Some generalizations of non-smoothness concepts for optimization problems and variational inequalities

based on some joint works with F. Stonyakin, M. Alkousa, A. Gasnikov, O. Savchuk and D. Pasechnyuk

Alexander Titov
a.a.titov@phystech.edu

Higher School of Economics
Moscow Institute of Physics and Technology

July 2021

1 Content

1 Joint Works

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

Stonyakin, F., Titov, A., Alkousa, M., Savchuk O., Pasechnyuk D. (2021). Gradient-Type Adaptive Methods for Relatively Lipschitz Convex Optimization Problem, arXiv preprint arXiv:2107.05765. https://arxiv.org/pdf/2107.05765.pdf

Definition 1 (Saddle Point Problem)

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

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

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

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

1 ε–solutions

Definition 3 (Saddle point problem)

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

Definition 4 (Variational inequality)

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

MIPT

NATIONAL RESEARCH UNIVERSITY

Smallest covering circle problem with non-smooth functional constraints.

$$\min_{x\in Q}\left\{f(x) := \max_{1\leq k\leq N}\|x-A_k\|_2^2;\ \varphi_p(x)\leq 0,\ p=1,...,m\right\}, \tag{5}$$

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:

$$\varphi_p(x) := \sum_{i=1}^{n} \alpha_{pi}x_i + \beta_{pi},\ p=1,...,m. \tag{6}$$

Smallest covering circle problem with non-smooth functional constraints.

$$\min_{x\in Q}\left\{f(x) := \max_{1\le k\le N}\|x-A_k\|_2^2;\ \varphi_p(x)\le 0,\ p=1,...,m\right\}, \quad (5)$$

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:

$$\varphi_p(x) := \sum_{i=1}^{n} \alpha_{pi}x_i + \beta_{pi},\ p = 1, ..., m. \quad (6)$$

The corresponding Lagrange saddle point problem:

$$\min_{x\in Q}\ \max_{\overrightarrow{\lambda}=(\lambda_1,\lambda_2,...,\lambda_m)^T\in\mathbb{R}^m_+} L(x,\lambda) := f(x) + \sum_{p=1}^{m}\lambda_p\varphi_p(x) - \frac{1}{2}\sum_{p=1}^{m}\lambda_p^2.$$

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

$$G(x,\lambda) = \begin{pmatrix} \nabla f(x) + \sum\limits_{p=1}^{m}\lambda_p\nabla\varphi_p(x), \\ (-\varphi_1(x)+\lambda_1, -\varphi_2(x)+\lambda_2, \ldots, -\varphi_m(x)+\lambda_m)^T \end{pmatrix},$$

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 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

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

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.

2 Content

Theorem 5

Consider the strongly convex-concave saddle point problem (1). *Define the function* $g(x) = \max\limits_{y \in Q_y} f(x,y)$. *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" 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

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

MIPT
NATIONAL RESEARCH UNIVERSITY

3 Relative Smoothness

Definition 6

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + LV(y, x) \quad (7)$$

Motivation and many examples

Lu, H., Freund, R. M., Nesterov, Y. (2018). Relatively smooth convex optimization by first-order methods, and applications. SIAM Journal on Optimization, 28(1), 333-354.

3 Relative Lipschitz Continuity

Definition 7

$$\|\nabla f(x)\|_* \leq \frac{M\sqrt{2V(y,x)}}{\|y-x\|} \quad \forall x, y \in Q,\ y \neq x, \tag{8}$$

Motivation

Lu, H. (2019). “Relative Continuity” for Non-Lipschitz Nonsmooth Convex Optimization Using Stochastic (or Deterministic) Mirror Descent. INFORMS Journal on Optimization, 1(4), 288-303.

3 Relative Lipschitz Continuity

Support Vector Machine problem

$$f(x) := \frac{1}{n}\sum_{i=1}^{n} \max\left\{0, 1 - y_i x^T w_i\right\} + \frac{\lambda}{2}\|x\|_2^2 \to \min_x \tag{9}$$

The intersection of n ellipsoids problem

$$f(x) := \max_{0\le i\le n}\left\{\frac{1}{2}x^T A_i x + b_i^T x + c_i\right\} \to \min_x \tag{10}$$

4 (α, L, δ)-relative smoothness

Definition 8

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + LV(y,x) + L\alpha V(x,y) + \delta, \tag{11}$$

$$\alpha \left(\langle \nabla f(x), y - x \rangle + LV(y,x) + \delta \right) \geqslant 0 \quad \forall x, y \in Q. \tag{12}$$

- Relative smoothness condition $\alpha = 0$
- Relative Lipschitz continuity $\alpha = 1$

5 Adaptive Algorithm for Relatively Lipschitz Optimization Problems

Require: $\varepsilon > 0, x_0, L_0 > 0$, R s.t. $V(x_*, x_0) \leqslant R^2$.

