**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2020 Т. 12 № 2 С. 301–317**

Ки&М

DOI: 10.20537/2076-7633-2020-12-2-301-317

ЧИСЛЕННЫЕ МЕТОДЫ И ОСНОВЫ ИХ РЕАЛИЗАЦИИ

УДК: 519.85

Метод зеркального спуска для условных задач оптимизации с большими значениями норм субградиентов функциональных ограничений

Ф. С. Стонякин[1,2,a], А. Н. Степанов[1,b], А. В. Гасников[2,3,4,c], А. А. Титов[2,d]

[1]Крымский федеральный университет им. В. И. Вернадского,
Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, д. 4
[2]Московский физико-технический институт,
Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., д. 9
[3]Институт проблем передачи информации им. А. А. Харкевича Российской академии наук,
Россия, 127051, г. Москва, Большой Каретный пер., д. 19, стр. 1
[4]Кавказский математический центр, Адыгейский государственный университет,
Россия, 385000, г. Майкоп, ул. Первомайская, 208

E-mail: [a] fedyor@mail.ru, [b] stepanov.student@gmail.com, [c] gasnikov@yandex.ru, [d] a.a.titov@phystech.edu

Получено 04.11.2019, после доработки — 08.12.2019.
Принято к публикации 24.12.2019.

В работе рассмотрена задача минимизации выпуклого и, вообще говоря, негладкого функционала f при наличии липшицевого неположительного выпуклого негладкого функционального ограничения g. При этом обоснованы оценки скорости сходимости методов адаптивного зеркального спуска также и для случая квазивыпуклого целевого функционала в случае выпуклого функционального ограничения. Предложен также метод и для задачи минимизации квазивыпуклого целевого функционала с квазивыпуклым неположительным функционалом ограничения. В работе предложен специальный подход к выбору шагов и количества итераций в алгоритме зеркального спуска для рассматриваемого класса задач. В случае когда значения норм (суб)градиентов функциональных ограничений достаточно велики, предложенный подход к выбору шагов и остановке метода может ускорить работу метода по сравнению с его аналогами. В работе приведены численные эксперименты, демонстрирующие преимущества использования таких методов. Также показано, что методы применимы к целевым функционалам различных уровней гладкости. В частности, рассмотрен класс гёльдеровых целевых функционалов. На базе техники рестартов для рассмотренного варианта метода зеркального спуска был предложен оптимальный метод решения задач оптимизации с сильно выпуклыми целевыми функционалами. Получены оценки скорости сходимости рассмотренных алгоритмов для выделенных классов оптимизационных задач. Доказанные оценки демонстрируют оптимальность рассматриваемых методов с точки зрения теории нижних оракульных оценок.

Ключевые слова: негладкая условная оптимизация, квазивыпуклый функционал, адаптивный зеркальный спуск, уровень гладкости, гёльдеров целевой функционал, оптимальный метод

Исследования Ф. С. Стонякина по разработке алгоритмов 2 и 4, замечаний 2 и 3, обоснованию теорем 2 и 5 выполнены при поддержке Российского научного фонда (проект 18-71-00048). Исследования Ф. С. Стонякина и А. Н. Степанова по вычислительным экспериментам (примеры 1 и 2) выполнены при поддержке Российского научного фонда (проект 18-71-00048). Исследования по разработке алгоритма 3, обоснованию теорем 3 и 4, а также по вычислительным экспериментам (примеры 3 и 4) выполнены при поддержке Российского фонда фундаментальных исследований (проект 18-31-00219 мол-а). Исследования Ф. С. Стонякина по разработке замечания 4 выполнены при поддержке гранта Президента Российской Федерации для государственной поддержки молодых российских ученых-кандидатов наук (проект МК-15.2020.1). Исследование А. В. Гасникова по разработке алгоритма 3 и обоснованию леммы 3 выполнено при поддержке Министерства науки и высшего образования Российской Федерации (Госзадание МФТИ, проект 075-00337-20-03).

