Algorithms for saddle point problems with some properties of generalized smoothness

joint work with F. Stonyakin, M. Alkousa and A. Gasnikov

Alexander Titov
a.a.titov@phystech.edu

Higher School of Economics
Moscow Institute of Physics and Technology

April 2021

1 Content

1 Joint Work

Titov, A., Stonyakin, F., Alkousa, M., Gasnikov, A. (2021). Algorithms for solving variational inequalities and saddle point problems with some generalizations of Lipschitz property for operators, arXiv preprint arXiv:2103.00961 https://arxiv.org/pdf/2103.00961.pdf

2 Content

2 Problem Classes

Definition 1 (Saddle Point Problem)

Consider (μ_x, μ_y)-strongly convex-concave saddle point problem:

$$\min_{x \in Q_x} \max_{y \in Q_y} f(x, y), \tag{1}$$

Q_x, Q_y are nonempty, convex, compact and bounded sets.

Definition 2 (Minty Variational Inequality)

For a given operator $g(x) : X \to \mathbb{R}$, where X is a closed convex subset of some finite-dimensional vector space, we need to find a vector $x_* \in X$, such that

$$\langle g(x), x_* - x \rangle \leq 0, \quad \forall x \in X. \tag{2}$$

2 ε–solutions

Definition 3 (Saddle point problem)

$$\max_{y} f(\tilde{x}, y) - \min_{x} f(x, \tilde{y}) \leq \varepsilon. \tag{3}$$

Definition 4 (Variaional inequality)

$$\max_{x \in Q} \langle g(x), \tilde{x} - x \rangle \leq \varepsilon + \sigma. \tag{4}$$

MIPT

NATIONAL RESEARCH UNIVERSITY

2 Connection between saddle point problem and VI

Any saddle point problem

$$\min_{x \in Q_x} \max_{y \in Q_y} f(x, y)$$

can be reduced to a variational inequality problem by considering the following operator:

$$g(z) = \begin{pmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{pmatrix}, \; z = (x, y) \in Q := Q_x \times Q_y. \tag{5}$$

2 Hölder continuity

Let f satisfy [1]

$$\|\nabla_x f(x,y) - \nabla_x f(x',y)\|_2 \le L_{xx}\|x - x'\|_2^\nu, \tag{6}$$

$$\|\nabla_x f(x,y) - \nabla_x f(x,y')\|_2 \le L_{xy}\|y - y'\|_2^\nu, \tag{7}$$

$$\|\nabla_y f(x,y) - \nabla_y f(x',y)\|_2 \le L_{xy}\|x - x'\|_2^\nu, \tag{8}$$

$$\|\nabla_y f(x,y) - \nabla_y f(x,y')\|_2 \le L_{yy}\|y - y'\|_2, \tag{9}$$

for any $x, x' \in Q_x, y, y' \in Q_y$ and for some $\nu \in [0,1]$.

Definition 5

Function $h(x) : Q_x \to \mathbb{R}$ is called μ-strongly convex, if

$$\langle \nabla h(x_1) - \nabla h(x_2), x_1 - x_2 \rangle \ge \mu\|x_1 - x_2\|_2^2 \quad \forall x_1, x_2 \in Q_x.$$

[1] Alkusa, M.S., Gasnikov, A.V., Dvinskikh, D.M., Kovalev, D.A., Stonyakin, F.S. (2020). Accelerated methods for saddle problems. Journal of Computational Mathematics and Mathematical Physics, 60 (11), 1843-1866.

Smallest covering circle problem with non-smooth functional constraints.

$$\min_{x\in Q}\left\{f(x) := \max_{1\leq k\leq N}\|x - A_k\|_2^2;\ \varphi_p(x)\leq 0,\ p=1,...,m\right\}, \qquad (10)$$

where $A_k \in \mathbb{R}^n, k = 1, ..., N$ are given points and Q is a convex compact set. Functional constraints φ_p, for $p = 1, ..., m$, have the following form:

$$\varphi_p(x) := \sum_{i=1}^{n} \alpha_{pi}x_i + \beta_{pi},\ p = 1, ..., m. \qquad (11)$$

Smallest covering circle problem with non-smooth functional constraints.

$$\min_{x\in Q}\left\{f(x) := \max_{1\le k\le N} \|x - A_k\|_2^2;\ \varphi_p(x) \le 0,\ p = 1, ..., m\right\}, \tag{10}$$

where $A_k \in \mathbb{R}^n, k = 1, ..., N$ are given points and Q is a convex compact set. Functional constraints φ_p, for $p = 1, ..., m$, have the following form:

$$\varphi_p(x) := \sum_{i=1}^{n} \alpha_{pi}x_i + \beta_{pi},\ p = 1, ..., m. \tag{11}$$

