Halpern's Iteration Method for Convex-Concave Minimax Problems

Hong-Kun Xu

School of Science
Hangzhou Dianzi University
Hangzhou 310018
China
E-mail: xuhk@hdu.edu.cn

Talk at
Workshop on Optimization and Operator Theory
Dedicated to Professor Lev Bregman on the Occasion of His 80th Birthday
Faculty of Mathematics, the Technion - Israel Institute of Technology
November 15-17, 2021

Abstract

Algorithmic approaches to minimax problems have recently been paid much attention, due to their important applications in machine learning, in particular, in generative adversarial nets (GANs). In this talk we will first review some recent progresses on the convergence rate of Halpern's iteration method, and then discuss several applications of Halpern's method in optimization problems, including variational inequalities, monotone inclusions, Douglas-Rachford splitting method. In particular, we will discuss how Halpern's method can be used to prove the strong convergence of recently introduced extra anchored gradient (EAG) algorithm for smooth convex-concave minimax problems in an even infinite-dimensional Hilbert space.

Outline

Outline

Minimax Problems

Minimax problems go back to von Neumann in the late 1920's and 30's:

J. von Neumann, Zur Theorie der Gesellschaftsspiele, Math. Ann. 100 (1928), 295-320.

—, Uber ein bkonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes, Ergebn. Math. Kolloqu. Wien 8 (1935-36), 73-83.

von Neumann's Minimax Theorem: If $f(x, y)$ is quasi-convex in x and quasi-concave in y, then

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y)$$

where X and Y are nonempty convex compact subsets of topological linear spaces, and $f : X \times Y \to \mathbb{R}$ is continuous in each variable.

Convex-Concave Minimax Problem

In this talk, we will focus on iterative methods for a (constrained) convex-concave minimax problem in a Hilbert space setting:

$$\min_{x\in Q}\max_{y\in S} f(x,y), \tag{1.1}$$

where Q and S are nonempty closed convex subsets of Hilbert spaces H_1 and H_2, respectively, and $f(x,y)$ is convex in x (for each fixed $y \in S$) and concave in y (for each fixed $x \in Q$).

Saddle Point

A solution of the minimax problem is interpreted as a saddle point defined as follows.

Definition

A pair of points $(u^*, v^*) \in Q \times S$ is said to be a saddle point of f (or a solution of the minimax problem (1.1)) if and only if the inequalities below are satisfied:

$$f(u^*, v) \leq f(u^*, v^*) \leq f(u, v^*), \quad u \in Q, \quad v \in S. \tag{1.2}$$

Namely, $u^* \in Q$ is a minimizer in Q of the function $f(\cdot, v^*)$, and $v^* \in S$ is a maximizer in S of the function $f(u^*, \cdot)$.

Set $G = Q \times S$ and $H = H_1 \times H_2$. Let $G^* := Q^* \times S^*$ denote the set of saddle points of the minimax problem (1.1).

Existence of Saddle Points

It is known that f has a saddle point if and only if

$$\min_{u\in Q}\max_{v\in S} f(u,v) = \max_{v\in S}\min_{u\in Q} f(u,v).$$

Theorem

Suppose

(i) *for each fixed $v \in S$, the function $f(\cdot, v)$ is convex and lower semicontinuous (l.s.c.),*

(ii) *for each fixed $u \in Q$, the function $f(u, \cdot)$ is concave and upper semicontinuous (u.s.c.).*

Assume, in addition, either that Q and S are bounded, or that

(iii) *there exists $\bar{v} \in S$ such that the function $f(\cdot, \bar{v})$ is coercive, i.e., $f(u, \bar{v}) \to \infty$ as $\|u\| \to \infty$ ($u \in Q$);*

(iv) *there exists $\bar{u} \in Q$ such that the function $-f(\bar{u}, \cdot)$ is coercive, i.e., $f(\bar{u}, v) \to -\infty$ as $\|v\| \to \infty$ ($v \in S$).*

Then f has at least one saddle point.