COMPUTER RESEARCH AND MODELING
2020 VOL. 12 NO. 2 P. 301–317
DOI: 10.20537/2076-7633-2020-12-2-301-317

Ки&М

NUMERICAL METHODS AND THE BASIS FOR THEIR APPLICATION

UDC: 519.85

Mirror descent for constrained optimization problems with large subgradient values of functional constraints

F. S. Stonyakin[1,2,a], A. N. Stepanov[1,b], A. V. Gasnikov[2,3,4,c], A. A. Titov[2,d]

[1]V. I. Vernadsky Crimean Federal University,
4 Academician Vernadsky av., Simferopol, 295007, Russia
[2]Moscow Institute of Physics and Technology,
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russia
[3]Institute for Information Transmission Problems,
19/1 Bolshoi Karetny per., Moscow, 127051, Russia
[4]Caucasus Mathematical Center, Adyghe State University,
208 Pervomayskaya st., Maykop, 385000, Russia

E-mail: [a] fedyor@mail.ru, [b] stepanov.student@gmail.com, [c] gasnikov@yandex.ru, [d] a.a.titov@phystech.edu

Received 04.11.2019, after completion — 08.12.2019.
Accepted for publication 24.12.2019.

The paper is devoted to the problem of minimization of the non-smooth functional f with a non-positive non-smooth Lipschitz-continuous functional constraint. We consider the formulation of the problem in the case of quasi-convex functionals. We propose new strategies of step-sizes and adaptive stopping rules in Mirror Descent for the considered class of problems. It is shown that the methods are applicable to the objective functionals of various levels of smoothness. Applying a special restart technique to the considered version of Mirror Descent there was proposed an optimal method for optimization problems with strongly convex objective functionals. Estimates of the rate of convergence for the considered methods are obtained depending on the level of smoothness of the objective functional. These estimates indicate the optimality of the considered methods from the point of view of the theory of lower oracle bounds. In particular, the optimality of our approach for Hölder-continuous quasi-convex (sub)differentiable objective functionals is proved. In addition, the case of a quasi-convex objective functional and functional constraint was considered. In this paper, we consider the problem of minimizing a non-smooth functional f in the presence of a Lipschitz-continuous non-positive non-smooth functional constraint g, and the problem statement in the cases of quasi-convex and strongly (quasi-)convex functionals is considered separately. The paper presents numerical experiments demonstrating the advantages of using the considered methods.

Keywords: non-smooth constrained optimization, quasi-convex functional, adaptive mirror descent, level of smoothness, Hölder-continuous objective functional, optimal method

Citation: *Computer Research and Modeling*, 2020, vol. 12, no. 2, pp. 301–317 (Russian).

The research of F. Stonyakin in Algorithms 2 and 4, Remarks 2 and 3, Theorems 2 and 5 was supported by Russian Science Foundation (project 18-71-00048). The research of F. Stonyakin and A. Stepanov in numerical experiments (Examples 1 and 2) was supported by Russian Science Foundation (project 18-71-00048). The research in Theorems 3 and 4, Algorithm 3 and in numerical experiments (Examples 3 and 4) was supported by Russian Foundation for Basic Research (project 18-31-00219 mol-a). The research of F. Stonyakin in Remark 4 was supported by the grant of the President of Russian Federation for young candidates of sciences (project MK-15.2020.1). The research of A. V. Gasnikov (Algorithm 3 and Lemma 3) is supported by the Ministry of Science and Higher Education of the Russian Federation (Goszadaniye in MIPT, project 075-00337-20-03).

Introduction

Non-smooth convex constrained optimization problems play an important role in modern large-scale optimization and its applications [Ben-Tal, 1997; Nesterov, 2015; Shpirko, 2014]. There are a lot of methods to solve such problems, among which one can mention the Mirror Descent Method [Beck, 2003; Nemirovsky, 1983].

