CARPATHIAN J. MATH.
Volume **39** (2023), No. 1,
Pages 315 - 323

Online version at https://semnul.com/carpathian/
Print Edition: ISSN 1584 - 2851; Online Edition: ISSN 1843 - 4401
DOI: https://doi.org/10.37193/CJM.2023.01.22

In memoriam Professor Charles E. Chidume (1947- 2021)

Stepsize Choice for Korpelevich's and Popov's Extragradient Algorithms for Convex-Concave Minimax Problems

JIAOJIAO WANG and HONG-KUN XU

ABSTRACT. We show that the choice of stepsize in Korpelevich's extragradient algorithm is sharp, while the choice of stepsize in Popov's extragradient algorithm can be relaxed. We also extend Korpelevich's extragradient algorithm and Popov's extragradient algorithm (with larger stepsize) to the infinite-dimensional Hilbert space framework, with weak convergence.

1. INTRODUCTION

Consider a convex-concave minimax problem

$$\min_{x\in C}\max_{y\in D} f(x,y), \tag{1.1}$$

where C and D are nonempty closed convex subsets of Hilbert spaces H_1 and H_2, respectively, and the objective function $f : C\times D\to\mathbb{R}$ is convex-concave, that is, (i) $f(\cdot,y)$ is convex for each fixed $y\in D$, and (ii) $f(x,\cdot)$ is concave for each fixed $x\in C$. Set $E := C\times D$ and $H := H_1\times H_2$. Let $E^* := C^*\times D^*$ denote the set of solutions of (1.1) (i.e., set of saddle points of f). Assume $E^*\neq\emptyset$. In addition, assume f is continuously Fréchet differentiable. Recall that the (saddle) gradient ∂f of f, which is denoted by g, is defined as

$$g(z)\equiv\partial f(z)=\begin{bmatrix}\nabla_x f(x,y)\\ -\nabla_y f(x,y)\end{bmatrix}. \tag{1.2}$$

Recall also that f is said to be L-smooth for some $L\geq 0$ if the gradient g is L-Lipschitz continuous, that is,

$$\|g(z)-g(z')\|\leq L\|z'-z\|,\quad z,z'\in E.$$

It is known that g is monotone and a point $z^*=(x^*,y^*)\in E$ is a solution of (1.1) (i.e., a saddle point of f) if and only if z^* is a solution to the variational inequality (VI):

$$\langle g(z^*),z-z^*\rangle\geq 0,\quad z\in E. \tag{1.3}$$

Note that VI (1.3) is equivalent to the fixed point equation

$$z^*=P_E(z^*-\lambda g(z^*)) \tag{1.4}$$

for every $\lambda>0$. Here P_E is the nearest point projection from H to E defined by

$$P_E z=\arg\min_{w\in E}\|w-z\|^2,\quad z\in H.$$

Received: 14.06.2022. In revised form: 27.07.2022. Accepted: 28.07.2022
2010 *Mathematics Subject Classification.* 47J25, 47H10, 90C47.
Key words and phrases. Extragradient algorithm, minimax, projection, inverse strongly monotone, nonexpansive, variational inequality, fixed point.
Corresponding author: Hong-Kun Xu; xuhk@hdu.edu.cn

Recently a lot of attention has been paid to algorithmic approaches to minimax problems (1.1), due to applications to machine learning (see [1, 3, 5, 10] and references therein).

Two classic methods, known as extragradient (EG) algorithms, are due to Korpelevich [4] and Popov [7].

1.1. **Korpelevich's Extragradient Algorithm.** In [4], Korpelevich introduced the following extragradient (EG) algorithm:

$$\begin{cases} \bar{z}_n = P_E(z_n - \alpha g(z_n)) & \text{(1.5a)} \\ z_{n+1} = P_E(z_n - \alpha g(\bar{z}_n)) & \text{(1.5b)} \end{cases}$$

where $z_0 \in E$ is an arbitrarily chosen initial point and $\alpha > 0$ is the stepsize. Korpelevich proved the following convergence result, where the stepsize $\alpha > 0$ is selected such that $\alpha < 1/L$.

