Theoretical bound of the complexity of some extragradient-type algorithms for variational inequalities in Banach spaces

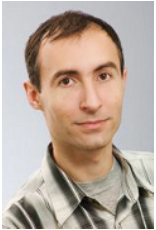
Vladimir Semenov, Sergey Denisov

Department of Computational Mathematics
Faculty of Computer Sciences and Cybernetics
TSNUK
{semenov.volodya, denisov.univ}@gmail.com

November 15 – November 19, 2021
MMOIT–2021

Our team

Vladimir Semenov



Sergey Denisov

Support

Ministry of Education and Science of Ukraine

Mathematical modeling and optimization of dynamic systems for defense, medicine and ecology,
0119U100337

National Academy of Sciences of Ukraine

New methods of research of correctness and solution search for discrete optimization problems, variational inequalities and their applications,
0119U101608

Variational Inequalities and Optimization

Let E be a real vector space with norm $\|\cdot\|$, we denote by E^* the dual of E (with dual norm $\|\cdot\|_*$) and $\langle f, x \rangle$ the value of $f \in E^*$ at $x \in E$.

Variational Inequality

$$\text{find} \quad x \in C \quad \text{such that} \quad \langle Ax, y - x \rangle \geq 0 \quad \forall\, y \in C \qquad (VI)$$

Smooth constrained optimization

$$\min_{x \in X} f(x)$$

Assumption: f is a convex smooth function, X is closed convex set

Optimality condition:

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X$$

Variational Inequalities and Optimization

Saddle point problems

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

Assumption: f is a convex-concave smooth function, X, Y are closed convex sets

Optimality condition:

$$\Bigg\langle \underbrace{\begin{pmatrix} \nabla_x f(x^*, y^*) \\ -\nabla_y f(x^*, y^*) \end{pmatrix}}_{Az^*}, \underbrace{\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}}_{z - z^*} \Bigg\rangle \geq 0 \quad \forall \underbrace{(x, y)}_{z} \in \underbrace{X \times Y}_{Z}$$

Variational Inequality

$$\text{find} \quad x \in Z \quad \text{such that} \quad \langle Az^*, z - z^* \rangle \geq 0 \quad \forall z \in Z$$

Assumption: A is monotone: $\langle Az_1 - Az_2, z_1 - z_2 \rangle \geq 0 \quad \forall z_1, z_2 \in Z$

Variational Inequalities and Optimization

1. Matrix games:

$$\min_{x\in\Delta^n}\max_{y\in\Delta^m}\langle Px, y\rangle$$

2. Constrained optimization:

$$\min_x f(x) \ \ s.t. \ \ g(x)\leq 0 \quad \longrightarrow \quad \min_x \max_{y\geq 0}(f(x)+yg(x))$$

3. Structural minimization:

$$f(x)+g(Ax)\to\min_x \quad \longrightarrow \quad \min_x\max_y(f(x)+\langle Ax, y\rangle - g^*(y))$$

4. Structural minimization (discrete maximum function):

$$\max_{i=1,\ldots,m} f_i(x)\to\min_x \quad \longrightarrow \quad \min_x\max_{y\in\Delta^m}\left(\sum_{i=1}^m y_i f_i(x)\right)$$

5. Machine Learning:

Adversarial training: $\min_x \sum_i L(x, a_i, b_i) \ \longrightarrow \ \min_x \max_{\omega\in\Omega}\sum_i L(x, a_i+\omega, b_i)$

Generative adversarial networks (GANs)

Some geometry of Banach spaces 1

Convexity:

$$\delta_E(\varepsilon) = \inf \left\{ 1 - \left\| \frac{x+y}{2} \right\| : \quad x, y \in S_E, \quad \|x - y\| = \varepsilon \right\} \quad \forall \varepsilon \in (0, 2]$$