Recently, in [Bayandina, 2018a] algorithms for Mirror Descent with both adaptive step selection and an adaptive stopping criterion were proposed. In addition, an optimal method was proposed for the special class of convex constrained optimization problems, when the gradient of the objective functional satisfies the Lipschitz property. For example, quadratic functionals do not satisfy the Lipschitz condition, but their gradient does. An adaptive Mirror Descent algorithm, based on the ideology of [Nesterov, 2004], was proposed to solve such problems in [Bayandina, 2018a, Section 3.3].

In this paper we develop the abovementioned research and consider some modifications of the algorithmic scheme [Bayandina, 2018a, Section 3.3]. More precisely, in proposed Algorithm 2 we consider a new approach to choosing a step in the method, as well as appropriate options for stopping criteria, which differ from [Bayandina, 2018a]. It is important that we choose the non-productive step ($\nabla g(x^k)$ is the subgradient g at the current point x^k) in the form $h_k = \dfrac{\varepsilon}{\|\nabla g(x^k)\|}$ instead of $h_k = \dfrac{\varepsilon}{\|\nabla g(x^k)\|^2}$ in [Bayandina, 2018a]. This circumstance, as well as the appropriate choice of the number of iterations (10), leads us to the fact that the method can run faster than the previous analogue [Bayandina, 2018a, Section 3.3] in the case when the values of the subgradients are large. Note that a method similar to Algorithm 2 was proposed in [Nemirovsky, 1983] for the case of convex Lipschitz-continuous functionals.

This paper substantiates the convergence rate estimates for the proposed version of the Mirror Descent method, proves its optimality from the point of view of the theory of lower bounds for objective functionals of various smoothness levels: which have a Lipschitz-continuous gradient or satisfy the Lipschitz (Hölder) condition. It is also shown that the obtained estimates of the convergence rate are preserved for a quasi-convex [Nesterov, 1984; Nesterov, 1989; Konnov, 2003] objective functional and constraint (see e.g. [Gasnikov, 2018, Exercise 2.7]). Using the restart technique, the optimal method for strongly (quasi-)convex objective functionals is considered. The paper ends with some numerical experiments for geometric problems with functional constraints, which illustrate that the proposed method can work faster compared to [Bayandina, 2018a, Section 3.3]. There are also given some examples of more efficient methods in the case of a large dimension.

The contribution of this paper is as follows:

– Two analogues of the Mirror Descent method [Bayandina, 2018, Section 3.3] are proposed for convex programming problems with another strategy for choosing a non-productive step. One of them (Algorithm 2) solves the optimization problem under the assumption of quasi-convexity of the objective functional and convexity of the constraint. The second method (Algorithm 3) is applicable in the case when both objective and constraint are quasi-convex, but assumes knowledge of the Lipschitz constant of the constraint M_g. Estimates of the rate of its convergence and optimality are obtained in terms of lower bounds for convex objective functionals of various smoothness levels.

– It is shown that the obtained convergence rate estimates are valid for the case of the minimization problems with quasi-convex objective functionals of different smoothness levels.

– It is shown that for the Hölder-continuous quasi-convex differentiable (or subdifferentiable by Clarke) objective functionals the convergence rate is equal to $O\left(\dfrac{1}{\varepsilon^2}\right)$.

– Using the restart technique, an optimal method was proposed for the class of minimization problems with strongly (quasi-)convex Hölder-continuous objective functionals with the complexity estimate equal to $O\left(\frac{1}{\varepsilon}\right)$.

– Numerical experiments for geometrical problems (some analogues of the Fermat–Torricelli–Steiner problem and the problem of the smallest covering ball) with convex constraints are presented. When (sub)gradient values of the functional constraints are large, the proposed method can work faster [Bayandina, 2018]. Some tests for high-dimensional problems are also considered.

– Numerical experiments for the minimization of quasi-convex functionals are given. A variant of the smallest covering ball problem with a quasi-convex objective functional is considered.

Problem Statement and Standard Mirror Descent Basics

Let $(E, \|\cdot\|)$ be a normed finite-dimensional vector space and E^* be its conjugate space with the norm:

$$\|y\|_* = \max_x \{\langle y, x\rangle, \|x\| \leqslant 1\},$$

where $\langle y, x\rangle$ is the value of the continuous linear functional y at $x \in E$.