Applications

Minimax problems find applications recently in machine learning, including

- Generative adversarial nets (GANs) (I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NeurIPS, pages 2672 – 2680, 2014.)
- Statistical learning (R. Giordano, T. Broderick, and M. I. Jordan, Covariances, robustness, and variational bayes. ArXiv Preprint: 1709.02536, 2017.)
- Certification of robustness in deep learning (A. Sinha, H. Namkoong, and J. Duchi, Certifiable distributional robustness with principled adversarial training. In ICLR, 2018)
- Distributed computing (G. Mateos, J. A. Bazerque, and G. B. Giannakis, Distributed sparse linear regression. IEEE Transactions on Signal Processing, 58(10):5262-5276, 2010)

Tianyi Lin, Chi Jin, and Michael I. Jordan, On Gradient Descent Ascent for Nonconvex-Concave Minimax Problems. arXiv.org/abs/1906.00331v1 (2019).

Outline

Characterization of Saddle Points via VIs

Theorem

Assume f is convex-concave and differentiable. Let

$$\partial f(u,v) = \begin{bmatrix} \nabla_u f(u,v) \\ -\nabla_v f(u,v) \end{bmatrix}. \tag{1.3}$$

[Note: ∂f is monotone.] Then $z^* := (u^*, v^*) \in Q \times S$ is a saddle point of f if and only if z^* solves the variational inequality (VI)

$$\langle \partial f(z^*), z - z^* \rangle \geq 0, \quad z = (u,v) \in G. \tag{1.4}$$

Equivalently,

$$\langle \nabla_u f(u^*, v^*), u - u^* \rangle \geq 0 \quad \text{and} \quad \langle \nabla_v f(u^*, v^*), v - v^* \rangle \leq 0, \quad (u,v) \in Q \times S. \tag{1.5}$$

For the unconstrained case (i.e., $Q = H_1$ and $S = H_2$), VI (1.4) *is reduced to the equation $\partial f(z^*) = 0$.*

Outline

Korpelevich's Extragradient (EG) Algorithm

Define $g : G \to H$ by

$$g(z) = \partial f(z) = \begin{bmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{bmatrix}. \tag{2.1}$$

Korpelevich (1976) introduced the following extragradient (EG) algorithm:

$$\begin{cases} \bar{z}_n = P_G(z_n - \alpha g(z_n)) & \text{(2.2a)} \\ z_{n+1} = P_G(z_n - \alpha g(\bar{z}_n)) & \text{(2.2b)} \end{cases}$$

where $z_0 \in G$ and $\alpha > 0$. In components, the algorithm (2.2) can be rewritten as

$$\begin{cases} \bar{x}_n = P_Q(x_n - \alpha \nabla_x f(x_n, y_n)) & \text{(2.3a)} \\ \bar{y}_n = P_S(y_n + \alpha \nabla_y f(x_n, y_n)) & \text{(2.3b)} \\ x_{n+1} = P_Q(x_n - \alpha \nabla_x f(\bar{x}_n, \bar{y}_n)) & \text{(2.3c)} \\ y_{n+1} = P_S(y_n + \alpha \nabla_y f(\bar{x}_n, \bar{y}_n)). & \text{(2.3d)} \end{cases}$$

Convergence of EG (2.2)

Theorem

(Korpelevich, 1976) Suppose $\dim H_1 < \infty$ *and* $\dim H_2 < \infty$*. Suppose also that* f *is convex-concave and* L*-smooth. Then for all* α*,* $0 < \alpha < \frac{1}{L}$*. Then the sequence* $\{z_n\}$ *generated by* (2.2) *converges to a saddle point of* (1.1)*.*

Recall that f is L-smooth if ∂f is L-Lipschitz:

$$\|\partial f(z) - \partial f(z')\| \leq L\|z - z'\|, \quad z, z' \in G.$$

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

Outline

Popov's Extragradient Algorithm

In 1980, Popov introduced the following extragradient method for solving the minimax problem (1.1):

$$\begin{cases} z_{n+1} = P_G(z_n - \tau g(\bar{z}_n)) & \text{(2.4a)} \\ \bar{z}_{n+1} = P_G(z_{n+1} - \tau g(\bar{z}_n)) & \text{(2.4b)} \end{cases}$$

for $n = 0, 1, \cdots$, where $z_0,\ \bar{z}_0 \in G$, and P_G is the projection onto G from H.

