset \{n\}$.

Then $\lim_{n\to\infty} b_n = 0$.

3. MAIN RESULTS

This section is dedicated to the self-adaptive selection of parameters, algorithmic acceleration and convergence analysis for solving convex bilevel optimization problems (1.1) and (1.3). For this purpose, we first list some assumption conditions in Assumption 3.1.

Assumption 3.1:

(a) $g \in \Gamma_0(H_1)$, $f \in \Gamma_0(H_2)$, and $A : H_1 \to H_2$ is a bounded linear operator.
(b) $\Omega = \{x^* | argmin\ g \cap A^{-1}(argmin\ f)\} \neq \emptyset$.
(c) $\omega : H_1 \to H_1$ is strongly convex with parameter $\varpi > 0$ and continuously differentiable such that its gradient is Lipschitz continuous with constant L_ω.

The basic algorithm for solving inner level problem is the split proximal algorithm, the method can also be seen as a fixed point algorithm where the iterated mapping is given by

$$T_\mu := prox_{\lambda\mu g}(I - \mu A^*(I - prox_{\lambda f})A). \tag{3.1}$$

Put

$$S_\gamma := I - \gamma\nabla\omega. \tag{3.2}$$

Supposing that Assumption 3.1 (c) holds, we obtain S_γ is a ρ-contraction mapping for all $\gamma \in (0, \frac{2}{L_\omega+\varpi}]$ and $\rho := \sqrt{1-\frac{2\gamma\varpi L_\omega}{L_\omega+\varpi}}$ (This result was obtained in [20]).

In this section, we propose a self-adaptive proximal split algorithm and an inertial version for approximating the unique fixed point of the following variational inequality problem:

$$\langle (I-S_\gamma)x^*, \tilde{x}-x^*\rangle \geq 0, \quad \forall \tilde{x} \in \Omega. \tag{3.3}$$

Set $\theta(x) = \sqrt{\|\nabla h(x)\|^2 + \|\nabla l_{\mu_n}(x)\|^2}$ with $h(x) = \frac{1}{2}\|(I-prox_{\lambda f})Ax\|^2$ and $l_{\mu_n}(x) = \frac{1}{2}\|(I-prox_{\mu_n\lambda g})x\|^2$.

First, we propose the following scheme for solving the bilevel optimization problem.

Algorithm 1
Step 0. Input $\beta \geq 3$, $\lambda > 0$, and $\gamma \in (0, \frac{2}{L_\omega+\varpi}]$. Give $\varepsilon > 0$, $\delta > 0$, and an initial point $x_1 \in H_1$. Set $n := 1$.
Step 1. Give x_n and compute

$$\begin{cases} s_n = T_{\mu_n}x_n, \\ z_n = S_\gamma x_n, \\ x_{n+1} = \beta_n z_n + (1-\beta_n)s_n, \end{cases} \tag{3.4}$$

where the stepsize

$$\mu_n = \tau_n \frac{h(x_n)+l_{\mu_n}(x_n)}{\theta^2(x_n)} \tag{3.5}$$

with $\delta \leq \tau_n \leq \frac{4h(x_n)}{h(x_n)+l_{\mu_n}(x_n)} - \delta$.
Step 2. If $h(x_n)+l_{\mu_n}(x_n) < \varepsilon$, then the iterative process stops. Otherwise, set $n := n+1$ and go to Step 1.

Lemma 3.1. *Let $\{\mu_n\}$ be generated by (3.5), then $\{\mu_n\}$ has a lower bound.*

Proof. By the definitions of $h(x)$ and $l_{\mu_n}(x)$, we obtain $\nabla h(x) = A^*(I-prox_{\lambda f})Ax$ and $\nabla l_{\mu_n}(x) = (I-prox_{\mu_n\lambda g})x$, respectively. We have

$$\|\nabla h(x)\|^2 = \|A^*(I-prox_{\lambda f})Ax\|^2 \leq \|A\|^2\|(I-prox_{\lambda f})Ax\|^2 = 2\|A\|^2h(x), \tag{3.6}$$

and

$$\|\nabla l_{\mu_n}(x)\|^2 = \|(I-prox_{\mu_{n_k}\lambda g})x\|^2 = 2l_{\mu_n}(x). \tag{3.7}$$

Putting together (3.6) and (3.7), we arrive at

$$\tau_n \frac{h(x_n)+l_{\mu_n}(x_n)}{\theta^2(x_n)} \geq \delta \frac{h(x_n)+l_{\mu_n}(x_n)}{2(\|A\|^2+1)(h(x_n)+l_{\mu_n}(x_n))} = \frac{\delta}{2(\|A\|^2+1)} > 0. \tag{3.8}$$

Setting $\mu_0 := \frac{\delta}{2(\|A\|^2+1)}$, we have $\mu_n \geq \mu_0 > 0$. □

Theorem 3.2. *Let H_1 and H_2 be two real Hilbert spaces and Assumption 3.1 holds. Assume that T_μ and S_γ are defined by (3.1) and (3.2), respectively. Given $x_1 \in H$ arbitrarily, suppose that*

(i) $\beta_n \in (0,1)$, $\lim_{n\to\infty}\beta_n = 0$ *and* $\sum_{n=1}^{\infty}\beta_n = \infty$;
(ii) $\delta \leq \tau_n \leq \frac{4h(x_n)}{h(x_n)+l_{\mu_n}(x_n)} - \delta$ *for some* $\delta > 0$ *small enough;*
(iii) $\lambda > 0$ *and* $\gamma \in (0, \frac{2}{L_\omega+\varpi}]$.

