Recent Progresses on Halpern's Iteration Method and Its Applications in Optimization

Hong-Kun Xu

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

Talk at
BIRS Workshop: New Frontiers in Proofs and Computation (Online)
Zhejiang University, Hangzhou, China
12-17 September 2021

Abstract

Halpern's iteration method, invented by Benjamin Halpern in 1967, is a fixed point algorithm for finding fixed points of a nonexpansive mapping. Since many optimization problems can be cast as fixed point problems of nonexpansive mappings, Halpern's method plays an important role in optimization methods. We will talk about some recent progresses on convergence and rate of convergence results of Halpern's method and its various applications in optimization problems, including variational inequalities, monotone inclusions, Douglas-Rachford splitting method, and minimax problems.

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method

2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method

3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

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

Projection

Let C be a nonempty closed convex subset of H. The projection from H onto C, P_C, is defined by

$$P_C x := \arg\min_{z \in C} \|x - z\|^2, \quad x \in H.$$

It is known that P_C is $\frac{1}{2}$-AV. That is, $2P_C - I$ is nonexpansive. Equivalently,

$$\langle x - y, P_C x - P_C y \rangle \geq \|P_C x - P_C y\|^2, \quad x, y \in H.$$

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Projected Gradient Method (PGM)

PGM is used to solve a constrained convex optimization problem of the form:

$$\min_{x \in C} \varphi(x), \tag{1.1}$$

where C is a nonempty closed convex subset of a Hilbert space H, and φ is a proper lower semicontinuous convex function; Notation: $\varphi \in \Gamma_0(H)$.
If φ is continuously differentiable, the minimization problem (1.1) can be solved by the projection-gradient method (PGM) which generates a sequence (x_n) by the iterative algorithm:

$$x_{n+1} = P_C(x_n - \lambda \nabla \varphi(x_n)) = P_C(I - \lambda \nabla \varphi) x_n, \quad n \geq 0, \tag{1.2}$$

where the initial guess $x_0 \in C$, and λ is the constant stepsize. The mapping

$$P_C(I - \lambda \nabla \varphi)$$

is averaged (hence, nonexpansive) if $\nabla \varphi$ is L-Lipschitz

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

and $0 < \lambda < 2/L$. And $P_C(I - \frac{2}{L} \nabla \varphi)$ is nonexpansive.

Proximal Gradient Method (ProxGM)

Consider the composite optimization problem

$$\min_{x \in H} f(x) + g(x) \tag{1.3}$$

where $f, g \in \Gamma_0(H)$.

(1.3) can be solved by ProxGM with constant stepsize:

$$x_{n+1} = \text{prox}_{\lambda g}(x_n - \lambda \nabla f(x_n)), \quad n = 0, 1, 2, \cdots . \tag{1.4}$$

Here

$$\text{prox}_{\lambda g}(x) := \min_{z \in H} \left(g(z) + \frac{1}{2\lambda} \|z - x\|^2 \right), \quad x \in H \tag{1.5}$$

is the proximal mapping of g.
$\text{prox}_{\lambda g}(I - \lambda \nabla f)$ is averaged if ∇f is L-Lipschitz and $0 < \lambda < 2/L$.
$\text{prox}_{\lambda g}(I - \frac{2}{L} \nabla f)$ is nonexpansive.

Lasso

The lasso may be stated as the optimization problem below, a special case of (1.3):

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \tag{1.6}$$

where $A \in \mathbb{R}^{m \times n}$ is an $m \times n$ matrix, $b \in \mathbb{R}^m$. Under certain conditions (e.g. RIP, restricted isometry property), (1.6) is equivalent to the sparsest recovery optimization problem

$$\min \|x\|_0 \quad \text{subject to } Ax = b + e \text{ with } \|e\|_2 \leq \mu. \tag{1.7}$$

$$x_{n+1} = \operatorname{prox}_{\lambda\|\cdot\|_1}(x_n - \lambda(A^\top(Ax_n - b)), \quad n = 0, 1, 2, \cdots . \tag{1.8}$$

Here $A^\top$ is the transpose of A.

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Halpern's Iteration Method

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

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) T x_n, \quad n = 0, 1, 2, \cdots, \tag{1.9}$$

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 (1.9) 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.