L. D. Popov, 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.

Convergence of Popov's Extragradient Method

Theorem

(Popov, 1980) Suppose $\dim H_1 < \infty$ *and* $\dim H_2 < \infty$. *Suppose also* f *is convex-concave and* L*-smooth. Then for all* τ, $0 < \tau < \frac{1}{3L}$, *the sequence* $\{z_n\}$ *generated by* (2.4) *converges to a saddle point of* (1.1).

Halpern's Iteration Method

Recently, Halpern's iteration method was applied to the minimax problem (1.1) and also to VIs, in general. This talk is motivated by the articles below:

- J. Diakonikolas, Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. arXiv:2002.08872v3. (Apr 2020)
- Jelena Diakonikolas, Constantinos Daskalakis, and Michael I. Jordan, Efficient Methods for Structured Nonconvex-Nonconcave Min-Max Optimization. arXiv:2011.00364v2 [math.OC]
- E. K. Ryu, K. Yuan, and W. Yin, ODE analysis of stochastic gradient methods with optimism and anchoring for minimax problems. arXiv:1905.10899v3 [cs.LG] 12 Oct 2020.
- T. Yoon and E. K. Ryu, 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, 2021.

Outline

Nonexpansive Mappings

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. A mapping $T : H \to H$ is said to be nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\| \quad \text{for all } x, y \in H.$$

A point $x \in H$ is a fixed point of T if $Tx = x$. $Fix(T) := \{x \in H : Tx = x\}$ denotes the set of fixed points of T (possibly empty).
Moreover, we say that T is α-averaged (α-AV, for short) if $\alpha \in (0, 1)$ and

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

with $V : X \to X$ nonexpansive. Clearly, averaged mappings are nonexpansive.

Notes:

- $Fix(V) = Fix(T)$;
- T has a fixed point if and only of there exists x such that the trajectory $\{T^n x\}$ is bounded.

Halpern's Iteration Method for Nonexpansive Mappings

Let C be a nonempty closed convex subset of a Banach space X and $T : C \to C$ be a nonexpansive mapping. Assume $Fix(T) \neq \emptyset$.
Halpern's iteration method generates a sequence (x_n) by the iteration process

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) T x_n, \quad n = 0, 1, 2, \cdots , \tag{3.1}$$

where (α_n) is a sequence in $[0, 1]$, $u \in C$ referred to as anchor, and $x_0 \in C$ an initial guess taken arbitrarily.

The algorithm (3.1) was first introduced by B. Halpern[1] in a Hilbert space H and for the special case where C is the closed unit ball of H and the anchor $u = 0$.

---

[1] B. Halpern, Fixed points of nonexpanding maps, Bull. Amer. Math. Soc. 73 (1967), 591-597.

Necessary Conditions for Convergence of Halpern's Method

Halpern noticed two necessary conditions for convergence of Halpern's method:

(C1) $\alpha_n \to 0$;

(C2) $\sum_{n=1}^{\infty} \alpha_n = \infty$.

These two conditions are not sufficient to guarantee convergence of Halpern's method unless T is AV.

Convergence of Halpern's Iteration Method

Table: Convergence of Halpern's Iteration Method under (C1)-(C2)+?

Author	Year	Journal	Add. Condition	Setting
B. Halpern	1967	Bull. Amer. Math. Soc.	(C3)	Hilbert
P.L. Lions	1977	C.R. Acad. Sci. Paris	(C4)	Hilbert
R. Wittmann	1992	Arch. Math.	(C5)	Hilbert
S. Reich	1994	Panamerican. Math. J.	(C6)	Hilbert
X.	2002	J. London Math. Soc.	(C7)	Banach

(C3) (α_n) is acceptable: there exists $(n(i))$ such that (i) $n(i+1) \geq n(i)$, (ii) $\lim_{i\to\infty} \frac{\alpha_{i+n(i)}}{\alpha_i} = 1$, (iii) $\lim_{i\to\infty} n(i)\alpha_i = \infty$;

(C4) $\lim_{n\to\infty} \frac{|\alpha_{n+1}-\alpha_n|}{\alpha_n^2} = 0$ (e.g., $\alpha_n = \frac{1}{(n+1)^\alpha}$, $0 < \alpha < 1$);