Then $\{x_n\}$ generated by Algorithm 1 converges strongly to x^*, where x^* is a solution to (1.3). And x^* is also the unique solution of the variational inequality

$$\langle \nabla \omega(x^*), \tilde{x} - x^* \rangle \geq 0, \forall\, \tilde{x} \in \Omega. \tag{3.9}$$

Proof. The proof is divided into three steps.

Step 1. Prove that $\{x_n\}$ is bounded.

For any $p \in \Omega$, we have $p \in Fix(T_{\mu_n})$. Using the nonexpasiveness of $prox_{\lambda\mu_n g}$, we obtain

$$\begin{aligned}\|s_n - p\|^2 &= \|prox_{\lambda\mu_n g}(I - \mu_n A^*(I - prox_{\lambda f})A)x_n - prox_{\lambda\mu_n g}(I - \mu_n A^*(I - prox_{\lambda f})A)p\|^2 \\ &\leq \|x_n - \mu_n A^*(I - prox_{\lambda f})Ax_n - (p - \mu_n A^*(I - prox_{\lambda f})Ap)\|^2 \\ &= \|x_n - p\|^2 - 2\mu_n \langle \nabla h(x_n), x_n - p\rangle + \mu_n^2 \|\nabla h(x_n)\|^2. \end{aligned} \tag{3.10}$$

Since minimizers of the convex function $h(x_n)$ are exactly zero points of its gradient mapping, using the definition of adjoint operators and the firmly nonexpansiveness of $I - prox_{\lambda f}$, we have

$$\langle \nabla h(x_n), x_n - p\rangle = \langle (I - prox_{\lambda f})Ax_n, Ax_n - Ap\rangle \geq \|(I - prox_{\lambda f})Ax_n\|^2 = 2h(x_n). \tag{3.11}$$

It follows from (3.5), (3.10), and (3.11) that

$$\begin{aligned}\|s_n - p\|^2 &\leq \|x_n - p\|^2 - 4\mu_n h(x_n) + \tau_n^2 \frac{(h(x_n) + l_{\mu_n}(x_n))^2}{\theta^4(x_n)} \|\nabla h(x_n)\|^2 \\ &= \|x_n - p\|^2 - 4\tau_n \frac{h(x_n) + l_{\mu_n}(x_n)}{\theta^2(x_n)} h(x_n) + \tau_n^2 \frac{(h(x_n) + l_{\mu_n}(x_n))^2}{\theta^2(x_n)} \frac{\|\nabla h(x_n)\|^2}{\theta^2(x_n)} \\ &\leq \|x_n - p\|^2 - \tau_n\Big(\frac{4h(x_n)}{h(x_n) + l_{\mu_n}(x_n)} - \tau_n\Big) \frac{(h(x_n) + l_{\mu_n}(x_n))^2}{\theta^2(x_n)}. \end{aligned} \tag{3.12}$$

By condition (ii), we derive

$$\|s_n - p\| \leq \|x_n - p\|. \tag{3.13}$$

According to the iterative process, we have

$$\begin{aligned}\|x_{n+1} - p\| &\leq \beta_n \|S_\gamma x_n - S_\gamma p\| + \beta_n \|S_\gamma p - p\| + (1 - \beta_n)\|s_n - p\| \\ &\leq (1 - \beta_n(1 - \rho))\|x_n - p\| + \beta_n(1 - \rho) \frac{\|S_\gamma p - p\|}{1 - \rho}. \end{aligned} \tag{3.14}$$

Then by the mathematical induction, it is easy to obtain

$$\|x_{n+1} - p\| \leq \max\{\|x_1 - p\|, \; \frac{\|S_\gamma p - p\|}{1 - \rho}\}.$$

Therefore, the sequence $\{x_n\}$ is bounded, so are $\{z_n\}$ and $\{s_n\}$.

Step 2. Prove that $\lim_{k\to\infty} \eta_{n_k} = 0$ implies $\lim_{k\to\infty} h(x_{n_k}) = 0$ and $\lim_{k\to\infty} l_{\mu_{n_k}}(x_{n_k}) = 0$ for any sequence $\{n_k\} \subset \{n\}$. The expression of η_n is under inequality (3.17).

Firstly, fixing $p \in \Omega$, we find from (3.4), (3.13), and Schwarz's inequality that

$$\begin{aligned}\|x_{n+1} - p\|^2 &= \beta_n^2 \|z_n - p\|^2 + (1 - \beta_n)^2 \|s_n - p\|^2 + 2\beta_n(1 - \beta_n)\langle z_n - p, s_n - p\rangle \\ &= \beta_n^2 \|S_\gamma x_n - S_\gamma p + S_\gamma p - p\|^2 + (1 - \beta_n)^2 \|s_n - p\|^2 \\ &\quad + 2\beta_n(1 - \beta_n)\langle S_\gamma x_n - S_\gamma p + S_\gamma p - p, s_n - p\rangle \\ &\leq 2\beta_n^2(\rho^2 \|x_n - p\|^2 + \|S_\gamma p - p\|^2) + (1 - \beta_n)^2 \|x_n - p\|^2 \end{aligned}$$

$$
\begin{aligned}
&+2\beta_n(1-\beta_n)(\rho\|x_n-p\|\|s_n-p\|+\langle S_\gamma p-p,s_n-p\rangle)\\
&\le 2\beta_n^2(\rho^2\|x_n-p\|^2+\|S_\gamma p-p\|^2)+(1-\beta_n)^2\|x_n-p\|^2\\
&+2\beta_n(1-\beta_n)\rho\|x_n-p\|^2+2\beta_n(1-\beta_n)\langle S_\gamma p-p,s_n-p\rangle\\
&\le (1-\beta_n(2-\beta_n(1+2\rho^2)-2\rho(1-\beta_n)))\|x_n-p\|^2+2\beta_n^2\|S_\gamma p-p\|^2\\
&+2\beta_n(1-\beta_n)\langle S_\gamma p-p,s_n-p\rangle. \qquad (3.15)
\end{aligned}
$$