The corresponding Lagrange saddle point problem:

$$\min_{x\in Q} \max_{\vec{\lambda}=(\lambda_1,\lambda_2,...,\lambda_m)^T\in\mathbb{R}^m_+} L(x,\lambda) := f(x) + \sum_{p=1}^{m} \lambda_p\varphi_p(x).$$

This problem is equivalent to the VI with monotone non-smooth operator

$$G(x,\lambda) = \begin{pmatrix} \nabla f(x) + \sum\limits_{p=1}^{m} \lambda_p \nabla\varphi_p(x), \\ (-\varphi_1(x), -\varphi_2(x), \dots, -\varphi_m(x))^T \end{pmatrix}.$$

Hölder continuity

- an important generalization of the Lipschitz condition
- if a function is uniformly convex, then its conjugated will necessarily have the Hölder-continuous gradient

Strong convexity

Strong convex-concavity for the smallest covering circle problem can be obtained with the term

$$-\frac{1}{2}\frac{\|\lambda\|_2^2}{R^2}$$

2 Mirror Descent Basics

Let E be a finite-dimensional real vector space and E^* be its dual. We denote the value of a linear function $g \in E^*$ at $x \in E$ by g, x. Let $\|\cdot\|_E$ be some norm on E, $\|\cdot\|_{E,*}$ be its dual, defined by

$$\|g\|_{E,*} = \max_x \left\{ \langle g, x \rangle, \|x\|_E \leqslant 1 \right\}$$

We use $\nabla f(x)$ to denote any subgradient of a function f at a point $x \in \mathrm{dom} f$.
We choose a *prox-function* $d(x)$, which is continuous, convex on X and

1. admits a continuous gradient $\nabla d(x)$, where $x \in X$;
2. Let $d(x)$ be 1-strongly convex on X with respect to $\|\cdot\|_E$

The corresponding *Bregman divergence*

$$V(x,z) = d(x) - d(z) - \langle \nabla d(z), x - z \rangle, \ x, z \in X$$

Given a vector $x \in X$, and a vector $g \in E^*$, the Mirror Descent step is defined as

$$\mathrm{Mirr}(x,g) := \arg\min_{y \in Q} \left\{ \langle g, y \rangle + V(y,x) \right\}.$$

Assume that ${}_{x \in X} d(x) = 0$ and $d(\cdot)$ is bounded on the unit ball in the chosen norm $\|\cdot\|$, more precisely

$$d(x) \leq \frac{\Omega}{2}, \quad \forall x \in X : \|x\| \leq 1,$$

where Ω is a known constant.

3 Universal Proximal method[2]

Definition 6 (Inexactly given operator)

Suppose that there exist some $\delta > 0, L(\delta) > 0$, such that, for any points $x, y, z \in X$, we are able to calculate $\tilde{g}(x,\delta), \tilde{g}(y,\delta) \in E^*$, satisfying

$$\langle \tilde{g}(y,\delta) - \tilde{g}(x,\delta), y - z \rangle \leq \frac{L(\delta)}{2}\left(\|y-x\|^2 + \|y-z\|^2\right) + \delta, \quad \forall z \in Q. \tag{12}$$

[2] Dvurechensky, P., Gasnikov, A., Stonyakin, F., Titov, A. (2018). Generalized Mirror Prox: Solving variational inequalities with monotone operator, inexact oracle, and unknown Hölder parameters

Algorithm 1 Universal Mirror Prox (UMP)

Require: $\varepsilon > 0$, $\delta > 0$, $x_0 \in X$, initial guess $L_0 > 0$, prox-setup: $d(x)$, $V(x, z)$.

1: Set $k = 0$, $z_0 = \arg\min_{u \in Q} d(u)$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Set $M_k = L_k/2$.
4: Set $\delta = \frac{\varepsilon}{2}$.
5: **repeat**
6: Set $M_k = 2M_k$.
7: Calculate $\tilde{g}(z_k, \delta)$ and

$$w_k = \arg\min_{x \in Q}{}^{\delta} \left\{ \langle \tilde{g}(z_k, \delta), x \rangle + M_k V(x, z_k) \right\}.$$

8: Calculate $\tilde{g}(w_k, \delta)$ and

$$z_{k+1} = \arg\min_{x \in Q}{}^{\delta} \left\{ \langle \tilde{g}(w_k, \delta), x \rangle + M_k V(x, z_k) \right\}.$$

9: **until**

$$\langle \tilde{g}(w_k, \delta) - \tilde{g}(z_k, \delta), w_k - z_{k+1} \rangle \le \frac{M_k}{2} \left(\|w_k - z_k\|^2 + \|w_k - z_{k+1}\|^2 \right) + \frac{\varepsilon}{2} + \delta.$$

3 Restarted Universal Mirror Prox

Definition 7 (μ–strongly monotonicity)

$$\langle g(x) - g(y), x - y \rangle \geq \mu \|x - y\|^2, \quad \forall x, y \in X.$$

Algorithm 2 Restarted Universal Mirror Prox (`Restarted UMP`).