(C5) $\sum_{n=1}^{\infty} |\alpha_{n+1} - \alpha_n| < \infty$ (e.g., $\alpha_n = \frac{1}{(n+1)^\alpha}$, $0 < \alpha \leq 1$);

(C6) (α_n) is decreasing;

(C7) $\lim_{n\to\infty} \frac{|\alpha_{n+1}-\alpha_n|}{\alpha_n} = 0$, i.e., $\frac{\alpha_{n+1}}{\alpha_n} \to 1$. (e.g., $\alpha_n = \frac{1}{(n+1)^\alpha}$, $0 < \alpha \leq 1$)

A Survey Article on Halpern, up to 2010

G. Lopez, V. Martin-Marquez, and H.K. Xu, Halpern's iteration for nonexpansive mappings, Contemporary Mathematics, 513 (2010), 211-231.

Outline

Lieder's $O(1/n)$ Rate of Convergence

It is remarkable that Lieder proved $O(1/n)$ rate of Halpern's method.

Theorem

(Felix Lieder [10]) Let H be a Hilbert space, let $T : H \to H$ be a nonexpansive mapping such that $Fix(T) \neq \emptyset$, and let (x_n) be generated by Halpern's method:

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) T x_n, \tag{3.2}$$

where $\alpha_n = \frac{1}{n+2}$ and $u = x_0$. Then

$$\|x_n - Tx_n\| \leq \frac{2\|x_0 - x^*\|}{n+1}, \quad n \geq 0, \; x^* \in Fix(T). \tag{3.3}$$

This bound is tight.

F. Lieder, On the convergence rate of the Halpern-iteration, Optiminzation Letters 15 (2021), 405-418.

Variational Inequalities (VIs)

Consider a nonempty closed convex subset C of a Hilbert space H and a monotone operator $F : C \to H$. The variational inequality (VI) is to seek a point $u^* \in C$ with the property

$$\langle F(u^*), u - u^* \rangle \geq 0, \quad u \in C. \tag{4.1}$$

VI (4.1) is denoted as $VI(F; C)$ and its solution set as $S(F; C)$, respectively.

It is known that $VI(F; C)$ is equivalent to the fixed point problem, for all $\gamma > 0$,

$$u^* = P_C(I - \gamma F)u^* \tag{4.2}$$

Hence, fixed point methods can be applied to solve VIs.

Inverse Strongly Monotone Operators (ISM)

A (single-valued) mapping on a Hilbert space H is said to be inverse strongly monotone (ISM) or cocoercive if, for some constant $\gamma > 0$,

$$\langle F(x) - F(y), x - y \rangle \geq \gamma \|F(x) - F(y)\|^2, \quad x, y \in H. \tag{4.3}$$

In this case, F is also said to be γ-ISM.

Projections P_C and proximal mappings $\mathrm{prox}_{\lambda g}$ are 1-ISM (note: P_C and $\mathrm{prox}_{\lambda g}$ are also $\frac{1}{2}$-AV). Also, $T : H \to H$ is nonexpansive if and only if $I - T$ is $\frac{1}{2}$-ISM.

Relations between AV and ISM

Proposition

Let $T : H \to H$ *be a mapping. Then* T *is* α*-AV for some* $\alpha \in (0,1)$ *if and only if* $I - T$ *is* $\frac{1}{2\alpha}$*-ISM (note that* $\frac{1}{2\alpha} > \frac{1}{2}$*).*

Lipschitzian Gradient $\nabla\varphi$ is ISM

Proposition

Let $\varphi : H \to H$ be a continuously Frechet differential, convex function. Suppose $\nabla\varphi$ is L-Lipschitz:

$$\|\nabla\varphi(x) - \nabla\varphi\| \leq L\|x - y\|, \quad x, y \in H.$$

Then $\nabla\varphi$ is $\frac{1}{L}$-ISM:

$$\langle\nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \geq \frac{1}{L}\|\nabla\varphi(x) - \nabla\varphi(y)\|^2, \quad x, y \in H.$$