- E is uniformly convex: $\delta_E(\varepsilon) > 0$ for all $\varepsilon \in (0, 2]$
- E is 2-uniformly convex: $\exists c > 0$ s. t. $\delta_E(\varepsilon) \geq c\varepsilon^2$ for all $\varepsilon \in (0, 2]$

Smoothness:

- E is smooth: $\lim_{t \to 0} \frac{\|x+ty\| - \|x\|}{t}$ exists for all x, $y \in S_E$.
- E is uniformly smooth: the limit is attained uniformly for x, $y \in S_E$.

$$\rho_E(t) = \sup \left\{ \frac{\|x+ty\| + \|x-ty\|}{2} - 1 : \quad x, y \in S_E \right\} \quad \forall t > 0$$

$$\lim_{t \to +0} \rho_E(t)/t = 0$$

- E is 2-uniformly smooth: $\exists c > 0$ s. t. $\rho_E(t) \leq ct^2$ for all $t > 0$

Example: L_p $(1 < p \leq 2)$ are 2-uniformly convex and uniformly smooth

Some geometry of Banach spaces 2[1]

Normalized duality mapping:

$$J : E \to 2^{E^*} : \quad Jx = \left\{ x^* \in E^* : \ \langle x^*, x \rangle = \|x\|^2 = \|x^*\|_*^2 \right\}.$$

If E is smooth, then J is single-valued ($J = \nabla \frac{1}{2} \| \cdot \|^2$).

- $Jx = \|x\|_{L_p}^{2-p} |x|^{p-2} x \in L_q, \ \ x \in L_p$
- $Jx = \|x\|_{W_p^m}^{2-p} \sum_{|\alpha| \le m} (-1)^{|\alpha|} D^\alpha (|D^\alpha x|^{p-2} D^\alpha x) \in W_q^{-m}, \ \ x \in W_p^m$

[1] I. Cioranescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, Kluwer Academic, Dordrecht, 1990.

Alber functional and Aoyama–Kohsaka inequality

Let E be a smooth Banach space. Consider the Alber functional[2]

$$\phi(x,y) = \|x\|^2 - 2\langle Jy, x\rangle + \|y\|^2 \quad \forall x, y \in E.$$

Hilbert space:

$$\phi(x,y) = \frac{1}{2}\|x-y\|^2.$$

Let E be a 2-uniformly convex and smooth Banach space. Then, for some $\mu \geq 1$, the inequality holds[3]

$$\phi(x,y) \geq \frac{1}{\mu}\|x-y\|^2 \quad \forall x, y \in E.$$

L_p, W_p^m $(1 < p \leq 2)$: $\mu = \frac{1}{p-1}$

---

[2] Y. I. Alber, Metric and generalized projection operators in Banach spaces: properties and applications. in: Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, vol. 178, Dekker, New York, 1996, pp. 15–50.

[3] K. Aoyama, F. Kohsaka, Strongly relatively nonexpansive sequences generated by firmly nonexpansive-like mappings, Fixed Point Theory Appl. 95 (2014). doi:10.1186/1687-1812-2014-95.

Xu inequality[4]

Let E be a 2-uniformly smooth Banach space. Then, for some $\kappa > 0$, the inequality holds

$$\|x + y\|^2 \leq \|x\|^2 + 2\langle Jx, y\rangle + \kappa\|y\|^2 \quad \forall x, y \in E.$$

L_p, W_p^m $(2 \leq p < +\infty)$: $\kappa = p - 1$

[4] H. K. Xu, Inequalities in Banach spaces with applications, Nonlinear Anal. 16(12) (1991) 1127–1138.

Alber generalized projection

Let K be a non-empty closed and convex subset of a reflexive, strictly convex and smooth space E. It is known that for each $x \in E$ there is a unique point $z \in K$ such that

$$\phi(z,x) = \inf_{y \in K} \phi(y,x)\,.$$

This point z is denoted by $\Pi_K x$, and the corresponding operator $\Pi_K : E \to K$ is called the generalized projection of E onto K.