Setting $b_n := \|x_n-p\|^2$,

$$\sigma_n := \beta_n(2-\beta_n(1+2\rho^2)-2\rho(1-\beta_n)),$$

and

$$a_n := \frac{2(\beta_n\|S_\gamma p-p\|^2+(1-\beta_n)\langle S_\gamma p-p,s_n-p\rangle)}{2-\beta_n(1+2\rho^2)-2\rho(1-\beta_n)},$$

we deduce from the above inequality that

$$b_{n+1}\le(1-\sigma_n)b_n+\sigma_n a_n. \qquad (3.16)$$

Applying Lemma 2.3 (i), (iii), and (3.12), we obtain

$$
\begin{aligned}
&\|x_{n+1}-p\|^2\\
&\le \beta_n\|z_n-p\|^2+(1-\beta_n)\|s_n-p\|^2\\
&=\beta_n\|S_\gamma x_n-S_\gamma p+S_\gamma p-p\|^2+(1-\beta_n)\|s_n-p\|^2\\
&\le \beta_n(\rho^2\|x_n-p\|^2+2\langle z_n-S_\gamma p,S_\gamma p-p\rangle+\|S_\gamma p-p\|^2)+(1-\beta_n)\|s_n-p\|^2\\
&\le \beta_n(\rho^2\|x_n-p\|^2+2\langle z_n-p+p-S_\gamma p,S_\gamma p-p\rangle+\|S_\gamma p-p\|^2)+(1-\beta_n)\|s_n-p\|^2\\
&\le \beta_n(\rho^2\|x_n-p\|^2+2\langle z_n-p,S_\gamma p-p\rangle-\|S_\gamma p-p\|^2)+(1-\beta_n)\|s_n-p\|^2\\
&\le \|x_n-p\|^2-(1-\beta_n)\tau_n(\frac{4h(x_n)}{h(x_n)+l_{\mu_n}(x_n)}-\tau_n)\frac{(h(x_n)+l_{\mu_n}(x_n))^2}{\theta^2(x_n)}+2\beta_n\langle z_n-p,S_\gamma p-p\rangle.
\end{aligned}
\qquad (3.17)
$$

Setting

$$\eta_n := (1-\beta_n)\tau_n(\frac{4h(x_n)}{h(x_n)+l_{\mu_n}(x_n)}-\tau_n)\frac{(h(x_n)+l_{\mu_n}(x_n))^2}{\theta^2(x_n)}$$

and

$$\varphi_n := 2\beta_n\langle z_n-p,S_\gamma p-p\rangle,$$

we obtain the second inequality required in Lemma 2.7

$$b_{n+1}\le b_n-\eta_n+\varphi_n. \qquad (3.18)$$

Since $\sum_{n=1}^\infty \beta_n=\infty$, it is easy to obtain $\sum_{n=1}^\infty \sigma_n=\infty$. By the boundedness of $\{z_n\}$ and $\beta_n\to 0$, we have $\lim\limits_{n\to\infty}\varphi_n=0$. The sequences of parameters $\{\sigma_n\}$ and $\{\varphi_n\}$ satisfy the conditions (i) and (ii) of Lemma 2.7. In order to complete the proof, it suffices to verify that $\eta_{n_k}\to 0(k\to\infty)$ implies that $\limsup_{k\to\infty}a_{n_k}\le 0$ for any subsequence $\{n_k\}\subset\{n\}$. Indeed, if $\eta_{n_k}\to 0$, then

$$\frac{(h(x_{n_k})+l_{\mu_{n_k}}(x_{n_k}))^2}{\theta^2(x_{n_k})}\to 0.$$

From the boundedness of the function $\theta^2(x_{n_k})$, we have $h(x_{n_k})+l_{\mu_{n_k}}(x_{n_k})\to 0$. Since both $h(x_{n_k})$ and $l_{\mu_{n_k}}(x_{n_k})$ are nonnegative, we further obtain

$$h(x_{n_k})=\frac{1}{2}\|(I-prox_{\lambda f})Ax_{n_k}\|^2\to 0 \tag{3.19}$$

and

$$l_{\mu_{n_k}}(x_{n_k})=\frac{1}{2}\|(I-prox_{\mu_{n_k}\lambda g})x_{n_k}\|^2\to 0. \tag{3.20}$$