Let $Q \subset E$ be a (simple) closed convex set. Consider the following problem:

$$f(x) \to \min_{x\in Q}, \tag{1}$$

s.t.

$$g(x) \leqslant 0. \tag{2}$$

Assume that convex functional g satisfies the Lipschitz condition with a constant M_g:

$$|g(x) - g(y)| \leqslant M_g\|x - y\| \ \forall x, y \in Q. \tag{3}$$

We consider the cases of convex and quasi-convex objective functional f. Let $d\colon Q \to \mathbb{R}$ be a distance generating function (d.g.f) which is continuously differentiable and 1-strongly convex w.r.t. the norm $\|\cdot\|$, i.e.

$$\forall x, y, \in Q \ \ \langle \nabla d(x) - \nabla d(y), x - y\rangle \geqslant \|x - y\|^2,$$

and assume that there is a constant Θ_0, such that $d(x_*) \leqslant \Theta_0^2$, where x_* is a solution of the problem (1)–(2) (we suppose that the considered problem is solvable).

For all $x, y \in Q \subset E$ consider the corresponding Bregman divergence

$$V(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle.$$

The proximal mapping operator is defined as follows:

$$\mathrm{Mirr}_x(p) = \arg\min_{u\in Q} \{\langle p, u\rangle + V(x, u)\} \ \text{ for each } \ x \in Q \ \text{ and } \ p \in E^*.$$

We assume for simplicity that $\mathrm{Mirr}_x(p)$ is easily computable.

Mirror Descent Algorithms: some step-sizes strategies

Two Mirror Descent methods for optimization problems with one convex subdifferentiable functional constraint were proposed in [Bayandina, 2018a]. The convergence of the first of them is obtained for the case of the Lipschitz-continuous objective functional (see [Bayandina, 2018a, Section 3.1]), while the convergence of the second is justified under the assumption that the gradient ∇f

satisfies the Lipschitz property (see [Bayandina, 2018a, Section 3.3]):

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\| \; \forall x, y \in Q. \tag{4}$$

Let us remind the method ([Bayandina, 2018a, Section 3.3]; see Algorithm 1).

Algorithm 1. Adaptive Mirror Descent

Require: $\varepsilon > 0, \Theta_0 : \; d(x_*) \leqslant \Theta_0^2$

$x^0 = argmin_{x \in Q} \, d(x)$

$I =: \varnothing$

$N \leftarrow 0$

repeat

  if $g(x^N) \leqslant \varepsilon$ **then**

    $M_N = \|\nabla f(x^N)\|_*, \; h_N = \dfrac{\varepsilon}{M_N}$

    $x^{N+1} = Mirr_{x^N}(h_N \nabla f(x^N))$ // *"productive steps"*

    $N \to I$

  else

    $M_N = \|\nabla g(x^N)\|_*, \; h_N = \dfrac{\varepsilon}{M_N^2}$

    $x^{N+1} = Mirr_{x^N}(h_N \nabla g(x^N))$ // *"non-productive steps"*

  end if

  $N \leftarrow N + 1$

until $2\dfrac{\Theta_0^2}{\varepsilon^2} \leqslant \sum\limits_{j \notin I} \dfrac{1}{M_j^2} + |I|$

Ensure: $\bar{x}^N := argmin_{x^k, \, k \in I} \, f(x^k)$

Lemma 1. *Let us define the following function:*

$$\omega(\tau) = \max_{x \in Q}\{f(x) - f(x_*) : \|x - x_*\| \leqslant \tau\}, \tag{5}$$

where τ is a positive number. Then for any $y \in Q$

$$f(y) - f(x_*) \leqslant \omega(v_f(y, x_*)), \tag{6}$$

where

$$v_f(y, x_*) = \left\langle \frac{\nabla f(y)}{\|\nabla f(y)\|}, y - x_* \right\rangle \; \text{for } \nabla f(y) \neq 0 \tag{7}$$

and $v_f(y, x_*) = 0$ for $\nabla f(y) = 0$.