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., $\alpha_n = \frac{1}{(n+1)^\alpha}$, $0 < \alpha \leq 1$)

Inertial Halpern

The inertial Halpern's method produces a sequence $\{x_n\}$ in the following recurrence manner:

$$\begin{cases} y_n = x_n + \beta_n(x_n - x_{n-1}), \quad n \geq 0, & \text{(1.10a)} \\ x_{n+1} = \alpha_n u + (1-\alpha_n)Ty_n, \quad n \geq 0, & \text{(1.10b)} \end{cases}$$

where $\beta_n \in [0,1)$ and $\alpha_n \in (0,1)$ for all $n \geq 0$, x_{-1}, $x_0 \in H$ are given, and $u \in H$ is an anchor.

Theorem

(H. Qi and X., 2021)[a] *Let H be a Hilbert space, $T : H \to H$ nonexpansive such that $Fix(T) \neq \emptyset$, and (x_n) be generated by* (1.10). *Assume*

(i) $\{\alpha_n\}$ *satisfies the conditions (C1), (C2) and (C7).*

(ii) $\beta_n\|x_n - x_{n-1}\| = o(\alpha_n)$, *that is,* $\lim_{n\to\infty} \beta_n\|x_n - x_{n-1}\|/\alpha_n = 0$.

Then $\{x_n\}$ converges in norm to $u^* := P_{Fix(T)}u$.

[a] H. Qi and H.K. Xu, Convergence of Halpern's iteration method with applications in optimization (preprint).

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.

Remark: Halpern for Averaged Mappings

A mapping $T : H \to H$ is said to be an averaged mapping (AV) if

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

where $\alpha \in (0, 1)$ and $V : H \to H$ is nonexpansive (i.e., $\|Vx - Vy\| \leq \|x - y\|$ for all $x, y \in H$). In this case, T is also called α-AV.

Remark: If T is averaged with $Fix(T) \neq \emptyset$, then Halpern's method (1.9) converges in norm under the conditions (C1) and (C2) to $P_{Fix(T)}u$.

Regularization

Halpern's method is implicit in nature. Indeed, Browder (1965-68)[2] was the first to regularize nonexpansive mappings by contractions. Given $C \neq \emptyset$, closed convex, and $t \in (0, 1)$. Define

$$T_t x := tu + (1 - t)Tx$$

for $x \in C$. Then T_t is a contraction on C and hence has a unique fixed point $x_t \in C$. Browder proved in a Hilbert space that

$$\text{strong} - \lim_{t \to 0} x_t = P_{Fix(T)} u. \tag{1.11}$$

Discretizing Browder's implicit regularization yields Halpern's iteration method.

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

Viscosity Approximation Method (VAM)

VAM is a generalization of Browder's regularization by replacing the anchor with a contraction mapping.

Let $f : C \to C$ be a ρ-contraction for some $\rho \in [0, 1)$:

$$\|f(x) - f(y)\| \le \rho\|x - y\|, \quad x, y \in H.$$

Then the mapping $x \mapsto tf(x) + (1 - t)Tx$ is a contraction on C with contraction coefficient $1 - (1 - \rho)t \in [0, 1)$ for $t \in (0, 1)$, and hence has a unique fixed point denoted x_t.

Theorem

(Moudafi, 2000; X., 2004) We have that the strong-$\lim_{t\to 0} x_t =: \tilde{x}$ *exists and is a solution to the variational inequality (VI)*

$$\langle (I - f)\tilde{x}, z - \tilde{x}\rangle \le 0, \quad z \in C.$$

A. Moudafi, Viscosity approximation methods for fixed-points problems, J. Math. Anal. Appl. 241 (2000), 46-55.

H. K. Xu, Viscosity approximation methods for nonexpansive mappings, J. Math. Anal. Appl. 298 (2004) 279-291.

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Krasnoselskii-Mann Method

A seemingly more popular fixed point method is the Krasnoselskii-Mann method (KM):

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n T x_n, \quad n = 0, 1, \cdots, \tag{1.12}$$

where $T : C \to C$ is nonexpansive, $(\alpha_n) \subset [0, 1]$.