Theorem 1.1. [4, Theorem 1] *Suppose $H_1 = \mathbb{R}^{d_1}$ and $H_2 = \mathbb{R}^{d_2}$ are Euclidean d_1- and d_2-spaces, respectively, and f is convex-concave and L-smooth. Suppose, in addition, the stepsize α is chosen in such a range that $0 < \alpha < \frac{1}{L}$. Then the sequence $\{z_n\}$ generated by* (1.5) *converges to a saddle point of f.*

1.2. **Popov's Extragradient Algorithm.** In [7], Popov introduced another EG algorithm for the minimax problem (1.1) as follows:

$$\begin{cases} z_{n+1} = P_E(z_n - \tau g(\bar{z}_n)) & \text{(1.6a)} \\ \bar{z}_{n+1} = P_E(z_{n+1} - \tau g(\bar{z}_n)) & \text{(1.6b)} \end{cases}$$

for $n = 0, 1, \cdots$, where $z_0,\ \bar{z}_0 \in E$ are the initial guesses, and $\tau > 0$ is the sepsize.

Popov's EG (1.6) differs from Korpelevich's (1.5) in the way of defining the midway point $\bar{z}_n$; Popov's way of defining $\bar{z}_n$ looks more complicated than Korpelevich's. Popov proved the following convergence result. Note that the stepsize τ is shrunk to the range $(0, 1/3L)$.

Theorem 1.2. [7, Theorem 1] *Suppose $H_1 = \mathbb{R}^{d_1}$ and $H_2 = \mathbb{R}^{d_2}$ are Euclidean d_1- and d_2-spaces, respectively, and f is convex-concave and L-smooth. Suppose, in addition, the stepsize τ is chosen in such a range that $0 < \tau < \frac{1}{3L}$. Then the sequence $\{z_n\}$ generated by* (1.6) *converges to a saddle point of f.*

The main aim of this paper is to discuss possible enlargement of the stepsizes α and τ in Korpelevich's EG (1.5) and Popov's EG (1.6), respectively. We will show that the range of Korpelevich's stepsize $\alpha \in (0, 1/L)$ is sharp (see Example 3.1), and the range of Popov's stepsize $\tau \in (0, 1/3L)$ can however be relaxed at least to the range $\tau \in (0, (\sqrt{2}-1)/L)$ (see Theorem 4.5). We will also include the infinite-dimensional versions of the convergence results of Korpelevich's EG (1.5) and Popov's EG (1.6); in the latter, new stepsize is used.

2. Preliminaries

In this section we present some basic notion and tools which are required in the discussion and proof of the main results in the next section. We will use H to denote a Hilbert space. The symbols $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ always stand for the inner product and norm of a given Hilbert space under any circumstances (no confusions would arise). Let K be a nonempty closed convex subset of H and P_K be the metric projection from H onto K. Recall that

$$P_K u = \arg\min_{v \in K} \|u - v\|^2$$

for all $u \in H$. The following properties of P_K will be utilized throughout the subsequent sections.

Lemma 2.1. *P_K satisfies the properties:*

(a) $\langle u - P_K u, v - P_K u\rangle \le 0$ *for all* $u \in H$ *and* $v \in K$;
(b) $\langle u - v, P_K u - P_K v\rangle \ge \|P_K u - P_K v\|^2$, *in particular,* $\|P_K u - P_K v\| \le \|u - v\|$, *for all* $u, v \in H$;
(c) $\|v - P_K u\|^2 \le \|v - u\|^2 - \|u - P_K u\|^2$ *for all* $u \in H$ *and* $v \in K$.

We assume that the objective function f in the minimax problem (1.1) is convex-concave and L-smooth, a consequence of which is that the (saddle) gradient $g = \partial f$ defined in (1.2) is $\frac{1}{L}$-inverse strongly monotone ($\frac{1}{L}$-ISM), that is,

$$\langle g(z) - g(z'), z - z'\rangle \ge \frac{1}{L}\|g(z) - g(z')\|^2, \quad z, z' \in H. \tag{2.7}$$

Closely related to inverse strongly monotone mappings are averaged mappings. Recall that a mapping $T : H \to H$ is said to be μ-averaged (μ-AV) if

$$T = (1 - \mu)I + \mu V$$

where $\mu \in (0, 1)$, I is the identity on H, and $V : H \to H$ is a nonexpansive mapping (i.e., $\|V(z) - V(z')\| \le \|z - z'\|$ for all $z, z' \in H$). It is known that a projection P_K is 1-ISM and $\frac{1}{2}$-AV.

A useful connection between inverse strongly monotone mappings and averaged mappings is given below.