For Algorithm 1 the following theorem holds.

Theorem 1. *Let $\varepsilon > 0$ be a fixed number and the stopping criterion of Algorithm 1 is satisfied. Then*

$$\min_{k \in I} v_f(x^k, x_*) < \varepsilon, \quad \max_{k \in I} g(x^k) \leqslant \varepsilon. \tag{8}$$

In addition, Algorithm 1 works no more than

$$N = \left\lceil \frac{2\max\{1, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil \tag{9}$$

iterations.

Now we will estimate the rate of convergence of the proposed method. For this we need the following auxiliary assumption [Nesterov, 2004, Lemma 3.2.1].

Basing on Lemma 1 and Theorem 1, we can estimate the rate of convergence of Algorithm 1 for a differentiable objective functional f with the Lipschitz-continuous gradient. Using the well-known inequality for an exact solution x_* we can get that (see, for example, [Nesterov, 2004])

$$f(x) \leqslant f(x_*) + \|\nabla f(x_*)\|_* \|x - x_*\| + \frac{1}{2}L\|x - x_*\|^2,$$

$$\min_{k \in I} f(x^k) - f(x_*) \leqslant \min_{k \in I} \left\{ \|\nabla f(x_*)\|_* \|x^k - x_*\| + \frac{1}{2}L\|x^k - x_*\|^2 \right\}.$$

Further, the following estimate is valid:

$$f(x) - f(x_*) \leqslant \varepsilon \|\nabla f(x_*)\|_* + \frac{1}{2}L\varepsilon^2.$$

Corollary 1. *Let f be differentiable on Q and* (4) *hold. Then, after the stopping of Algorithm 1, the next inequality holds:*

$$\min_{1 \leqslant k \leqslant N} f(x^k) - f(x_*) \leqslant \varepsilon \cdot \|\nabla f(x_*)\|_* + \frac{L\varepsilon^2}{2}.$$

Let us observe a new version of the adaptive Mirror Descent method with another step selection strategy. A resembling idea was researched in [Juditsky, 2010], but only for the case of Lipschitz-continuous functional. Note, that the following modification can be used to minimize functionals with different levels of smoothness. As earlier, we will consider the method for a fixed accuracy $\varepsilon > 0$, an initial approximation x^0, and some value Θ_0, such that $V(x^0, x_*) \leqslant \Theta_0^2$.

Algorithm 2. Adaptive Mirror Descent

Require: $\varepsilon > 0, \Theta_0 : d(x_*) \leqslant \Theta_0^2$
$x^0 = argmin_{x \in Q}\, d(x)$
$I =: \varnothing$
$N \leftarrow 0$
repeat
  if $g(x^N) \leqslant \varepsilon \|\nabla g(x^N)\|_*$ **then**
    $M_N = \|\nabla f(x^N)\|_*,\ h_N = \dfrac{\varepsilon}{M_N}$
    $x^{N+1} = Mirr_{x^N}(h_N \nabla f(x^N))$ // *"productive steps"*
    $N \to I$
  else
    $M_N = \|\nabla g(x^N)\|_*,\ h_N = \dfrac{\varepsilon}{M_N}$
    $x^{N+1} = Mirr_{x^N}(h_N \nabla g(x^N))$ // *"non-productive steps"*
  end if
  $N \leftarrow N + 1$
until $2\dfrac{\Theta_0^2}{\varepsilon^2} \leqslant N$
Ensure: $\bar{x}^N := argmin_{x^k,\, k \in I}\, f(x^k)$

The following theorem holds.