Require: $\varepsilon > 0$, $\mu > 0$, $\Omega : d(x) \leq \frac{\Omega}{2}\ \forall x \in Q : \|x\| \leq 1$; x_0, $R_0\ : \|x_0 - x_*\|^2 \leq R_0^2$.

1: Set $p = 0, d_0(x) = R_0^2 d\left(\frac{x - x_0}{R_0}\right)$.

2: **repeat**

3: Set x_{p+1} as the output of `UMP` for monotone case with accuracy $\frac{\mu\varepsilon}{2}$, prox-function $d_p(\cdot)$ and stopping criterion $\sum_{i=0}^{k-1} M_i^{-1} \geq \frac{\Omega}{\mu}$.

4: Set $R_{p+1}^2 = R_0^2 \cdot 2^{-(p+1)} + 2(1 - 2^{-(p+1)})(\frac{\varepsilon}{4} + 3\delta)$.

5: Set $d_{p+1}(x) \leftarrow R_{p+1}^2 d\left(\frac{x - x_{p+1}}{R_{p+1}}\right)$.

6: Set $p = p + 1$.

7: **until** $p > \log_2\left(\frac{2R_0^2}{\varepsilon}\right)$.

Ensure: x_p.

Applying the `Restarted UMP` method to solve the strongly convex-concave saddle point problem (1), one will get the following complexity estimate:

$$\inf_{\nu\in[0,1]}\left[\left(\frac{L_\nu}{\mu}\right)^{\frac{2}{1+\nu}}\cdot\frac{2^{\frac{2}{1+\nu}}\Omega}{\varepsilon^{\frac{1-\nu}{1+\nu}}}\cdot\log_2\frac{2R_0^2}{\varepsilon}\right].$$

For $\nu=0$ the convergence rate of the `Restarted UMP` and the accelerated method coincides, while for $\nu>0$ the asymptotic of the proposed accelerated method is better:

$$\mathcal{O}\left(\sqrt{\frac{L}{\mu_x}}\cdot\sqrt{\frac{L_{yy}}{\mu_y}}\cdot\log\frac{2L_{yy}R^2}{\varepsilon}\cdot\log\frac{2LD^2}{\varepsilon}\right),$$

where

$$L=\tilde{L}\left(\frac{\tilde{L}}{2\varepsilon}\frac{(1-\nu)(2-\nu)}{2-\nu}\right)^{\frac{(1-\nu)(1+\nu)}{2-\nu}},\tilde{L}=\left(L_{xy}\left(\frac{2L_{xy}}{\mu_y}\right)^{\frac{\nu}{2-\nu}}+L_{xx}D^{\frac{\nu-\nu^2}{2-\nu}}\right).$$

4 Content

$$g(x) = \max_{y \in Q_y} f(x, y) \to \min_{x \in Q_x} \tag{13}$$

Lemma 8

Consider the problem 1 under assumptions (6)–(9). Define $g(x) : Q_x \to \mathbb{R}$ *according to (13). Then* $g(x)$ *has the Hölder continuous gradient with Hölder constant* $\left(L_{xy} \left(\frac{2L_{xy}}{\mu_y} \right)^{\frac{\nu}{2-\nu}} + L_{xx} D^{\frac{\nu-\nu^2}{2-\nu}} \right)$ *and Hölder exponent* $\frac{\nu}{2-\nu}$.

Definition 9 ((δ, L, μ)-model)

A function $h(x) : Q_x \to \mathbb{R}$ admits (δ, L, μ)-model, if, for any $x_1, x_2 \in Q_x$:

$$\frac{\mu}{2}\|x_2 - x_1\|_2^2 + \langle \nabla h(x_1), x_2 - x_1 \rangle + h(x_1) - \delta \leq h(x_2) \leq$$

$$\leq h(x_1) + \langle \nabla h(x_1), x_2 - x_1 \rangle + \frac{L}{2}\|x_2 - x_1\|_2^2 + \delta$$

Hölder continuity and (δ, L, μ)-model

Note, that if a function $h(x)$ has the Hölder-continuous gradient with Hölder constant $\tilde{L}_{\tilde{\nu}}$ and Hölder exponent $\tilde{\nu}$, then $h(x)$ admits (δ, L, μ)-model. More precisely, inequalities in Definition (9) hold with $L = \tilde{L}_{\tilde{\nu}} \left(\frac{\tilde{L}_{\tilde{\nu}}}{2\delta} \frac{1-\tilde{\nu}}{1+\tilde{\nu}} \right)^{\frac{1-\tilde{\nu}}{1+\tilde{\nu}}}$.