Lemma 2.2. (cf. [8]) *Given a mapping $T : H \to H$ and $\mu \in (0, 1)$. Then T is μ-AV if and only $I - T$ is $\frac{1}{2\mu}$-ISM. In particular, if f is L-smooth (i.e., g is $\frac{1}{L}$-ISM), then $P_K(I - \lambda g)$ is averaged for λ such that $0 < \lambda < \frac{2}{L}$.*

We will use the following notation:

- $\omega_w(z_n) := \{\xi \in H : z_{n_k} \to \xi$ weakly for some subsequence $\{z_{n_k}\}$ of $\{z_n\}\}$ is the set of weak accumulation points of a sequence $\{z_n\}$ in H;
- $Fix(T) := \{z \in H : Tz = z\}$ is the set of fixed points of a mapping $T : H \to H$.

For the purpose of proving weak (or strong) convergence of a sequence in H, we will use the following tools listed in the lemmas below.

Lemma 2.3. *(Demiclosedness principle for nonexpansive mappings.)* [2] *Let $T : K \to K$ be a nonexpasive mapping with a fixed point, where K is a nonempty closed convex subset of a Hilbert space H. Then $I - T$ is demiclosed (at 0), that is, for any sequence $\{z_n\}$ in K, the condition $\|z_n - Tz_n\| \to 0$ implies that $\omega_w(z_n) \subset Fix(T)$.*

Lemma 2.4. [6] *Suppose K is a nonempty subset of a Hilbert space H and $\{z_n\}$ is sequence in H such that the following conditions are satisfied:*

- $\lim_{n\to\infty} \|z_n - z\|$ *exists for each* $z \in K$, *and*
- $\omega_w(z_n) \subset K$.

Then the sequence $\{z_n\}$ weakly converges to a point in K.

Lemma 2.5. [9] *Suppose a real nonnegative sequence $\{a_n\}_{n=0}^\infty$ satisfies the condition:*

$$a_{n+1} \le a_n + \sigma_n, \quad n = 0, 1, \cdots,$$

where $\{\sigma_n\}$ is a nonnegative real sequence such that $\sum_{n=0}^\infty \sigma_n < \infty$. Then $\lim_{n\to\infty} a_n$ exists.

Lemma 2.6. [9] *Suppose a real nonnegative sequence $\{a_n\}_{n=0}^\infty$ satisfies the condition:*

$$a_{n+1} \le (1 - \gamma_n)a_n + \delta_n a_n + \sigma_n, \quad n = 0, 1, \cdots,$$

where $\{\gamma_n\}$, $\{\delta_n\}$, and $\{\sigma_n\}$ fulfil the properties:

(i) $\{\gamma_n\} \subset [0, 1]$ *and* $\sum_{n=0}^\infty \gamma_n = \infty$,

(ii) $\limsup_{n\to\infty} \delta_n \le 0$,

(iii) $\sigma_n \ge 0$ *for all* n, *and* $\sum_{n=0}^{\infty} \sigma_n < \infty$.

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

3. Stepsize Choice of Korpelvich's and Popov's Extragradient Algorithms

We begin with a brief discussion on the projection-gradient algorithm (PGA) for solving a convex minimization problem of the form:

$$\min_{x\in K} \varphi(z), \tag{3.8}$$

where K is a nonempty closed convex subset of a Hilbert space H, and φ is a convex, continuously Frechet differentiable function. Recall that PGA generates a sequence $\{z_n\}$ through the iteration process:

$$z_{n+1} = P_K(z_n - \lambda\nabla\varphi(z_n)) = P_K(I - \lambda\nabla\varphi)z_n, \quad n \ge 0, \tag{3.9}$$

where the initial guess $z_0 \in K$, and $\lambda > 0$ is a (constant) stepsize. The following convergence result is well known in the literature.

Theorem 3.3. *Suppose φ is convex and L-smooth (i.e., $\nabla\varphi$ is L-Lipschitz) and suppose* (3.8) *has a solution. If the stepsize λ is chosen in the range $0 < \lambda < \frac{2}{L}$, then the sequence $\{z_n\}$ generated by PGA* (3.9) *converges weakly to a solution of* (3.8).

Note that φ being convex and L-smooth implies that the mapping $P_C(I - \lambda\nabla\varphi)$ is averaged if $0 < \lambda < 2/L$ and nonexpansive if $\lambda = 2/L$ (cf. [8]).

Comparing the stepsizes α, τ and λ chosen in Korpelvich's EG (Theorem 1.1), Popov's EG (Theorem 1.2), and PGA (Theorem 3.3), respectively, it is natural and reasonable to wonder if the stepsizes α and τ in Korpelvich's and Popov's EGs can be made larger (close to $\frac{2}{L}$). We are concerned with this issue. More precisely, we will show that the stepsize $\alpha < \frac{1}{L}$ in Korpelvich's EG is sharp (see example 3.1), while the stepsize $\tau < \frac{1}{3L}$ in Popov's EG can be relaxed (see Theorem 4.5).