Step 3. Prove that

$$\omega_w(x_{n_k})\subset\Omega. \tag{3.21}$$

Since $\{x_{n_k}\}$ is bounded, we see that $\omega_w(x_{n_k})\neq\emptyset$. Take $\bar{x}\in\omega_w(x_{n_k})$ and assume that $\{x_{n_{k_j}}\}$ is a subsequence of $\{x_{n_k}\}$ weakly converging to $\bar{x}$. Without loss of generality, we still use $\{x_{n_k}\}$ to denote $\{x_{n_{k_j}}\}$. By the lower semi-continuity of h, we have $0\le h(\bar{x})\le\liminf_{k\to\infty}h(x_{n_k})=0$. Thus we obtain

$$h(\bar{x})=\frac{1}{2}\|(I-prox_{\lambda f})A\bar{x}\|^2=0. \tag{3.22}$$

Choosing a fixed positive constant μ, using the proximal identity of Lemma 2.5 and the nonexpansiveness of proximal operators, we deduce that

$$\begin{aligned}
l_{\mu}(x_{n_k})&=\frac{1}{2}\|(I-prox_{\lambda\mu_{n_k}g})x_{n_k}+(prox_{\lambda\mu_{n_k}g}-prox_{\lambda\mu g})x_{n_k}\|^2\\
&\le\|(I-prox_{\lambda\mu_{n_k}g})x_{n_k}\|^2+\|(prox_{\lambda\mu_{n_k}g}-prox_{\lambda\mu g})x_{n_k}\|^2\\
&=2l_{\mu_{n_k}}(x_{n_k})+\|prox_{\lambda\mu g}(\frac{\mu}{\mu_{n_k}}x_{n_k}+(1-\frac{\mu}{\mu_{n_k}})prox_{\lambda\mu_{n_k}g}x_{n_k})-prox_{\lambda\mu g}x_{n_k}\|^2\\
&\le 2l_{\mu_{n_k}}(x_{n_k})+\|\frac{\mu}{\mu_{n_k}}x_{n_k}+(1-\frac{\mu}{\mu_{n_k}})prox_{\lambda\mu_{n_k}g}x_{n_k}-x_{n_k}\|^2\\
&\le 2l_{\mu_{n_k}}(x_{n_k})+(1-\frac{\mu}{\mu_{n_k}})^2\|prox_{\lambda\mu_{n_k}g}x_{n_k}-x_{n_k}\|^2\\
&\le 2l_{\mu_{n_k}}(x_{n_k})+2(1-\frac{\mu}{\mu_{n_k}})^2 l_{\mu_{n_k}}(x_{n_k}).
\end{aligned} \tag{3.23}$$

By Lemma 3.1, $1-\frac{\mu}{\mu_0}\le 1-\frac{\mu}{\mu_{n_k}}\le 1$. From (3.23) and (3.20), we conclude $l_{\mu}(x_{n_k})\to 0$ as $k\to\infty$. It follows that

$$l_{\mu}(\bar{x})=\frac{1}{2}\|(I-prox_{\mu\lambda g})\bar{x}\|^2=0. \tag{3.24}$$

Combining (3.22) and (3.24), we obtain $\bar{x}\in\Omega$, which yields $\omega_w(x_{n_k})\subset\Omega$. Since $prox_{\lambda\mu_{n_k}g}$ is nonexpansive, we have

$$\begin{aligned}
\|T_{\mu_{n_k}}x_{n_k}-x_{n_k}\|&\le\|T_{\mu_{n_k}}x_{n_k}-prox_{\lambda\mu_{n_k}g}x_{n_k}\|+\|prox_{\lambda\mu_{n_k}g}x_{n_k}-x_{n_k}\|\\
&\le\|\mu_{n_k}A^*(I-prox_{\lambda}f)Ax_{n_k}\|+\|prox_{\lambda\mu_{n_k}g}x_{n_k}-x_{n_k}\|\\
&\le\mu_{n_k}\|A\|\cdot\|(I-prox_{\lambda f})Ax_{n_k}\|+\|prox_{\lambda\mu_{n_k}g}x_{n_k}-x_{n_k}\|
\end{aligned} \tag{3.25}$$

According to (3.19) and (3.20), we obtain

$$\|T_{\mu_{n_k}}x_{n_k}-x_{n_k}\|\to 0 \tag{3.26}$$

as $k \to \infty$. Since S_γ is a contraction and P_Ω is a nonexpansive mapping, we obtain that $P_\Omega S_\gamma$ is also contractive. Hence there exists the unique fixed point x^* of $P_\Omega S_\gamma$ ($x^* = P_\Omega S_\gamma x^*$), and $x^* \in \Omega$ is the projection of $S_\gamma x^*$ onto Ω. By Lemma 2.2, x^* is the unique solution of the following variational inequality problem

$$\langle S_\gamma x^* - x^*, \tilde{x} - x^* \rangle \leq 0, \forall\, \tilde{x} \in \Omega. \tag{3.27}$$

Meanwhile, by invoking (3.26) and Schwarz's inequality, we have

$$\begin{aligned}
&\limsup_{k\to\infty} \langle S_\gamma x^* - x^*, s_{n_k} - x^* \rangle \\
\leq &\limsup_{k\to\infty} \langle S_\gamma x^* - x^*, x_{n_k} - x^* \rangle + \lim_{k\to\infty} \|T_{\mu_{n_k}} x_{n_k} - x_{n_k}\| \cdot \|S_\gamma x^* - x^*\| \\
= &\limsup_{k\to\infty} \langle S_\gamma x^* - x^*, x_{n_k} - x^* \rangle
\end{aligned}$$