Outline

Halpern Method for Constrained Optimization

Consider the convex minimization problem

$$\min_{x \in C} \varphi(x)$$

with solution set $S \neq \emptyset$. Note: $S = Fix(P_C(I - \lambda\nabla\varphi))$ for any $\lambda > 0$. Suppose $\nabla\varphi$ is L-Lipschitz. Set $T_\lambda = P_C(I - \lambda\nabla\varphi)$. Then T_λ is $\frac{2+\lambda L}{4}$-AV for $0 < \lambda < \frac{2}{L}$. It turns out that Halpern's iteration ($0 < \lambda < \frac{2}{L}$):

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) P_C(I - \lambda\nabla\varphi)x_n, \quad n = 0, 1, \cdots \tag{4.4}$$

strongly converges to the solution $u^* := P_S(u)$, where (α_n) satisfies (C1) and (C2). In particular, take $\alpha_n = \frac{1}{n+2}$ and $\lambda = \frac{2}{L}$, we get that

$$x_{n+1} = \frac{x_0}{n+2} + \frac{n+1}{n+2} P_C(I - \frac{2}{L}\nabla\varphi)x_n$$

converges in norm to $P_S u$ with the rate of convergence (for any $x^* \in S$):

$$\|x_n - P_C(I - \frac{2}{L}\nabla\varphi)x_n\| \leq \frac{2\|x_0 - x^*\|}{n+1}.$$

Halpern Method for Composite Optimization

Consider the composite convex minimization problem

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

with solution set $S \neq \emptyset$. Note: $S = Fix(\text{prox}_{\lambda g}(I - \lambda \nabla f))$ for any $\lambda > 0$. Suppose ∇f is L-Lipschitz and set $T_\lambda = \text{prox}_{\lambda g}(I - \lambda \nabla f)$. Then T_λ is $\frac{2+\lambda L}{4}$-AV for $0 < \lambda < \frac{2}{L}$. It turns out that Halpern's iteration ($0 < \lambda < \frac{2}{L}$):

$$x_{n+1} = \alpha_n u + (1 - \alpha_n)\text{prox}_{\lambda g}(I - \lambda \nabla f))x_n, \quad n = 0, 1, \cdots \tag{4.5}$$

strongly converges to the solution $u^* := P_S(u)$, where (α_n) satisfies the conditions (C1) and (C2):
(C1) $\alpha_n \to 0$;
(C2) $\sum_{n=0}^{\infty} \alpha_n = \infty$.

Outline

Consider IV:

$$\langle F(u^*), u - u^* \rangle \geq 0, \quad u \in C, \tag{4.6}$$

where C is a nonempty closed convex subset C of a Hilbert space H and $F : C \to H$ is $\frac{1}{L}$-ISM. Assume its solution set as $S \neq \emptyset$. Note that $S = Fix(P_C(I - \gamma F))$ for any $\gamma > 0$. Now define a mapping G_η by

$$G_\eta := \eta \left(I - P_C \left(I - \frac{1}{\eta} F \right) \right).$$

Then it is not hard to find that G_η is $\frac{2}{2\eta + L}$-ISM for $\eta > \frac{L}{2}$. Note that $G^{-1}0 = F^{-1}0$. [Nesterov called G_η gradient mapping when $F = \nabla \varphi$.]

Halpern Method Applied to $I - (2/L)G_{L/2}$

Let $\eta > \frac{L}{2}$. Since G_η is $\frac{2}{2\eta+L}$-ISM, $I - \frac{4}{2\eta+L}G_\eta$ is nonexpansive. It turns out that $I - \frac{2}{L}G_{L/2}$ is nonexpansive. Applying Halpern to $I - \frac{2}{L}G_{L/2}$ yields

$$x_{n+1} = \alpha_n u + (1-\alpha_n)\left(x_n - \frac{2}{L}G_{L/2}x_n\right), \quad n = 0, 1, \cdots$$

strongly converges to the solution $u^* := P_S(u)$, where (α_n) satisfies the conditions (C1) and (C2) plus one of (C3)-(C6). In particular, if $\alpha_n = \frac{1}{n+2}$ and $u = x_0$,

$$\|G_{L/2}(x_n)\| \le \frac{L\|x_0 - x^*\|}{n+1} = O\left(\frac{1}{n}\right)$$

and $\|G_{L/2}(x_n)\| \le \varepsilon$ after at most $(\frac{2L\|x_0 - x^*\|}{\varepsilon} + 1)$ iterations. Our results slightly refine those of J. Diakonikolas [2].