3.1. Stepsize of $1/L$ is sharp for Korpelvich's extragradient algorithm.

Example 3.1. If the stepsize α is taken to be $\frac{1}{L}$, then Korpelevich's algorithm (1.5) may not converge to an optimal solution, as shown by the following simple example. Let $H_1 = H_2 = \mathbb{R}$ and set $C = D = [0,1]$ Define the objective function f on $E := [0,1]^2$ by

$$f(x,y) = \frac{1}{2}x^2 - \frac{1}{2}y^2, \quad 0 \le x, y \le 1.$$

Clearly, f is convex-concave and differentiable. Since $\frac{\partial f}{\partial x} = x$ and $\frac{\partial f}{\partial y} = -y$, we get $g(z) = \partial f(z) = z$ for $z = (x,y) \in E$; hence g is 1-Lipschitz. It is easily seen that

$$\min_{x\in[0,1]} \max_{y\in[0,1]} f(x,y) = \max_{y\in[0,1]} \min_{x\in[0,1]} f(x,y) = f(0,0) = 0. \tag{3.10}$$

Korpelevich's algorithm (1.5) generates a sequence $\{z_n\}$ by

$$\bar{z}_n = P_E(z_n - \alpha g(z_n)), \quad z_{n+1} = P_E(z_n - \alpha g(\bar{z}_n)). \tag{3.11}$$

Now since $\alpha = \frac{1}{L} = 1$ and $g(z_n) = z_n$ for all n, it easily follows that $\bar{z}_n = (0,0)$ and $z_{n+1} = z_n$. So, if we take the initial guess $z_0 \in E$ with $z_0 \ne (0,0)$, then $z_n \equiv z_0$, not convergent to the saddle point (0,0).

However, if the stepsize $\alpha < 1$, then the sequence $\{z_n\}$ generated by Korpelevich's algorithm (3.11) satisfies

$$\bar{z}_n = (1-\alpha)z_0, \quad z_{n+1} = (1 - \alpha(1-\alpha))z_n, \quad n \ge 0.$$

It turns out that $z_n = (1-\alpha(1-\alpha))^n z_0 \to (0,0)$, the optimal solution of (3.10). This verifies Korpelevich's convergence theorem.

3.2. **Relaxation of Stepsize for Popov's extragradient algorithm.** Recall that Popov's EG produces a sequence $\{z_n\}$ by making use of the iteration procedure:

$$\begin{cases} z_{n+1} = P_E(z_n - \tau g(\bar{z}_n)) & (3.12a) \\ \bar{z}_{n+1} = P_E(z_{n+1} - \tau g(\bar{z}_n)) & (3.12b) \end{cases}$$

for $n = 0, 1, \cdots$, where $\tau > 0$ is the sepsize.

Popov [7] proved convergence of the algorithm (3.12) in the finite-dimensional setting when the stepsize τ is taken in the range $\tau \in (0, \frac{1}{3L})$. We now extend this range to $\tau \in (0, \frac{\sqrt{2}-1}{L})$.

Theorem 3.4. *Suppose $H_1 = \mathbb{R}^{d_1}$ and $H_2 = \mathbb{R}^{d_2}$ are Euclidean d_1- and d_2-spaces, respectively, and f is convex-concave and L-smooth. Then for stepsize τ such that $0 < \tau < \frac{\sqrt{2}-1}{L}$, the sequence $\{z_n\}$ generated by Popov's EG* (3.12) *converges to a saddle point of f.*

Proof. Take $\hat{z} \in E^*$ and use Lemma 2.1 to derive that, for $l \geq 0$,

$$\begin{aligned} \|z_{l+1} - \hat{z}\|^2 &= \|P_E(z_l - \tau g(\bar{z}_l)) - \hat{z}\|^2 \\ &\leq \|z_l - \hat{z} - \tau g(\bar{z}_l)\|^2 - \|z_l - \tau g(\bar{z}_l) - z_{l+1}\|^2 \\ &= \|z_l - \hat{z}\|^2 - 2\tau\langle z_l - \hat{z}, g(\bar{z}_l)\rangle + \tau^2\|g(\bar{z}_l)\|^2 \\ &\quad - \{\|z_l - z_{l+1}\|^2 - 2\tau\langle z_l - z_{l+1}, g(\bar{z}_l)\rangle + \tau^2\|g(\bar{z}_l)\|^2\} \qquad (3.13) \\ &= \|z_l - \hat{z}\|^2 - \|z_{l+1} - z_l\|^2 - 2\tau\langle z_{l+1} - \hat{z}, g(\bar{z}_l)\rangle. \qquad (3.14) \end{aligned}$$