1: Set $k = k+1, L_{k+1} = \frac{L_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q}\{\langle \nabla f(x_k), x\rangle + L_{k+1}V(x, x_k)\}. \tag{13}$$

3: **if**

$$0 \leqslant \langle \nabla f(x_k), x_{k+1} - x_k\rangle + L_{k+1}V(x_{k+1}, x_k) + \frac{\varepsilon}{2}, \tag{14}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2 \cdot L_{k+1}$ and go to item 2.

5: **end if**

Ensure: $\widehat{x} = \frac{1}{S_N}\sum_{k=0}^{N-1}\frac{x_{k+1}}{L_{k+1}}$.

5 Adaptive Algorithm for Relatively Lipschitz Optimization Problems

Theorem 9

Let $f : Q \to \mathbb{R}$ *be a convex and* M*-relatively Lipschitz continuous function, i.e.* (11) *and* (12) *take place with* $\alpha = 1, \delta \leq \frac{\varepsilon}{2}$. *Then after the stopping of the Algorithm, the following inequality holds*

$$f(\widehat{x}) - f(x_*) \leqslant \varepsilon.$$

Moreover, the total number of iterations will not exceed $N = \left\lceil \frac{4M^2R^2}{\varepsilon^2} \right\rceil$.

5 Adaptation to Inexactness for Relatively Lipschitz Continuous Minimization Problems

Require: $\varepsilon > 0, x_0, L_0 > 0, \delta_0 > 0,$ R s.t. $V(x_*, x_0) \leqslant R^2$.

1: Set $k = k+1, L_{k+1} = \frac{L_k}{2}, \delta_{k+1} = \frac{\delta_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q}\{\langle \nabla f(x_k), x\rangle + L_{k+1}V(x, x_k)\}. \tag{15}$$

3: **if**

$$0 \leqslant \langle \nabla f(x_k), x_{k+1} - x_k\rangle + L_{k+1}V(x_{k+1}, x_k) + \delta_{k+1}, \tag{16}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2 \cdot L_{k+1}, \delta_{k+1} = 2 \cdot \delta_{k+1}$ and go to item 2.

5: **end if**

Ensure: $\widehat{x} = \frac{1}{S_N}\sum_{k=0}^{N-1}\frac{x_{k+1}}{L_{k+1}}$.

5 Adaptation to Inexactness for Relatively Lipschitz Continuous Minimization Problems

Theorem 10

Let $f : Q \to \mathbb{R}$ be a convex and M-relatively Lipschitz continuous function, i.e. (11) and (12) take place with $\alpha = 1$. Then after the stopping of the Algorithm, the following inequality holds

$$f(\widehat{x}) - f(x_*) \leqslant \frac{R^2}{S_N} + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1}}{L_{k+1}}. \tag{17}$$

6 Universal Method for α-Relatively Smooth Convex Optimization Problems with Adaptation to Inexactness

Require: $\varepsilon > 0, x_0, L_0 > 0, \delta_0 > 0,\ R$ s.t. $V(x_*, x_0) \leqslant R^2$.

1: Set $k = k+1, L_{k+1} = \frac{L_k}{2}, \delta_{k+1} = \frac{\delta_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q}\{\langle \nabla f(x_k), x\rangle + L_{k+1}V(x, x_k)\}. \tag{18}$$

3: **If**

$$f(x_{k+1}) \leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k\rangle + L_{k+1}V(x_{k+1}, x_k) + \delta_{k+1}, \tag{19}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2 \cdot L_{k+1}, \delta_{k+1} = 2 \cdot \delta_{k+1}$ and go to item 2.

5: **end if**

Ensure: $\widehat{x} = \frac{1}{S_N}\sum_{k=0}^{N-1}\frac{x_{k+1}}{L_{k+1}}$.

Theorem 11

Let $f : Q \to \mathbb{R}$ be a convex and α-relatively smooth function, i.e. (11), (12) hold. Then after N iterations of the Algorithm, the following inequality holds

$$f(\widehat{x}) - f(x_*) \leqslant \frac{R^2}{S_N} + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1}}{L_{k+1}}, \tag{20}$$

where $S_N = \sum\limits_{k=0}^{N-1} \frac{1}{L_{k+1}}$. Note that the auxiliary problem (18) in Algorithm is solved no more than $3N$ times.

6 Universal Method for α-Relatively Smooth Convex Optimization Problems

Require: $\varepsilon > 0, x_0, L_0 > 0$, R s.t. $V(x_*, x_0) \leqslant R^2$.

1: Set $k = k+1, L_{k+1} = \frac{L_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q} \{\langle \nabla f(x_k), x \rangle + L_{k+1} V(x, x_k)\}. \tag{21}$$

3: **If**

$$f(x_{k+1}) \leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + L_{k+1} V(x_{k+1}, x_k) + \frac{3\varepsilon}{4}, \tag{22}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2 \cdot L_{k+1}$ and go to item 2.

5: **end if**

6: Stopping criterion

$$S_N = \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \geqslant \frac{4R^2}{\varepsilon}. \tag{23}$$

Ensure: $\widehat{x} = \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{x_{k+1}}{L_{k+1}}$.