J. Diakonikolas, Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. arXiv:2002.08872v3 (Apr 2020).

Outline

Zeros of the Sum of Two Maximal Operators

Consider the problem of finding a zero of the sum of two maximal monotone operators:

$$0 \in (A+B)x, \tag{4.7}$$

where A and B are two maximal (multivalued) monotone operators in a Hilbert space H. Assume the solution set of (4.7) is nonempty. Recall the resolvent $J_\lambda^A := (I+\lambda A)^{-1}$ for $\lambda > 0$.

Lemma

We have that $v \in Fix((2J_\lambda^A - I)(2J_\lambda^B - I))$ or $v \in Fix(J_\lambda^A(2J_\lambda^B - I) + (I - J_\lambda^B))$ if and only if $u := J_\lambda^B v$ is a solution of (4.7).

Remark

(i) $(2J_\lambda^A - I)(2J_\lambda^B - I)$ is nonexpansive (not AV, in general); (ii) $J_\lambda^A(2J_\lambda^B - I) + (I - J_\lambda^B)$ is $\frac{1}{2}$-AV (or firmly nonexpansive).

Douglas-Rachford (DR) Mappings

- The iterates $x_{n+1} = (2J_\lambda^A - I)(2J_\lambda^B - I)x_n$ fail to converge;
- The iterates $x_{n+1} = (J_\lambda^A(2J_\lambda^B - I) + (I - J_\lambda^B))x_n$ converge weakly to some point v such that $u = J_\lambda^B v$ is a solution of (4.7) [12].

The generalized DR mapping:

$$V_\beta := (1-\beta)I + \beta R_\lambda^A R_\lambda^B, \quad 0 < \beta \leq 1; \; R_\lambda^A = 2J_\lambda^A - I, R_\lambda^B = 2J_\lambda^B - I.$$

- $V_1 = (2J_\lambda^A - I)(2J_\lambda^B - I)$ is the Peaceman-Rachford (PR) mapping;
- $V_{1/2} = J_\lambda^A(2J_\lambda^B - I) + (I - J_\lambda^B)$ is the DR mapping.

So V_β unifies the two mappings.

P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM Journal on Numerical Analysis 16 (1979), no. 6, 964-979.

Douglas-Rachford-Halpern Method

Douglas-Rachford-Halpern is Halpern applied to the generalized DR mapping, which yields

$$x_{n+1} = \alpha_n u + (1-\alpha_n)((1-\beta)x_n + \beta R_\lambda^A R_\lambda^B x_n), \quad n = 0, 1, \cdots$$

strongly converges to $v := P_{Fix(R_\lambda^A R_\lambda^B)} u$ and the solution $u^* := J_\lambda^B(v)$ is a solution of (4.7), where (α_n) satisfies the conditions (C1) and (C2) plus one of (C3)-(C7) [if $0 < \beta < 1$, then (C1) and (C2) are sufficient]. The $O(1/n)$ rate of convergence can be obtained by taking $\alpha_n = \frac{1}{n+2}$ and $u = x_0$:

$$\|x_n - R_\lambda^A R_\lambda^B x_n\| \leq \frac{2\|x_0 - x^*\|}{n+1}.$$

Outline

Halpern for Minimax

Proposition

Assume the convex-concave objective function f *is* L*-smooth. Given an initial point* $z_0 = (x_0, y_0) \in H$ *and anchor* $w = (u, v) \in H$*. Define a sequence* $z_n = (x_n, y_n)$ *in* H *by the Halpern iteration method:*

$$z_{n+1} = \alpha_n w + (1 - \alpha_n)(z_n - \frac{2}{L}\partial f(z_n)), \quad n = 0, 1, \cdots ;$$

alternatively, in components,

$$x_{n+1} = \alpha_n u + (1 - \alpha_n)(x_n - \frac{2}{L}\nabla_x f(x_n, y_n)),$$
$$y_{n+1} = \alpha_n v + (1 - \alpha_n)(y_n + \frac{2}{L}\nabla_y f(x_n, y_n)).$$