Since $\hat{z}$ is a solution to VI (1.3) and g is $\frac{1}{L}$-ISM, it follows that

$$\langle g(z), z - \hat{z}\rangle - \frac{1}{L}\|g(z) - g(\hat{z})\|^2 \geq \langle g(\hat{z}), z - \hat{z}\rangle \geq 0, \quad z \in E.$$

In particular, we have

$$\langle g(\bar{z}_l), \bar{z}_l - \hat{z}\rangle - \frac{1}{L}\|g(\bar{z}_l) - g(\hat{z})\|^2 \geq 0.$$

Adding this to the right side of (3.13), we get

$$\begin{aligned} \|z_{l+1} - \hat{z}\|^2 &\leq \|z_l - \hat{z}\|^2 - \|z_{l+1} - z_l\|^2 - 2\tau\langle z_{l+1} - \bar{z}_l, g(\bar{z}_l)\rangle - \frac{2\tau}{L}\|g(\bar{z}_l) - g(\hat{z})\|^2 \\ &= \|z_l - \hat{z}\|^2 - \|z_{l+1} - \bar{z}_l\|^2 - \|\bar{z}_l - z_l\|^2 - 2\langle z_{l+1} - \bar{z}_l, \bar{z}_l - z_l\rangle \\ &\quad - 2\tau\langle z_{l+1} - \bar{z}_l, g(\bar{z}_l)\rangle - \frac{2\tau}{L}\|g(\bar{z}_l) - g(\hat{z})\|^2 \\ &= \|z_l - \hat{z}\|^2 - \|z_{l+1} - \bar{z}_l\|^2 - \|\bar{z}_l - z_l\|^2 \\ &\quad + 2\langle z_{l+1} - \bar{z}_l, z_l - \tau g(\bar{z}_l) - \bar{z}_l\rangle - \frac{2\tau}{L}\|g(\bar{z}_l) - g(\hat{z})\|^2 \\ &= \|z_l - \hat{z}\|^2 - \|z_{l+1} - \bar{z}_l\|^2 - \|\bar{z}_l - z_l\|^2 - \frac{2\tau}{L}\|g(\bar{z}_l) - g(\hat{z})\|^2 \\ &\quad + 2\langle z_{l+1} - \bar{z}_l, z_l - \tau g(\bar{z}_{l-1}) - \bar{z}_l\rangle + 2\tau\langle z_{l+1} - \bar{z}_l, g(\bar{z}_{l-1}) - g(\bar{z}_l)\rangle. \qquad (3.15) \end{aligned}$$

However, since $\bar{z}_l = P_G(z_l - \tau g(\bar{z}_{l-1}))$, by Lemma 2.1 we have

$$\langle z_{l+1} - \bar{z}_l, z_l - \tau g(\bar{z}_{l-1}) - \bar{z}_l\rangle \leq 0.$$

On the other hand, since g is L-Lipschitz continuous,

$$|\langle z_{l+1} - \bar{z}_l, g(\bar{z}_{l-1}) - g(\bar{z}_l)\rangle| \leq L\|z_{l+1} - \bar{z}_l\| \cdot \|\bar{z}_{l-1} - \bar{z}_l\|.$$

It turns out from (3.15) that

$$\begin{aligned}\|z_{l+1}-\hat z\|^2 &\le \|z_l-\hat z\|^2-\|z_{l+1}-\bar z_l\|^2-\|\bar z_l-z_l\|^2\\ &\quad+2\tau L\|\bar z_{l-1}-\bar z_l\|\|z_{l+1}-\bar z_l\|-\frac{2\tau}{L}\|g(\bar z_l)-g(\hat z)\|^2\\ &\le \|z_l-\hat z\|^2-\|z_{l+1}-\bar z_l\|^2-\|\bar z_l-z_l\|^2\\ &\quad+2\tau L(\|\bar z_{l-1}-z_l\|+\|z_l-\bar z_l\|)\|z_{l+1}-\bar z_l\|-\frac{2\tau}{L}\|g(\bar z_l)-g(\hat z)\|^2. \end{aligned}\tag{3.16}$$