- M. A. Krasnoselskii, Two remarks on the method of successive approximations, (Russian), Uspehi Math. Nauk. 10 (1955), 123-127.
- W. R. Mann, Mean value methods in iteration, Proc. Amer. Math. Soc. 4 (1953), 506–510.

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method

2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method

3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Rate of Convergence

Ulrich Kohlenbach and his group, including Genaro Lopez-Acedo, Adriana Nicolae, L. Leustean, Felix Lieder, Pedro Pinto, have, using proof mining, made significant contributions to the study of the rate of convergence of Halpern iteration method in Hilbert, Banach and CAT(0) spaces, in a series of articles, a small number of which are listed below (details are however not discussed here).

- U. Kohlenbach, On quantitative versions of theorems due to F.E. Browder and R. Wittmann, Adv. Math. 226 (2011), 2764-2795.
- U. Kohlenbach and L. Leustean, Effective metastability of Halpern iterates in CAT(0) spaces, Adv. Math. 231 (2012), 2526-2556.
- U. Kohlenbach, On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces, Israel J. Math. 216 (2016), 215-246.
- L. Leustean, Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings, J. Universal Computer Sci. 13 (2007), no. 11 (2007), 1680-1691.
- Pedro Pinto, A rate of metastability for the Halpern type proximal point algorithm, Numer. Funct. Anal. Optim. 42 (2021), no. 3, 320-343.

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:

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) T x_n, \tag{2.1}$$

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

$$\|x_n - Tx_n\| \leq \frac{2\|x_0 - x^*\|}{n+1}, \quad n \geq 0, \; x^* \in Fix(T). \tag{2.2}$$

This bound is tight.

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

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Convergence Rate of KM

Recall KM:

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T x_n.$$

Theorem

(R. Cominetti and J. A. Soto, 2014) Let X be a Banach space, C a nonempty closed convex bounded subset of X, and $T : C \to C$ nonexpansive. Then

$$\|x_n - Tx_n\| \leq \frac{\operatorname{diam}(C)}{\sqrt{\pi \sum_{i=0}^{n} \alpha_i (1 - \alpha_i)}}. \tag{2.3}$$

- R. Cominetti and J. A. Soto, On the rate of convergence of Krasnosel'skii-Mann iterations and their connections with sum of Bernoullis, Israel J. Math. 199 (2014), 757-772.
- M. Bravo, R. Cominetti, and M. Pavez-Signe, Rates of convergence for inexact Krasnosel'skii-Mann iterations in Banach spaces, Math. Program. 175 (2019), 241-262.

Sharpness

Theorem

(M. Bravo and R. Cominetti, 2018) The constant $\kappa := 1/\sqrt{\pi}$ *in the bound (2.3) is tight. Specifically, for each* $\kappa < 1/\sqrt{\pi}$*, there exists a nonexpansive map* T *defined on the unit cube* $C = [0,1]^N \subset \ell^\infty(N)$*, an initial point* $x_0 \in C$*, and a constant sequence* $\alpha_n = \alpha$*, such that the corresponding KM iterates* (x_n) *satisfy for some* $n \in \mathbb{N}$

$$\|x_n - Tx_n\| > \kappa \frac{\mathrm{diam}(C)}{\sqrt{\sum_{i=0}^{n} \alpha_i(1-\alpha_i)}}. \tag{2.4}$$

M. Bravo and R. Cominetti, Sharp convergence rates for averaged nonexpansive maps, Israel J. Math. 227 (2018), 163-188.

Remark

Remark

At least, in the Hilbert space setting, we have infinitesimal $a_n \to 0$ *satisfying*

$$\|x_n - Tx_n\| \leq a_n \cdot \frac{\mathrm{diam}(C)}{\sqrt{\pi \sum_{i=0}^{n} t_i(1-t_i)}}.$$

That is,

$$\|x_n - Tx_n\| = o\left(\frac{\mathrm{diam}(C)}{\sqrt{\pi \sum_{i=0}^{n} t_i(1-t_i)}}\right).$$

We however do not know how to find a_n.