Let K be a closed and convex subset of a reflexive, strictly convex and smooth space E, $x \in E$, $z \in K$. Then

$$z = \Pi_K x \quad \Leftrightarrow \quad \langle Jz - Jx, y - z\rangle \geq 0 \quad \forall y \in K \quad \Leftrightarrow$$
$$\Leftrightarrow \quad \phi(y, \Pi_K x) + \phi(\Pi_K x, x) \leq \phi(y,x) \quad \forall y \in K.$$

Variational Inequality[5]

Let E be a 2-uniformly convex and uniformly smooth Banach space, $C \subseteq E$, $A : E \to E^*$.

$$\text{find } \ x \in C \ \text{ such that } \quad \langle Ax, y - x \rangle \geq 0 \quad \forall y \in C \tag{1}$$

We denote the solution set of the problem (1) by S.

Assumption:

- set $C \subseteq E$ is convex and closed;
- operator $A : E \to E^*$ is monotone and Lipschitzian-type with $L > 0$ on C;
- set S is non-empty.

We can formulate (1) as fixed point problem:

$$x = \Pi_C J^{-1} \left(Jx - \lambda Ax \right), \tag{2}$$

where $\lambda > 0$. Formulation (2) is useful because is contains an obvious algorithmic idea (**but it doesn't work here**).

[5] Y. Alber, I. Ryazantseva, Nonlinear Ill Posed Problems of Monotone Type, Springer, Dordrecht, 2006.

Algorithms

Alg. 1. Y. Shehu, 2020: Banach version of the Forward-Backward-Forward Algorithm (P. Tseng, 2000)

$$
\begin{aligned}
y_n &= \Pi_C J^{-1} \left(J x_n - \lambda_n A x_n \right), \\
x_{n+1} &= J^{-1} \left(J y_n - \lambda_n \left(A y_n - A x_n \right) \right).
\end{aligned}
$$

Alg. 2. Extrapolation from the Past (Popov, S.–Malitsky, Gidel,...)

$$
\begin{aligned}
y_n &= \Pi_C J^{-1} \left(J x_n - \lambda_n A y_{n-1} \right), \\
x_{n+1} &= \Pi_C J^{-1} \left(J x_n - \lambda_n A y_n \right).
\end{aligned}
$$

Alg. 3. Operator extrapolation (Alber projection setting) (Malitsky–Tam, Vedel–S.–Denisov)

$$
x_{n+1} = \Pi_C J^{-1} \left(J x_n - \lambda_n A x_n - \lambda_n \left(A x_n - A x_{n-1} \right) \right).
$$

ε-solution

Definition 1. $z \in C$ is ε-solution if

$$Gap(z) \le \varepsilon,$$

where

$$Gap(z) = \sup_{y \in C} \langle Ay, z - y \rangle .$$

Alg. 1

Theorem 1. $\lambda_n \in \left(0, \frac{1}{\sqrt{\mu\kappa}L}\right]$, $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$:

$$Gap(z_N) \leq \frac{\sup_{y \in C} \phi(y, x_1)}{2\sum_{n=1}^N \lambda_n}.$$

Corollary 1. $\lambda = \frac{1}{2\sqrt{\mu\kappa}L}$, $z_N = \frac{1}{N}\sum_{n=1}^N y_n$:

$$Gap(z_N) \leq \frac{L\sqrt{\mu\kappa}\sup_{y \in C} \phi(y, x_1)}{N}.$$

Alg. 2

Theorem 2. $\lambda_n \in \left(0, \frac{\sqrt{2}-1}{\mu L}\right]$, $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$:

$$Gap(z_N) \leq \frac{\sup_{y \in C} \phi(y, x_1) + \mu \lambda_1 L \phi(x_1, y_0)}{2 \sum_{n=1}^N \lambda_n}.$$

Corollary 2. $\lambda = \frac{1}{3\mu L}$, $z_N = \frac{1}{N} \sum_{n=1}^N y_n$:

$$Gap(z_N) \leq \mu L \frac{\frac{3}{2} \sup_{y \in C} \phi(y, x_1) + \frac{1}{2} \phi(x_1, y_0)}{N}.$$