An ingredient of this proof lies in the observation below: we have, for $\theta>0$ and $\beta>0$,

$$2\|\bar z_{l-1}-z_l\|\cdot\|z_{l+1}-\bar z_l\|\le\theta\|\bar z_{l-1}-z_l\|^2+\frac{1}{\theta}\|z_{l+1}-\bar z_l\|^2,$$

$$2\|z_l-\bar z_l\|\cdot\|z_{l+1}-\bar z_l\|\le\frac{1}{\beta}\|z_l-\bar z_l\|^2+\beta\|z_{l+1}-\bar z_l\|^2.$$

Substituting them into the right side of (3.16) yields

$$\begin{aligned}\|z_{l+1}-\hat z\|^2\le \|z_l-\hat z\|^2&\\ &-(1-\tau L(\theta+\frac{1}{\theta}+\beta))\|z_{l+1}-\bar z_l\|^2\\ &-(1-\frac{\tau L}{\beta})\|z_l-\bar z_l\|^2\\ &+\frac{\tau L}{\theta}(\|z_l-\bar z_{l-1}\|^2-\|z_{l+1}-\bar z_l\|^2)\\ &-\frac{2\tau}{L}\|g(\bar z_l)-g(\hat z)\|^2.\end{aligned}\tag{3.17}$$

A key point is to choose θ and β appropriately to ensure that the constants c_1 and c_2 set in (3.18) are positive, which is made success, due to our choice of the stepsize τ. As a matter of fact, since $0<\tau<\frac{\sqrt{2}-1}{L}<\frac{1}{2L}$, it is possible to take β such that $\tau L<\beta<\frac{1}{\tau L}-2$. We then take $\theta>0$ such that $\theta_-<\theta<\theta_+$, where

$$\theta_\pm=\frac{\frac{1}{\tau L}-\beta\pm\sqrt{(\frac{1}{\tau L}-\beta)^2-4}}{2}>0.$$

With such choices of β and θ we surely have

$$c_1:=1-\tau L(\theta+\frac{1}{\theta}+\beta)>0,\quad c_2:=1-\frac{\tau L}{\beta}>0.\tag{3.18}$$

Setting $c_3:=\frac{\tau L}{\theta}$ and $c_4:=\frac{2\tau}{L}$, we can rewrite (3.17) as

$$\begin{aligned}\|z_{l+1}-\hat z\|^2&\le\|z_l-\hat z\|^2-c_1\|z_{l+1}-\bar z_l\|^2-c_2\|z_l-\bar z_l\|^2\\ &\quad+c_3(\|z_l-\bar z_{l-1}\|^2-\|z_{l+1}-\bar z_l\|^2)-c_4\|g(\bar z_l)-g(\hat z)\|^2.\end{aligned}\tag{3.19}$$

Summing up (3.19) from $l=p$ and $l=q$ for each $q>p\ge 0$ yields

$$\begin{aligned}\|z_{q+1}-\hat z\|^2&\le\|z_p-\hat z\|^2-c_1\sum_{l=p}^{q}\|z_{l+1}-\bar z_l\|^2-c_2\sum_{l=p}^{q}\|z_l-\bar z_l\|^2\\ &\quad+c_3(\|z_p-\bar z_{p-1}\|^2-\|z_{q+1}-\bar z_q\|^2)-c_4\sum_{l=p}^{q}\|g(\bar z_l)-g(\hat z)\|^2.\end{aligned}\tag{3.20}$$

An immediate consequence of (3.20) is the following

$$\|z_{q+1}-\hat z\|^2\le\|z_p-\hat z\|^2+c_3\|z_p-\bar z_{p-1}\|^2\tag{3.21}$$

for all $q > p \geq 0$. In particular, (z_l) is bounded. Moreover, (3.20) also implies that

- $\sum_{l=0}^{\infty} \|z_{l+1} - \bar{z}_l\|^2 < \infty$,
- $\sum_{l=0}^{\infty} \|z_l - \bar{z}_l\|^2 < \infty$,
- $\sum_{l=0}^{\infty} \|g(\bar{z}_l) - g(\hat{z})\|^2 < \infty$.