Suppose (α_n) *satisfies the conditions (C1)-(C2) and one of (C3)-(C7). Then* (z_n) *converges in norm to* $z^* = P_S w$*. Moreover, if we take* $\alpha_n = \frac{1}{n+2}$ *and* $w = z_0$*, then, for any* $\hat{z} \in S$*,*

$$\|\partial f(z_n)\| \le \frac{L\|z_0 - \hat{z}\|}{n+1}.$$

Extra Anchored Gradient (EAG)

Consider the (unconstrained) convex-concave minimax problem:

$$\min_{x \in H_1} \max_{y \in H_2} f(x, y), \tag{4.8}$$

where $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$.
T. Yoon and E.K. Ryu[2] introduced EAG as an accelerated algorithm for solving (4.8)

$$\begin{cases} z_{k+1/2} = z_k + \beta_k(z_0 - z_k) - \alpha_k g(z_k), & \text{(4.9a)} \\ z_{k+1} = z_k + \beta_k(z_0 - z_k) - \alpha_k g(z_{k+1/2}) & \text{(4.9b)} \end{cases}$$

for $k \geq 0$, where $\beta_k \in [0, 1)$, known as anchoring coefficients, and $\alpha_k \in (0, 1)$, the step-size, for all $k \geq 0$, $z_0 \in H := H_1 \times H_2$ is an initial point. [Recall $g(z) = \partial f(z)$.]

[2] T. Yoon and E. K. Ryu, 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, 2021.

Extra Anchored Gradient (EAG)

Assuming f is L-smooth, Yoon and Ryu studied the convergence rate of $\|\nabla f(z^k)\|^2 = O(1/k^2)$ for two variants of EAG (both with $\beta_k = \frac{1}{k+2}$):

- EAG with constant step-size (EAG-C): $\alpha_k = \alpha$ for all k;
- EAG with varying step-size (EAG-V).

They however have not discussed the convergence of the iterates $\{z_k\}$.

EAG-C and EAG-V

EAG-C:

$$\begin{cases} z_{k+1/2} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha g(z_k), & (4.10a) \\ z_{k+1} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha g(z_{k+1/2}). & (4.10b) \end{cases}$$

EAG-V:

$$\begin{cases} z_{k+1/2} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha_k g(z_k), & (4.11a) \\ z_{k+1} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha_k g(z_{k+1/2}). & (4.11b) \end{cases}$$

Recall again $g(z) = \partial f(z)$.

Halpern standardization of EAG

We observe that the algorithm EAG (4.9) can be written as a standard two-step Halpern iteration. In fact, we have EAG (4.9) is equivalent to, using the notation $\bar{z}_k := z_{k+1/2}$,

$$\begin{cases} \bar{z}_k = \beta_k z_0 + (1-\beta_k)(z_k - \gamma_k g(z_k)), & \text{(4.12a)} \\ z_{k+1} = \beta_k z_0 + (1-\beta_k)(z_k - \gamma_k g(\bar{z}_k)), & \text{(4.12b)} \end{cases}$$

where $\gamma_k = \frac{\alpha_k}{1-\beta_k}$.
We notice that the unconstrained Korpelvich extragradient (EG) algorithm updates the the iterates via

$$\begin{cases} \bar{z}_k = z_k - \alpha g(z_k) & \text{(4.13a)} \\ z_{k+1} = z_k - \alpha g(\bar{z}_k). & \text{(4.13b)} \end{cases}$$

We may view the EAG (4.12) as a Halperned (or anchored) EG with varying stepsizes.

Anchored EG

For the constrained minimax problem (1.1), the anchored EG turns out to bo of the form

$$\begin{cases} \bar{z}_k = \beta_k z_0 + (1-\beta_k) P_G(z_k - \gamma_k g(z_k)), & \text{(4.14a)} \\ z_{k+1} = \beta_k z_0 + (1-\beta_k) P_G(z_k - \gamma_k g(\bar{z}_k)) & \text{(4.14b)} \end{cases}$$

or

$$\begin{cases} \bar{z}_k = P_G[\beta_k z_0 + (1-\beta_k)(z_k - \gamma_k g(z_k))], & \text{(4.15a)} \\ z_{k+1} = P_G[\beta_k z_0 + (1-\beta_k)(z_k - \gamma_k g(\bar{z}_k))] & \text{(4.15b)} \end{cases}$$

These two algorithms have the same convergence.