Now, we can take subsequence $\{x_{n_{k_i}}\}$ of $\{x_{n_k}\}$ such that $\{x_{n_{k_i}}\} \rightharpoonup \hat{x}$ as $i \to \infty$ and

$$\limsup_{k\to\infty} \langle S_\gamma x^* - x^*, x_{n_k} - x^* \rangle = \lim_{i\to\infty} \langle S_\gamma x^* - x^*, x_{n_{k_i}} - x^* \rangle = \langle S_\gamma x^* - x^*, \hat{x} - x^* \rangle.$$

Since $\hat{x} \in \omega_w(x_{n_k}) \subset \Omega$ and x^* is the solution of the variational inequality problem (3.27), we conclude that

$$\limsup_{k\to\infty} \langle S_\gamma x^* - x^*, s_{n_k} - x^* \rangle \leq \langle S_\gamma x^* - x^*, \hat{x} - x^* \rangle \leq 0.$$

Hence, $\limsup_{k\to\infty} a_{n_k} \leq 0$ combing conditions (i). Applying Lemma 2.7 yields that $\{x_n\}$ converges strongly to a point x^* in Ω. Indeed, using the fact that $S_\gamma = I - \gamma \nabla \omega$, we see that (3.27) is equivalent to

$$\langle x^* - (x^* - \gamma \nabla \omega(x^*)), \tilde{x} - x^* \rangle = \gamma \langle \nabla \omega(x^*), \tilde{x} - x^* \rangle \geq 0, \quad \forall\, \tilde{x} \in \Omega,$$

which directly implies that (3.9) holds true due to $\gamma > 0$. This means that x^* satisfies the first order optimality condition and therefore x^* is also the optimal solution to problem (1.1). This completes the proof. □

Next, we introduce an inertial acceleration version for solving convex bilevel optimization based on the inertial technique proposed by Nesterov [15].

Algorithm 2

Step 0. Input $\beta \geq 3$, $\lambda > 0$ and $\gamma \in (0, \frac{2}{L_\omega + \varpi}]$. Give $\varepsilon > 0$, $\delta > 0$, and two initial points $x_0,\ x_1 \in H_1$. Set $n := 1$.

Step 1. Given x_{n-1}, x_n, compute

$$\begin{cases}
y_n = x_n + \Delta_n (x_n - x_{n-1}), \\
s_n = T_{\mu_n} y_n, \\
z_n = S_\gamma y_n, \\
x_{n+1} = \beta_n z_n + (1 - \beta_n) s_n,
\end{cases} \tag{3.28}$$

where Δ_n satisfies $0 \leq |\Delta_n| \leq \overline{\Delta}_n$ with $\overline{\Delta}_n$ defined by

$$\overline{\Delta}_n = \begin{cases}
min\{\dfrac{n-1}{n+\beta-1}, \dfrac{\varepsilon_n}{\|x_n - x_{n-1}\|}\}, & x_n \neq x_{n-1}, \\
\dfrac{n-1}{n+\beta-1}, & x_n = x_{n-1}
\end{cases}$$

and the stepsize $\mu_n = \tau_n \frac{h(y_n)+l_{\mu_n}(y_n)}{\theta^2(y_n)}$ with $\delta \le \tau_n \le \frac{4h(y_n)}{h(y_n)+l_{\mu_n}(y_n)} - \delta$.

Step 2. If $h(y_n)+l_{\mu_n}(y_n) < \varepsilon$, then the iterative process stops. Otherwise, set $n := n+1$ and go to Step 1.

Theorem 3.3. *Let H_1 and H_2 be two real Hilbert spaces and Assumption 3.1 holds. Assume that T_μ and S_γ are defined by (3.1) and (3.2), respectively. Given $x_0, x_1 \in H_1$ arbitrarily, suppose that*

(i) $\beta_n \in (0,1)$, $\lim_{n\to\infty} \beta_n = 0$ *and* $\sum_{n=1}^{\infty} \beta_n = \infty$;

(ii) $\delta \le \tau_n \le \frac{4h(y_n)}{h(y_n)+l_{\mu_n}(y_n)} - \delta$ *for some* $\delta > 0$ *small enough;*

(iii) $\lambda > 0$ *and* $\gamma \in (0, \frac{2}{L_\omega+\varpi}]$;

(iv) $\varepsilon_n = o(\beta_n), i.e., \lim_{n\to\infty} \frac{\varepsilon_n}{\beta_n} = 0$.

Then $\{x_n\}$ generated by Algorithm 2 converges strongly to x^*, where x^* is a solution to (1.3), and x^* is also the unique solution of the variational inequality

$$\langle \nabla\omega(x^*), \tilde{x} - x^* \rangle \ge 0, \forall\, \tilde{x} \in \Omega. \tag{3.29}$$

Proof. The proof is divided into three steps.

Step 1. Prove that $\{x_n\}$ is bounded.