Theorem 12

Let $f : Q \to \mathbb{R}$ be a convex and α-relatively smooth function, i.e. (11) and (12) hold with $\delta \leq \frac{3\varepsilon}{4}$. Then after the stopping of the Algorithm, the following inequality holds

$$f(\widehat{x}) - f(x_*) \leqslant \varepsilon.$$

If f is M-relatively Lipschitz continuous, i.e. (11) and (12) take place with $\alpha = 1$, the number of iterations of Algorithm does not exceed

$$N = \left\lceil \frac{16M^2R^2}{\varepsilon^2} \right\rceil$$

.

7 Content

7 Mirror Descent for Variational Inequalities with Relatively Bounded Operator

Definition 13 (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 boundedness of an operator by the so-called Relative boundedness condition as following.

Definition 14 (The Relative boundedness)

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

$$\langle g(x), y - x \rangle \leq M\sqrt{2V(y,x)}, \quad \forall x, y \in X, \tag{24}$$

7 Mirror Descent for Variational Inequalities with Relatively Bounded Operator

Definition 15 (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 16 (σ-monotonicity)

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

$$\langle g(y) - g(x), y - x \rangle \geq -\sigma, \quad \forall x, y \in X. \tag{25}$$

7 Adaptive Algorithm for VI's

Require: $\varepsilon > 0, x_0, L_0 > 0,\ R > 0$ s.t. $\max\limits_{x \in Q} V(x, x_0) \leqslant R^2$.

1: Set $k = k + 1, L_{k+1} = \frac{L_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q} \{\langle g(x_k), x\rangle + L_{k+1} V(x, x_k)\}. \tag{26}$$

3: **if**

$$\frac{\varepsilon}{2} + \langle g(x_k), x_{k+1} - x_k\rangle + L_{k+1} V(x_{k+1}, x_k) \geqslant 0, \tag{27}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2L_{k+1}$, and go to item 2.

5: **end if**

6: Stopping criterion

$$S_N = \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \geqslant \frac{2R^2}{\varepsilon}. \tag{28}$$

Ensure: $\widehat{x} = \frac{1}{S_N} \sum\limits_{k=0}^{N-1} \frac{x_k}{L_{k+1}}$.

7 Adaptive Algorithm for VI's.

Theorem 17

Let $g : Q \to \mathbb{R}$ be a relatively bounded and monotone operator, i.e. (24) and (25) hold. Then after the stopping of the Algorithm, the following inequality holds

$$\max_{x \in Q} \langle g(x), \widehat{x} - x \rangle \leqslant \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \langle g(x), x_k - x \rangle \leqslant \varepsilon.$$

Moreover, the total number of iterations will not exceed $N = \left\lceil \frac{4M^2R^2}{\varepsilon^2} \right\rceil$.

7 Adaptation to Inexactness for Relatively Bounded VI's

Require: $\varepsilon > 0, x_0, L_0 > 0, \delta_0 > 0,$ R s.t. $\max V(x, x_0) \leqslant R^2$.

1: Set $k = k+1, L_{k+1} = \frac{L_k}{2}, \delta_{k+1} = \frac{\delta_k}{2}$.

2: Find

$$x_{k+1} = \arg\min_{x \in Q} \{\langle g(x_k), x \rangle + L_{k+1} V(x, x_k)\}. \tag{29}$$

3: **if**

$$0 \leqslant \langle g(x_k), x_{k+1} - x_k \rangle + L_{k+1} V(x_{k+1}, x_k) + \delta_{k+1}, \tag{30}$$

then go to the next iteration (item 1).

4: **else**

set $L_{k+1} = 2 \cdot L_{k+1}, \delta_{k+1} = 2 \cdot \delta_{k+1}$ and go to item 2.

5: **end if**

Ensure: $\widehat{x} = \frac{1}{S_N} \sum\limits_{k=0}^{N-1} \frac{x_k}{L_{k+1}}$.

Theorem 18

Let $g : Q \to \mathbb{R}$ be a relatively bounded and monotone operator, i.e. (24) and (25) hold. Then after N steps of the Algorithm the following inequality holds

$$\max_{x\in Q}\langle g(x), \widehat{x} - x\rangle \leqslant \frac{R^2}{S_N} + \frac{1}{S_N}\sum_{k=0}^{N-1}\frac{\delta_{k+1}}{L_{k+1}}. \tag{31}$$

Note that the auxiliary problem (29) is solved no more than $3N$ times.

Remark

The condition of the relative boundedness is essential only for justifying (30). For $L_{k+1} \geq L = \frac{M^2}{\varepsilon}$ and $\delta_{k+1} \geq \frac{\varepsilon}{2}$, (30) certainly holds. So, for $C = \max\{\frac{2L}{L_0}; \frac{2\delta}{\delta_0}\}$, $L_{k+1} \leq CL$ and $\delta_{k+1} \leq C\delta = \frac{C\varepsilon}{2}$ $\forall k \geq 0$. Thus, $\max\limits_{x\in Q}\langle g(x), \widehat{x} - x\rangle \leq \varepsilon$ after $N = O(\varepsilon^{-2})$ iterations of the Algorithm.

Thank you for your attention!