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

$$\langle F(u^*), u - u^* \rangle \geq 0, \quad u \in C. \tag{3.1}$$

VI (3.1) 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 $\gamma > 0$,

$$u^* = P_C(I - \gamma F)u^* \tag{3.2}$$

Hence, fixed point methods can be applied to solve VIs. It is also known that if F is strongly monotone and Lipschitz, then for appropriate $\gamma > 0$, $I - \gamma F$ and hence, $P_C(I - \gamma F)$, are contractions. In this case,

$$x_n := P_C(I - \gamma F)x_{n-1} \to u^*.$$

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 $\gamma > 0$,

$$\langle F(x) - F(y), x - y \rangle \geq \gamma \| F(x) - F(y) \|^2, \quad x, y \in H. \tag{3.3}$$

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

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method

2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method

3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

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 ($0 < \lambda < \frac{2}{L}$):

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

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

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

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

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Halpern Method for Monotone Operators

Consider the equation

$$Fx = 0$$

where F is $\frac{1}{L}$-ISM with zeros (i.e., $F^{-1}0 = \{x \in H : Fx = 0\} \neq \emptyset$). Halpern's method applied to the nonexpansive mapping $I - \frac{2}{L}F$ yields

$$x_{n+1} = \alpha_n u + (1-\alpha_n)\left(x_n - \frac{2}{L}Fx_n\right), \quad n = 0, 1, 2, \cdots . \tag{3.4}$$

It turns out that (x_n) converges in norm to $P_{F^{-1}0}u$ if (α_n) satisfies conditions (C1)-(C2) and one of (C3)-(C6). Moreover, if we take $\alpha_n = \frac{1}{n+2}$ and $u = x_0$, then, for any $x^* \in F^{-1}0$,

$$\|F(x_n)\| \leq \frac{L\|x_0 - x^*\|}{n+1} = O\left(\frac{1}{n}\right).$$

Hence, an ε-approximate solution (i.e., $\|F(x_n)\| \leq \varepsilon$) can be obtained after at most $(\frac{2L\|x_0 - x^*\|}{\varepsilon} + 1)$ iterations .

Halpern Method for Monotone VIs

Consider IV:

$$\langle F(u^*), u - u^* \rangle \geq 0, \quad u \in C, \tag{3.5}$$

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

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method
2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method
3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Zeros of the Sum of Two Maximal Operators

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

$$0 \in (A+B)x, \tag{3.6}$$

where A and B are two maximal (multivalued) monotone operators in a Hilbert space H. Assume the solution set of (3.6) 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 (3.6).

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 (3.6) [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 (3.6), 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 $u = x_0$:

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

Outline

1. Introduction
   - Nonexpansive Mappings
   - Projected Gradient Method
   - Halpern's Iteration Method
   - Krasnoselskii-Mann Method

2. Rate of Convergence
   - Rate of Convergence of Halpern Method
   - Rate of Convergence of KM Method

3. Variational Inequalities (VIs)
   - Halpern Method for Optimization
   - Halpern Method for Monotone IVs
   - Douglas-Rachford-Halpern Method
   - Halpern Method for Minimax Problems

Halpern Method for Minimax Problems

A minimax problem (game) is a optimization problem of the form

$$\min_{x \in H_1} \max_{y \in H_2} L(x, y), \tag{3.7}$$

where H_1 and H_2 are Hilbert spaces, and $L : H_1 \times H_2 \to \mathbb{R}$ is a real-valued loss function. Set $H := H_1 \times H_2$ equipped with the inner product and norm

$$\langle z, w \rangle = \langle x, u \rangle_{H_1} + \langle y, v \rangle_{H_2}, \quad \|(x, y)\| = \sqrt{\|x\|_{H_1}^2 + \|y\|_{H_2}^2}$$

for $z = (x, y)$ and $w = (u, v)$ in H.
The minimax problem (3.7) has recently been paid significant attention due to its important application in machine learning, in particular, generative adversarial networks (GANs)[3]

[3] 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.