For any $p \in \Omega$, we have

$$\|y_n - p\| = \|x_n + \Delta_n(x_n - x_{n-1}) - p\| \le \|x_n - p\| + |\Delta_n| \cdot \|x_n - x_{n-1}\|. \tag{3.30}$$

Using a technique similar to Theorem 3.2, we have

$$\|s_n - p\|^2 \le \|y_n - p\|^2 - 2\mu_n \langle y_n - p, \nabla h(y_n)\rangle + \mu_n^2 \|\nabla h(y_n)\|^2 \tag{3.31}$$

and

$$\langle y_n - p, \nabla h(y_n)\rangle = \langle (I - prox_{\lambda f})Ay_n, Ay_n - Ap\rangle \ge \|(I - prox_{\lambda f})Ay_n\|^2 = 2h(y_n). \tag{3.32}$$

It follows that

$$\|s_n - p\|^2 \le \|y_n - p\|^2 - \tau_n\left(\frac{4h(y_n)}{h(y_n)+l_{\mu_n}(y_n)} - \tau_n\right)\frac{(h(y_n)+l_{\mu_n}(y_n))^2}{\theta^2(x_n)}. \tag{3.33}$$

Condition (ii) together with the nonnegativity of $\frac{(h(y_n)+l_{\mu_n}(y_n))^2}{\theta^2(x_n)}$ ensures that

$$\|s_n - p\| \le \|y_n - p\|. \tag{3.34}$$

Combining (3.30) and (3.34), we obtain

$$\begin{aligned}
\|x_{n+1} - p\| &\le \beta_n \|S_\gamma y_n - S_\gamma p\| + \beta_n \|S_\gamma p - p\| + (1-\beta_n)\|s_n - p\| \\
&\le (1 - \beta_n(1-\rho))\|y_n - p\| + \beta_n \|S_\gamma p - p\| \\
&\le (1 - \beta_n(1-\rho))\|x_n - p\| + \beta_n(1-\rho)\frac{\|S_\gamma p - p\| + |\Delta_n| \cdot \|x_n - x_{n-1}\|/\beta_n}{1-\rho}.
\end{aligned} \tag{3.35}$$

Using the definition of $\overline{\Delta}_n$, condition (iv), and the mathematical induction, we conclude that $\{x_n\}$ is bounded.

Step 2. Prove that $\lim_{k\to\infty} \eta_{n_k} = 0$ implies $\lim_{k\to\infty} h(y_{n_k}) = 0$ and $\lim_{k\to\infty} l_{\mu_{n_k}}(y_{n_k}) = 0$ for any sequence $\{n_k\} \subset \{n\}$. The expression of η_n is under inequality (3.39).

Firstly, fixing $p \in \Omega$, we have from Lemma 2.3 (ii) that

$$\|y_n - p\|^2 \le \|x_n - p\|^2 + 2\langle x_n - p + \Delta_n(x_n - x_{n-1}), \Delta_n(x_n - x_{n-1})\rangle$$

$$\leq \|x_n - p\|^2 + 2(\|x_n - p\| + |\Delta_n| \cdot \|x_n - x_{n-1}\|)|\Delta_n| \cdot \|x_n - x_{n-1}\|. \tag{3.36}$$

According to (3.36) and condition (iv), we obtain similar results to those in Theorem 3.2.

$$\begin{aligned}
\|x_{n+1} - p\|^2 &\leq (1 - \beta_n(2 - \beta_n(1 + 2\rho^2) - 2\rho(1 - \beta_n)))\|y_n - p\|^2 \\
&\quad + 2\beta_n^2\|S_\gamma p - p\|^2 + 2\beta_n(1 - \beta_n)\langle S_\gamma p - p, s_n - p\rangle \\
&\leq (1 - \beta_n(2 - \beta_n(1 + 2\rho^2) - 2\rho(1 - \beta_n)))\|x_n - p\|^2 + 2M|\Delta_n| \cdot \|x_n - x_{n-1}\| \\
&\quad + 2\beta_n^2\|S_\gamma p - p\|^2 + 2\beta_n(1 - \beta_n)\langle S_\gamma p - p, s_n - p\rangle \\
&\leq (1 - \beta_n(2 - \beta_n(1 + 2\rho^2) - 2\rho(1 - \beta_n)))\|x_n - p\|^2 + 2M\varepsilon_n \\
&\quad + 2\beta_n^2\|S_\gamma p - p\|^2 + 2\beta_n(1 - \beta_n)\langle S_\gamma p - p, s_n - p\rangle,
\end{aligned} \tag{3.37}$$

where $M := \sup_{n\geq 1}\{\|x_n - p\| + |\Delta_n| \cdot \|x_n - x_{n-1}\|\}$. Set $b_n := \|x_n - p\|^2$,

$$\sigma_n := \beta_n(2 - \beta_n(1 + 2\rho^2) - 2\rho(1 - \beta_n)),$$

and

$$a_n := \frac{2(M\varepsilon_n/\beta_n + \beta_n\|S_\gamma p - p\|^2 + (1 - \beta_n)\langle S_\gamma p - p, s_n - p\rangle)}{2 - \beta_n(1 + 2\rho^2) - 2\rho(1 - \beta_n)},$$

we deduce from the above inequality that

$$b_{n+1} \leq (1 - \sigma_n)b_n + \sigma_n a_n. \tag{3.38}$$

Combining (3.28), (3.33), and (3.36), we have

$$\begin{aligned}
\|x_{n+1} - p\|^2 &\leq \|y_n - p\|^2 - (1 - \beta_n)\tau_n\left(\frac{4h(y_n)}{h(y_n) + l_{\mu_n}(y_n)} - \tau_n\right)\frac{(h(y_n) + l_{\mu_n}(y_n))^2}{\theta^2(y_n)} \\
&\quad + 2\beta_n\langle z_n - p, S_\gamma p - p\rangle \\
&\leq \|x_n - p\|^2 - (1 - \beta_n)\tau_n\left(\frac{4h(y_n)}{h(y_n) + l_{\mu_n}(y_n)} - \tau_n\right)\frac{(h(y_n) + l_{\mu_n}(y_n))^2}{\theta^2(y_n)} \\
&\quad + 2M\varepsilon_n + 2\beta_n\langle z_n - p, S_\gamma p - p\rangle.
\end{aligned} \tag{3.39}$$