(δ_0, L, μ_x)-model for the problem

According to Lemma (8), $g(x)$ has the Hölder continuous gradient, so $g(x)$ admits (δ_0, L, μ_x)-model (δ was replaced by δ_0 to simplify notation) with

$$L = \tilde{L}\left(\frac{\tilde{L}}{2\delta_0}\frac{(1-\nu)(2-\nu)}{2-\nu}\right)^{\frac{(1-\nu)(1+\nu)}{2-\nu}},$$

where $\tilde{L} = \left(L_{xy}\left(\frac{2L_{xy}}{\mu_y}\right)^{\frac{\nu}{2-\nu}} + L_{xx}D^{\frac{\nu-\nu^2}{2-\nu}}\right)$.

Theorem 10

Consider the strongly convex-concave saddle point problem (1) under assumptions (6)-(9). Define the function $g(x)$ according to (13). Then $g(x)$ admits an inexact (δ, L, μ_x)-model with $\delta = (D\Delta + \delta_0)$. Applying k steps of the Fast Gradient Method to the "outer" problem (13) and solving the "inner" problem in linear time, we obtain an ε-solution to the problem (1), where $\delta = \mathcal{O}(\varepsilon)$. The total number of iterations does not exceed

$$\mathcal{O}\left(\sqrt{\frac{L}{\mu_x}} \cdot \sqrt{\frac{L_{yy}}{\mu_y}} \cdot \log\frac{2L_{yy}R^2}{\varepsilon} \cdot \log\frac{2LD^2}{\varepsilon}\right),$$

where

$$L = \tilde{L}\left(\frac{\tilde{L}}{2\varepsilon}\frac{(1-\nu)(2-\nu)}{2-\nu}\right)^{\frac{(1-\nu)(1+\nu)}{2-\nu}}, \tilde{L} = \left(L_{xy}\left(\frac{2L_{xy}}{\mu_y}\right)^{\frac{\nu}{2-\nu}} + L_{xx}D^{\frac{\nu-\nu^2}{2-\nu}}\right).$$

5 Content

MIPT

NATIONAL RESEARCH UNIVERSITY

5 Mirror Descent for Variational Inequalities with Relatively Bounded Operator

Definition 11 (The classical boundedness)

$g(x)$ is bounded on X, if there exists $M > 0$, such that

$$\|g(x)\|_* \leq M, \quad \forall x \in X.$$

We can replace the classical concept of the boundendness of an operator by so-called Relative boundedness condition as following.

Definition 12 (The Relative boundedness)

$g(x) : X \to E^*$ is Relatively bounded, if there exists $M > 0$, such that

$$\langle g(x), y - x \rangle \leq M\sqrt{2V(y, x)}, \quad \forall x, y \in X, \tag{14}$$

5 Mirror Descent for Variational Inequalities with Relatively Bounded Operator

Definition 13 (Special case of the definition)

The Relative boundedness condition can be rewritten in the following way:

$$\|g(x)\|_* \leq \frac{M\sqrt{2V(y,x)}}{\|y-x\|}, \quad y \neq x.$$

Definition 14 (σ-monotonicity)

Let $\sigma > 0$. The operator $g(x) : X \to E^*$ is σ-monotone, if

$$\langle g(y) - g(x), y - x \rangle \geq -\sigma, \quad \forall x, y \in X. \tag{15}$$

5 Mirror Descent method for variational inequalities

Input: $\varepsilon > 0, \sigma > 0, M > 0; x_0$ and R such that $V(x_*, x_0) \leq R^2$.

1. Set $h = \frac{\varepsilon}{M^2}$.
2. Initialization $k = 0$.
3. **REPEAT**
4. $x^{k+1} = Mirr_{x^k}\Big(hg(x^k)\Big)$.
5. Set $k = k + 1$.
6. **UNTIL** $k \geq N = \dfrac{2RM^2}{\varepsilon^2}$.
7. **ENSURE** $\tilde{x} = \frac{1}{N}\sum_{k=0}^{N-1} x^k$.

5 Convergence of Algorithm

The following Theorem describes the effectiveness of the proposed Algorithm 24.

Theorem 15

Let $g : X \to E^*$ be Relatively bounded and σ–monotone operator, i.e. (14) and (15) hold. Then after no more than

$$N = \frac{2RM^2}{\varepsilon^2}$$

iterations of Algorithm 24, one can obtain an $(\varepsilon + \sigma)$–solution of the problem (2).

Thank you for your attention!