Consequently,

$$\lim_{l\to\infty} \|z_{l+1} - \bar{z}_l\| = \lim_{l\to\infty} \|z_l - \bar{z}_l\| = 0 \tag{3.22}$$

and

$$\lim_{l\to\infty} \|g(\bar{z}_l) - g(\hat{z})\| = 0. \tag{3.23}$$

It turns out that $\lim_{l\to\infty} \|z_{l+1} - z_l\| = 0$. On the other hand, a special case of (3.21) is

$$\|z_{l+1} - \hat{z}\|^2 \leq \|z_l - \hat{z}\|^2 + \mu_l, \quad l \geq 0, \tag{3.24}$$

where $\mu_l = \tau L\|z_l - \bar{z}_{l-1}\|^2$. Since $\sum_{l=1}^{\infty} \mu_l < \infty$, we can apply Lemma 2.5 to (3.24) to get that

$$\lim_{l\to\infty} \|z_l - \hat{z}\| \quad \text{exists for each } \hat{z} \in E^*. \tag{3.25}$$

By virtue of boundedness, (z_l) has a subsequence $(z_{l'})$ convergent to some point z'. We now claim that $z' \in E^*$ and then by (3.25), the full sequence (z_l) must be convergent to z'. Note from (3.22) that $\bar{z}_{l'} \to z'$ and $z_{l'+1} \to z'$. By (3.12a), we have $z_{l'+1} = P_G(z_{l'} - \tau g(\bar{z}_{l'}))$. Taking the limit as $l' \to \infty$ in yields $z' = P_E(z' - \tau g(z'))$. This is equivalent to the VI

$$\langle g(z'), z - z' \rangle \geq 0$$

for all $z \in E$. Hence $z' \in E^*$. This completes the proof. $\square$

Remark 3.1. Our range of stepsize $\tau \in (0, (\sqrt{2}-1)/L)$ relaxed Popov's range $\tau \in (0, 1/3L)$. It remains an open question whether the stepsize can be enlarged to the range $\tau \in (0, 2/L)$.

Also in our proof of Theorem 1.2, we made use of the fact that g is $\frac{1}{L}$-ISM by including the terms $\|g(\bar{z}_l)-g(\hat{z})\|^2$ in (3.16) and $\sum_{l=p}^{q} \|g(\bar{z}_l)-g(\hat{z})\|^2$ in (3.20), respectively. However, we are unaware if these terms may help to enlarge the choice range of the stepsize τ. However, from (3.23), we get at least that $g(\bar{z}_l) \to g(\hat{z})$, a consequence of which is that $g(\tilde{z}) = g(\hat{z})$ for all $\tilde{z}$, $\hat{z} \in E^*$. This is no surprising because in the unconstrained case (i.e., $E = H$), $g(\hat{z}) = 0$ for all $\hat{z} \in E^*$.

4. Infinite-dimensional setting

We begin by establishing the infinite-dimensional version of convergence of Popov's EG algorithm (3.12).

Theorem 4.5. *Suppose H_1 and H_2 are infinite-dimensional Hilbert spaces, and f is convex-concave and L-smooth. Then for stepsize τ such that $0 < \tau < \frac{\sqrt{2}-1}{L}$, the sequence $\{z_n\}$ generated by Popov's EG* (3.12) *converges weakly to a saddle point of f.*

Proof. We will use Lemma 2.4 to prove the weak convergence of $\{z_n\}$. Because of (3.25) which remains valid in the infinite-dimensional scenario, it only remains to verify that $\omega_w(z_n) \subset E^*$. This is not straightforward since the gradient operator g is, in general, not weakly continuous. To overcome the difficulty, we manipulate the technique of nonexpansive mappings. Set $T := P_G(I - \tau g)$; note that $Fix(T) = E^*$. Since $0 < \tau < \frac{\sqrt{2}-1}{L} < \frac{2}{L}$,

T is a nonexpansive mapping; hence $I - T$ is demiclosed. We claim that $\|z_n - Tz_n\| \to 0$. Indeed, we have

$$\begin{aligned}\|z_n - Tz_n\| &\le \|z_n - z_{n+1}\| + \|z_{n+1} - Tz_n\| \\ &= \|z_n - z_{n+1}\| + \|P_G(z_n - \tau g(\bar{z}_n)) - P_G(z_n - \tau g(z_n))\| \\ &\le \|z_n - z_{n+1}\| + \tau\|g(\bar{z}_n) - g(z_n)\| \\ &\le \|z_n - z_{n+1}\| + \tau L\|\bar{z}_n - z_n\| \\ &\le \|z_{n+1} - \bar{z}_n\| + (1+\tau L)\|\bar{z}_n - z_n\| \to 0 \quad \text{by (3.22)}.\end{aligned}$$

Consequently, we can use the demiclosedness of $I - T$ (Lemma 2.3) to obtain $\omega_w(z_n) \subset Fix(T) = E^*$. This together with (3.25) ensures that (z_l) converges weakly to a point in E^*. □

We end this section by proving the infinite-dimensional version of Korpelvich's EG algorithm (1.5).