Set

$$\eta_n := (1 - \beta_n)\tau_n\left(\frac{4h(y_n)}{h(y_n) + l_{\mu_n}(y_n)} - \tau_n\right)\frac{(h(y_n) + l_{\mu_n}(y_n))^2}{\theta^2(y_n)}$$

and

$$\varphi_n := 2M\varepsilon_n + 2\beta_n\langle z_n - p, S_\gamma p - p\rangle,$$

we have

$$b_{n+1} \leq b_n - \eta_n + \varphi_n. \tag{3.40}$$

It is easy to verify that $\{\sigma_n\}$ and $\{\varphi_n\}$ satisfy the conditions (i) and (ii) of Lemma 2.7. Assume that $\eta_{n_k} \to 0$. With similar processes, we can derive that $\lim_{k\to\infty} h(y_{n_k}) = 0$ and $\lim_{k\to\infty} l_{\mu_{n_k}}(y_{n_k}) = 0$.

Step 3. Prove that

$$\omega_w(x_{n_k}) \subset \Omega. \tag{3.41}$$

Take $\bar{x} \in \omega_w(x_{n_k})$. Since $\|y_{n_k} - x_{n_k}\| = |\Delta_{n_k}| \cdot \|x_{n_k} - x_{n_{k-1}}\| \leq \varepsilon_{n_k} \to 0$ as $k \to \infty$, we obtain $\bar{x} \in \omega_w(y_{n_k})$ immediately. The following proof can be carried out by a similar argument to Step 3 in Theorem 3.2. Applying Lemma 2.7 yields that $\{x_n\}$ converges strongly to a point x^* in Ω. This completes the proof. □

4. NUMERICAL ILLUSTRATIONS

In this section, we present numerical comparisons with our proposed algorithms and that of the algorithm (called Bilevel Gradient Sequential Averaging Method, abbreviated as BiG-SAM) in [20]. Our numerical experiments are coded using software MATLAB.

Example 4.1. Let $H_1 = \mathbb{R}^m$ and $H_2 = \mathbb{R}^k$. The inner objective function is taken here as

$$F(x) := \delta_C(x) + \delta_Q(Ax),$$

where $A \in \mathbb{R}^{k\times m}$ is a random matrix whose elements are normally distributed, δ_C and δ_Q denote the indicator functions of two nonempty closed convex sets C, Q of H_1 and H_2, respectively. δ_C is defined as $\delta_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise. Problem (1.3) reduces to

$$\min_{x\in H_1} F(x) = \min_{x\in C}\{\frac{1}{2}\|(I-P_Q)Ax\|^2\}.$$

We take the outer objective function as $\omega(x) = \frac{1}{2}x^T Px$ with P being a given positive definite matrix.

In this example, let $C = \{x \in H_1 | \|x\| \leq r\}$ with a random $r > 0$, $Q = \{y \in H_2 | \langle a, y\rangle \leq b\}$ with a random vector $a \in H_2$ and a random $b \in \mathbb{R}$. Set the parameters $\lambda = 1,$, $\gamma = \frac{2}{L_\omega+\varpi}$, $\beta_n = \frac{1}{n+1}$, $\varepsilon_n = \beta_n/n^{0.1}$ (Algorithm 2), and $\mu = 1/L_f$(BIG-SAM). For simplicity, we set $P = I$. Take arbitrary $x_1, x_0 \in H_1$. Tables 1 and 2 demonstrate that Algorithm 1 and Algorithm 2 are better than BiG-SAM in terms of averaged running time and number of iterations under different error limitations. Because Algorithm 1 and Algorithm 2 do not require to calculate or estimate the norm of A, they have obvious advantages in running time.

TABLE 1. Comparison of Algorithm 1, Algorithm 2 and BIG-SAM. $\beta = 3, \quad \varepsilon = 10^{-3}$.

	Algorithm 1		Algorithm 2		BiG-SAM	
Parameters	Iterations	$Time(s)$	Iterations	$Time(s)$	Iterations	$Time(s)$
m=1000,k=300	5	0.2989	5	0.2031	4	1.3428
m=2000,k=500	6	1.5020	5	1.4197	13	2.4682
m=5000,k=1000	8	1.5909	8	1.4577	15	3.8438
m=10000,k=2000	9	1.7500	11	1.5725	12	12.8906
m=20000,k=2000	8	5.5781	7	5.0844	17	30.2500

In Example 4.1, settomg $m = 10000, k = 3000$, we compare the running time of the three algorithms in the case of different ε. From Figure 1, one sees that the test values of our proposed algorithms are faster to reach the stop criteria, especially Algorithm 2 with inertia step.

Funding
This paper was supported by the Innovation and Entrepreneurship Training Program for College Students of Civil Aviation University of China (Grant No. 202210059021).

Acknowledgements
The authors would like to thank the referee for valuable suggestions to improve the manuscript.

TABLE 2. Comparison of Algorithm 1, Algorithm 2 and BIG-SAM. $\beta = 3, \quad \varepsilon = 10^{-5}$.