Alg. 3

Theorem 3. $\lambda_n \in \left(0, \frac{1}{2\mu L}\right]$, $z_{N+1} = \frac{\sum_{n=1}^{N} \lambda_n x_{n+1}}{\sum_{n=1}^{N} \lambda_n}$:

$$Gap(z_{N+1}) \leq \frac{\sup_{y \in C} \phi(y, x_1)}{2 \sum_{n=1}^{N} \lambda_n}.$$

Corollary 3. $\lambda = \frac{1}{2\mu L}$, $z_{N+1} = \frac{1}{N} \sum_{n=1}^{N} x_{n+1}$:

$$Gap(z_{N+1}) \leq \frac{\mu L \sup_{y \in C} \phi(y, x_1)}{N}.$$

Numerical experiment with toy saddle point problem

We compare Alg. 1, Alg. 2, and Alg. 3 (Euclidean setting) with their Bregman versions (Entropy setting) (Alg. 1*, Alg. 2*, and Alg. 3*).

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

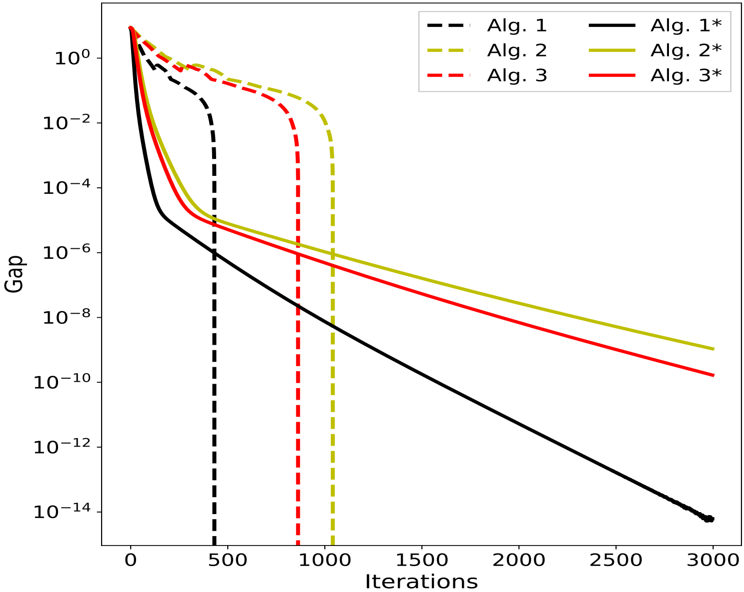
P is an $m \times n$ matrix, $X = \Delta^n = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$ and $Y = \Delta^m$.

$$Az = A(x, y) = \begin{pmatrix} P^* y \\ -Px \end{pmatrix}, \quad C = X \times Y.$$

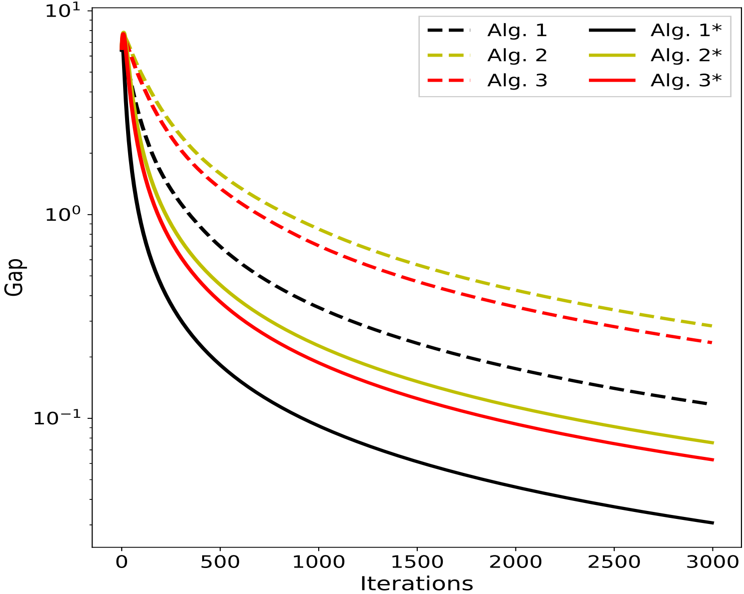
$$G(z) = \max_{v\in C}(Av, z - v) = \max_i (Px)_i - \min_j (P^* y)_j$$