Theorem 4.6. *Suppose H_1 and H_2 are infinite-dimensional Hilbert spaces, and f is convex-concave and L-smooth. Suppose the stepsize τ is chosen in the range $0 < \alpha < \frac{1}{L}$. Then the sequence $\{z_n\}$ generated by Korpelvich's EG algorithm* (1.5) *converges weakly to a saddle point of f.*

Proof. Observe that Korpelvich's finite-dimensional argument in the proof of [4, Theorem 1] remains valid for the infinite-dimensional setting; hence [4, Eq. (14), p. 39] remains true, that is,

$$\|z_{n+1} - z^*\|^2 \le \|z_n - z^*\|^2 - (1 - \alpha^2 L^2)\|z_n - \bar{z}_n\|^2. \tag{4.26}$$

Since $0 < \alpha L < 1$, it follows from (4.26) that

- $\|z_{n+1} - z^*\|^2 \le \|z_n - z^*\|^2$; hence $\{z_n\}$ is bounded and

$$\lim_{n\to\infty} \|z_n - z^*\| \quad \text{exists for each } z^* \in E^*. \tag{4.27}$$

- $\lim_{n\to\infty} \|z_n - \bar{z}_n\| = 0$.

Similar to the proof of Theorem 4.5, it remains to verify that $\omega_w(z_n) \subset E^*$. Define the mapping $T := P_G(I - \alpha g)$. Again we have $Fix(T) = E^*$; moreover, T is averaged since $\alpha < \frac{1}{L}$. By (1.5a), $\bar{z}_n$ can be rewritten as $\bar{z}_n = Tz_n$. Hence, we arrive at

$$\|z_n - Tz_n\| = \|z_n - \bar{z}_n\| \to 0.$$

Consequently, Lemma 2.3 (the demiclosedness principle for nonexpansive mappings) is applicable to obtain $\omega_w(z_n) \subset Fix(T) = E^*$. This ends the proof. □

5. Conclusion

In this paper we have discussed the problem of stepsize choice in two classic extragradient algorithms - Korpelevich's and Popov's methods [4, 7], where the objective function is convex-concave and L-Lipschitz. We have discovered that the range of the stepsize $\alpha \in (0, 1/L)$ is best possible for Korpelevich's EG algorithm, and relaxed the range of the stepsize $\tau \in (0, 1/3L)$ in Popov's EG algorithm to the slightly larger range $\tau \in (0, (\sqrt{2}-1)/L)$. It is an open question what is the best possible range for the stepsize τ for Popov's EG algorithm. In addition, we have also proved weak convergence of both Korpelevich's and Popov's EG algorithms in an infinite-dimensional Hilbert space.

Acknowledgments. The authors were indebted to the anonymous referees for their suggestions and ccomments which improved the presentation of this manuscript.

HKX was supported in part by National Natural Science Foundation of China (grant no. U1811461) and by Australian Research Council (grant no. DP 200100124).

REFERENCES

[1] Diakonikolas, J. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. arXiv:2002.08872v3. *Proceedings of Thirty Third Conference on Learning Theory* PMLR 125:1428-1451, 2020.

[2] Goebel, K.; Kirk, W. A. Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics 28, *Cambridge University Press, Cambridge*, 1990.

[3] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville; A.; Bengio, Y. Generative adversarial nets. In Advances in neural information processing systems, pp. 2672–2680, 2014.

[4] Korpelevich, G. M. The extragradient method for finding saddle points and other problems. *Ekonomika i matematicheskie metody* **12** (1976), no. 4, 747–756.

[5] Lin, T.; Jin, C.; Jordan, M. I. *On gradient descent ascent for nonconvex-concave minimax problems*. arXiv.org/abs/1906.00331v1 (2019).

[6] Opial, Z. Weak convergence of successive approximations for nonexpansive mappins. *Bull. Amer. Math. Soc.* **37** (1967), no. 4, 591–597.

[7] Popov, L. D. A modification of the Arrow-Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR* **28** (1980), no. 5, 777–784.

[8] Xu, H. K. Averaged mappings and the gradient-projection algorithm. *J. Optim. Theory Appl.* 150 (2011), 360-378.

[9] Xu, H. K. Iterative algorithms for nonlinear operators. *J. London Math. Soc.* 66 (2002), 240-256.

[10] Yoon, T.; Ryu, E. K. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm, *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139, 2012, 12098–12109.

DEPARTMENT OF MATHEMATICS
HANGZHOU DIANZI UNIVERSITY
SCHOOL OF SCIENCE
HANGZHOU 310018, CHINA
Email address: 1912454791@qq.com, xuhk@hdu.edu.cn