Another EAG

If the z_k in the EAG (4.12b) is updated also by the midway point $\bar{z}_k$, we obtain another EAG as follows:

$$\begin{cases} \bar{z}_k = \beta_k z_0 + (1-\beta_k)(z_k - \gamma_k g(z_k)), & \text{(4.16a)} \\ z_{k+1} = \beta_k z_0 + (1-\beta_k)(\bar{z}_k - \gamma_k g(\bar{z}_k)). & \text{(4.16b)} \end{cases}$$

Convergence of EAG

Theorem

(X., 2021) Assume (i) f is L-smooth, (ii) $0 < \gamma_* \le \gamma_k \le \gamma^* < \frac{2}{L}$, and (iii) (β_k) satisfies the conditions:

(C1) $\beta_k \to 0$

(C2) $\sum_k \beta_k = \infty$.

Then the sequence $\{z_k\}$ generated by EAG (4.16) *converges in norm to the solution $z^* = P_{G^*} z_0$.*

Remark: The anchor w may differ from the initial point z_0. In that case, the limit $z^* = P_{G^*} w$.

References I

F. E. Browder, Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces, Arch. Rational Mech. Anal. 24 (1967), 82-90.

J. Diakonikolas, Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. arXiv:2002.08872v3. (Apr 2020)

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

B. Halpern, Fixed points of nonexpanding maps, Bull. Amer. Math. Soc. 73 (1967), 591-597.

Ulrich Kohlenbach, On quantitative versions of theorems due to F.E. Browder and R. Wittmann, Advances in Mathematics 226 (2011), 2764-2795.

References II

- U. Kohlenbach, On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces, Israel J. Math. 216 (2016), 215-246.
- U. Kohlenbach and L. Leustean, Effective metastability of Halpern iterates in CAT(0) spaces, Adv. Math. 231 (2012), 2526-2556.
- D. Körnlein, Quantitative results for Halpern iterations of nonexpansive mappings, J. Math. Anal. Appl. 428 (2015), 1161-1172.
- L. Leustean, Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings, Journal of Universal Computer Science, vol. 13, no. 11 (2007), 1680-1691.
- F. Lieder, On the convergence rate of the Halpern-iteration, Optiminzation Letters 15 (2021), 405-418.
- P. L. Lions, Approximation des points fixes de contractions, C. R. Acad. Sci. Ser. A-B Paris 284 (1977), 1357-1359.

References III

- P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM Journal on Numerical Analysis 16 (1979), no. 6, 964-979.
- G. Lopez, V. Martin-Marquez, and H.K. Xu, Halpern's iteration for nonexpansive mappings, Contemporary Mathematics, 513 (2010), 211-231.
- A. Moudafi, Viscosity approximation methods for fixed-points problems, J. Math. Anal. Appl. 241 (2000), 46-55.
- P. Pinto, A rate of metastability for the Halpern type proximal point algorithm, Numer. Funct. Anal. Optim. 42 (2021), no. 3, 320-343.
- R. Wittmann, Approximation of fixed points of nonexpansive mappings, *Arch. Math.* **58** (1992), 486-491.
- S. Reich, Approximating fixed points of nonexpansive mappings, Panamerican. Math. J. 4 (1994), no. 2, 23-28.

References IV

- H. K. Xu, Iterative algorithms for nonlinear operators, *J. London Math. Soc.* **66** (2002), 240-256.
- H. K. Xu, Viscosity approximation methods for nonexpansive mappings, J. Math. Anal. Appl. 298 (2004) 279-291.
- A. Yadav, S. Shah, Z. Xu, D. Jacobs, and T. Goldstein, Stabilizing adversarial nets with prediction methods. arXiv:1705.07364v3 [cs.LG] 8 Feb 2018.
- T. Yoon and E. K. Ryu, 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, 2021.

Thank you for your attention!