	Algorithm 1		Algorithm 2		BiG-SAM	
Parameters	Iterations	$Time(s)$	Iterations	$Time(s)$	Iterations	$Time(s)$
m=1000,k=300	5	0.5033	5	0.3781	6	2.3973
m=2000,k=500	8	2.5702	8	1.53702	9	3.2823
m=5000,k=1000	8	5.0076	7	4.1271	8	8.8536
m=10000,k=2000	11	6.5862	10	6.3281	12	14.3095
m=20000,k=2000	16	7.4688	12	6.8312	18	39.5938

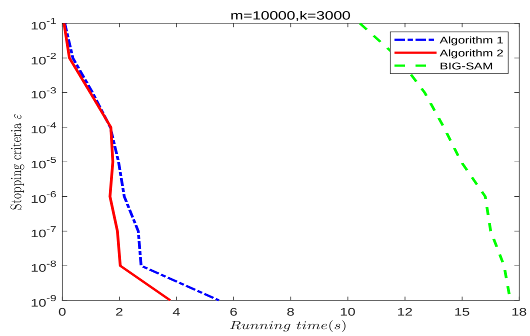

FIGURE 1. Comparison of running time of three algorithms

REFERENCES

[1] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009) 183-202.
[2] M. Cerulli, C. D'Ambrosio, L. Liberti, M. Pelegrín, Detecting and solving aircraft conflicts using bilevel programming, J. Global Optim. 81 (2021) 529-557.
[3] L.C. Ceng, A modified inertial subgradient extragradient method for solving pseudomonotone variational inequalities and common fixed point problems, Fixed Point Theory 21 (2020), 93-108.
[4] P. Duan, X. Zheng, Bounded perturbation resilience and superiorization techniques for a modified proximal gradient method, Optimization 69 (2020) 1219-1235.
[5] P. Duan, Y. Zhang, Alternated and multi-step inertial approximation methods for solving convex bilevel optimization problems, Optimization DOI: 10.1080/02331934.2022.2069022.
[6] S. Dempe, N. Dinh, J. Dutta, Optimality conditions for a simple convex bilevel programming problem, in Variational Analysis and Generalized Differentiation in Optimization and Control, R.S. Burachik and J.C. Yao, (eds). pp. 149-161, New York, Springer, 2010.
[7] G. Eichfelder, Multiobjective bilevel optimization, Math. Program. 123 (2010) 419-449.
[8] M. Fampa, L.A. Barroso, D. Candal, L. Simonetti, Bilevel optimization applied to strategic pricing in competitive electricity markets, Comput. Optim. Appl. 39 (2008) 121-142.
[9] K. Geobel, W.A. Kirk, Topics in Metric Fixed Point Theory, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1990.
[10] S. He, C. Yang, Solving the variational inequality problem defined on intersection of finite level sets, Abstr. Appl. Anal. 2013 (2013) 942315.
[11] A.I. Kibzun, A. Naumov, S.V. Ivanov, Bilevel optimization problem for railway transport hub planning, UBS 38 (2012) 140-160.

[12] L. Liu, S.Y. Cho, J.C. Yao, Convergence analysis of an inertial Tseng's extragradient algorithm for solving pseudomonotone variational inequalities and applications, J. Nonlinear Var. Anal. 5 (2021), 627-644.
[13] J.J. Moreau, Proprietes des applications 'prox', C.R. Acad. Sci. Paris Sér. A Math. 256 (1963) 1069-1071.
[14] A. Moudafi, B.S. Thakur, Solving proximal split feasibility problems without prior knowledge of operator norm, Optim. Lett. 8 (2014) 2099-2110.
[15] Y. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Dokl. Akad. Nauk SSSR 269 (1983) 543-547.
[16] E.H. Neto, A.D. Pierro, On perturbed steepest descent methods with inexactline search for bilevel convex optimization, Optimization 60 (2011) 991-1008.
[17] G.N. Ogwo, T.O. Alakoya, O.T. Mewomo, Iterative algorithm with self-adaptive step size for approximating the common solution of variational inequality and fixed point problems, Optimization DOI: 10.1080/02331934.2021.1981897.
[18] M.A. Olona, T.O. Alakoya, A.O.E. Owolabi, O.T. Mewomo, Inertial shrinking projection algorithm with self-adaptive step size for split generalized equilibrium and fixed point problems for a countable family of nonexpansive multivalued mappings, Demonstr. Math. 54 (2021) 47-67.
[19] X. Qin, A. Petrusel, J.C. Yao, CQ iterative algorithms for fixed points of nonexpansive mappings and split feasibility problems in Hilbert spaces, J. Nonlinear Convex Anal. 19 (2018) 157-165.
[20] S. Sabach, S. Shtern, A first order method for solving convex bilevel optimization problems, SIAM J. Optim. 27 (2017) 640-660.
[21] Y. Shehu, F.U. Ogbuisi, Convergence analysis for proximal split feasibility problems and fixed point problems, J. Appl. Math. Comput. DOI: 10.1007/s12190-014-0800-7.
[22] Y. Shehu, P.T. Vuong, A. Zemkoho, An inertial extrapolation method for convex simple bilevel optimization, Optim. Methods Softw. doi: 10.1080/10556788.2019.1619729.
[23] M. Solodov, An explicit descent method for bilevel convex optimization, J. Convex Anal. 14 (2007), 227-238.
[24] B. Tan, X. Qin, S.Y. Cho, Revisiting subgradient extragradient methods for solving variational inequalities, Numer. Algo. 90 (2022), 1593-1615.
[25] H.K. Xu, Properties and iterative methods for the lasso and its variants, Chin. Ann. Math. 35 (2014) 1-18.