Saddle Point

We will consider the convex-concave case, that is, for each fixed $y \in H_2$, the function $L(\cdot, y)$ is a convex function on H_1, and for each fixed $x \in H_1$, the function $L(x, \cdot)$ is a concave function on H_2. A point $(x^*, y^*) \in H$ is said to be a saddle point of L (also known as a solution of the minimax game (3.7)) if

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*) \tag{3.8}$$

for all $x \in H_1$ and $y \in H_2$. For the sake of simplicity, we put $z = (x, y)$ and $L(z) = L(x, y)$. Assume L is Fréchet differentiable. Recall that the saddle differential of L at z is defined as

$$L'(z) = \begin{bmatrix} \nabla_x L(x, y) \\ -\nabla_y L(x, y) \end{bmatrix}. \tag{3.9}$$

Since $-L(x, \cdot)$ is convex for each fixed $x \in H_1$, L' is a maximal monotone operator on $H := H_1 \times H_2$. Moreover, $L'(z) = 0$ if and only if z is a saddle point of L. Consequently, the problem of finding a saddle point of L is equivalent to finding a zero of a maximal monotone operator. The latter can be solved by fixed point methods. Recall also that L is said to be R-smooth if the saddle differential L' of L is R-Lipschitz on H.

Halpern for Minimax

Proposition

Assume the convex-concave objective function L *is* R*-smooth for some* $R \geq 0$*. 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:*

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

alternatively, in components,

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

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* $\hat{z} \in S$*,*

$$\|L'(z_n)\| \leq \frac{R\|z_0 - \hat{z}\|}{n+1}.$$

Extra Anchored Gradient (EAG)

T. Yoon and E.K. Ryu[4] introduced EAG as an accelerated algorithm for solving the convex-concave minimax problem (3.7):

$$\begin{cases} z_{k+1/2} = z_k + \beta_k(z_0 - z_k) - \alpha_k L'(z_k), & (3.10a) \\ z_{k+1} = z_k + \beta_k(z_0 - z_k) - \alpha_k L'(z_{k+1/2}) & (3.10b) \end{cases}$$

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 $k \geq 0$, $z_0 \in H := \mathbb{R}^n \times \mathbb{R}^m$ is an initial point.
Assuming L is R-smooth, Yoon and Ryu studied the convergence rate of $\|\nabla L(z^k)\|^2$ of 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).

[4] 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.

EAG-C and EAG-V

EAG-C:

$$\begin{cases} z_{k+1/2} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha L'(z_k), & \text{(3.11a)} \\ z_{k+1} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha L'(z_{k+1/2}). & \text{(3.11b)} \end{cases}$$

EAG-V:

$$\begin{cases} z_{k+1/2} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha_k L'(z_k), & \text{(3.12a)} \\ z_{k+1} = z_k + \frac{1}{k+2}(z_0 - z_k) - \alpha_k L'(z_{k+1/2}). & \text{(3.12b)} \end{cases}$$

Halpern standardization of EAG

We observe that the algorithm EAG (3.10) can be written as a standard two-step Halpern iteration. In fact, we have EAG (3.10) is equivalent to

$$\begin{cases} z_{k+1/2} = \beta_k z_0 + (1-\beta_k)(I - \gamma_k F) z_k, & \text{(3.13a)} \\ z_{k+1} = \beta_k z_0 + (1-\beta_k)(I - \gamma_k F) z_{k+1/2}, & \text{(3.13b)} \end{cases}$$

where $F = L'$ and $\gamma_k = \frac{\alpha_k}{1-\beta_k}$.

Convergence of EAG

Using the two-step Halpern formulation (3.13), we can prove the strong convergence of EAG (3.10) even in infinite-dim framework.

Theorem

(X., 2021) Assume (i) L *is* R*-smooth, (ii)* $0 < \gamma_* \le \gamma_k \le \gamma^* < \frac{2}{R}$*, and (iii)* (β_k) *satisfies (C1)-(C2). Then the sequence* $\{z_k\}$ *generated by EAG* (3.13) *converges in norm to the solution* $z^* = P_S z_0$*.*

Proof.

Omitted. □

Remark: Our result applies to EAG-C (3.11) and EAG-V (3.12). For instance, for EAG-C (3.11), we have $\gamma_k = \frac{\alpha_k}{1-\beta_k} = \frac{k+2}{k+1}\alpha$ since $\alpha_k = \alpha$ and $\beta_k = \frac{1}{k+2}$ for all $k \ge 1$. So, if $\alpha < \frac{2}{R}$, condition (ii) is satisfied for all k big enough, with $\gamma_* = \alpha$ and $\gamma^* = \alpha + \varepsilon$, $0 < \varepsilon < \frac{2}{R} - \alpha$.

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.

Thanks

Thank you for your attention!