$n = 100$, $m = 150$. Starting points $(1/n, ..., 1/n) \in \Delta^n$, $(1/m, ..., 1/m) \in \Delta^m$.

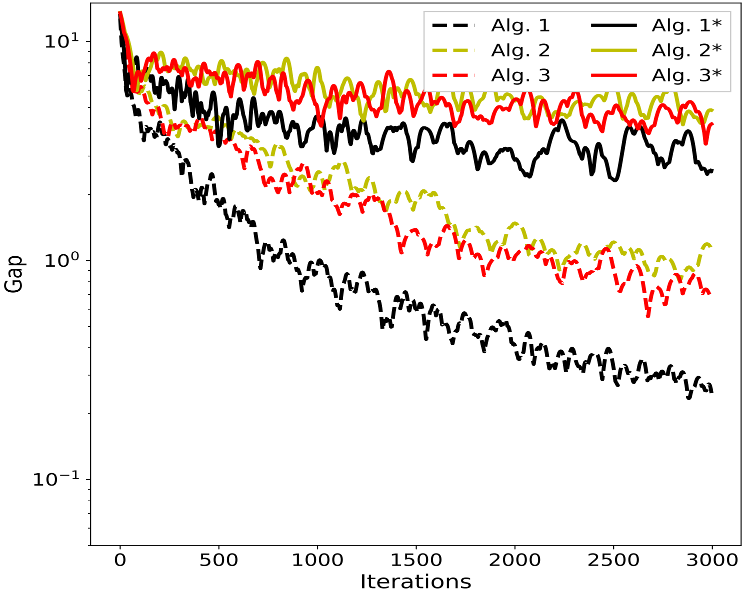
Numerical experiment with toy saddle point problem



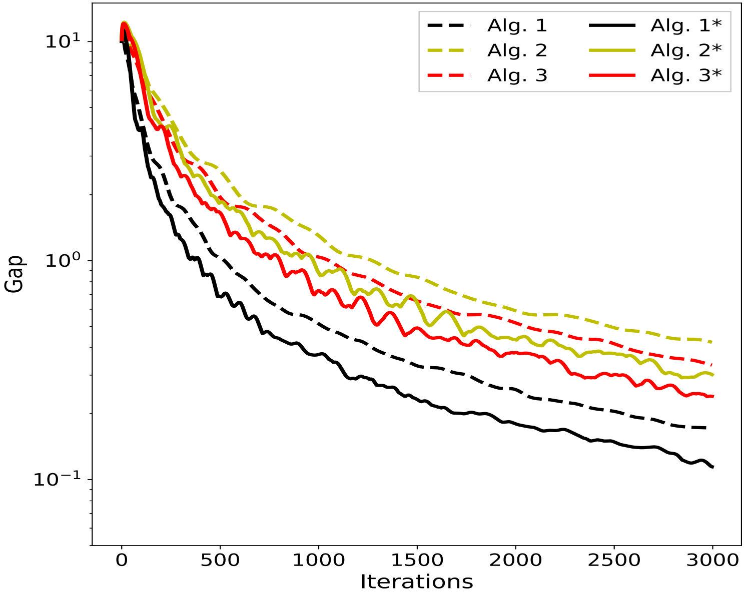
Numerical experiment with toy saddle point problem



Numerical experiment with toy saddle point problem



Numerical experiment with toy saddle point problem



Thanks for your attention!