Theorem 2. *Let $\varepsilon > 0$ be a fixed number and Algorithm 2 work during a fixed number of steps*

$$N = \left\lceil \frac{2\Theta_0^2}{\varepsilon^2} \right\rceil. \tag{10}$$

Then we have

$$\min_{k \in I} v_f(x^k, x_*) \leqslant \varepsilon, \qquad \max_{k \in I} g(x^k) \leqslant \varepsilon M_g. \tag{11}$$

Proof. 1. If $k \in I$,

$$
\begin{aligned}
h_k\left\langle \nabla f(x^k), x^k - x_* \right\rangle = \varepsilon v_f(x^k, x_*) \leqslant \\
\leqslant \frac{h_k^2}{2}\|\nabla f(x^k)\|_*^2 + V(x^k, x_*) - V(x^{k+1}, x_*) = \\
= \frac{\varepsilon^2}{2} + V(x^k, x_*) - V(x^{k+1}, x_*).
\end{aligned}
\tag{12}
$$

2. If $k \in J := N \setminus I$, then $\dfrac{g(x^k)}{\|\nabla g(x^k)\|_*} > \varepsilon$ and $\dfrac{g(x^k) - g(x_*)}{\|\nabla g(x^k)\|_*} \geqslant \dfrac{g(x^k)}{\|\nabla g(x^k)\|_*} > \varepsilon$. Therefore, the following inequalities hold:

$$
\begin{aligned}
\varepsilon^2 < h_k\left(g(x^k) - g(x_*)\right) \leqslant \frac{h_k^2}{2}\|\nabla g(x^k)\|_*^2 + \\
+V(x^k, x_*) - V(x^{k+1}, x_*) = \frac{\varepsilon^2}{2} + V(x^k, x_*) - V(x^{k+1}, x_*), \text{ or} \\
\frac{\varepsilon^2}{2} < V(x^k, x_*) - V(x^{k+1}, x_*).
\end{aligned}
\tag{13}
$$

3. After summing up the inequalities (12) and (13) one can get

$$
\sum_{k\in I} \varepsilon v_f(x^k, x_*) \leqslant \frac{\varepsilon^2|I|}{2} - \frac{\varepsilon^2|J|}{2} + V(x^0, x_*) - V(x^{k+1}, x_*) = \varepsilon^2|I| - \frac{\varepsilon^2 N}{2} + \Theta_0^2.
$$

After (10) iterations of Algorithm 2 the next inequality holds:

$$
\min_{k\in I} v_f(x^k, x_*) \leqslant \varepsilon.
$$

Clear, for each $k \in I \quad g(x^k) \leqslant \varepsilon\|\nabla g(x^k)\|_* \leqslant \varepsilon M_g$.

Now we have to show that the set of productive steps I is non-empty. If $I = \varnothing$, then $|J| = N$ and (3) means that $N \geqslant \dfrac{2\Theta_0^2}{\varepsilon^2}$. On the other hand, from (13) we have

$$
\frac{\varepsilon^2 N}{2} < V(x^0, x_*) \leqslant \Theta_0^2,
$$

which leads us to the controversy, so $I \neq \varnothing$. □

Let us show how to estimate the quality of the solution by the function basing on the previous theorem. Note, that it is possible to take into account different levels of smoothness of the objective functional.

Corollary 2. *Let f satisfy the Lipschitz condition*

$$
|f(x) - f(y)| \leqslant M_f\|x - y\| \quad \forall x, y \in Q. \tag{14}
$$

Then, after the stopping of Algorithm 2, the following inequality holds:

$$
\min_{k\in I} f(x^k) - f(x_*) \leqslant M_f \varepsilon.
$$

The case of quasi-convex functionals

Let us consider the optimization problem (1) under the assumption of quasi-convexity of the objective functional f. Recall (see [Hazan, 2015]) that function $f\colon Q \to \mathbb{R}$ is called quasi-convex, if

$$f((1-\alpha)x + \alpha y) \leqslant \max\{f(x), f(y)\} \quad \forall \alpha \in [0;1] \quad \forall x, y \in Q.$$

As earlier, let g satisfy the Lipschitz condition (3) with the constant M_g.

Let us remind the definition of Clarke subdifferential [Clarke, 1983]. Let $x_0 \in \mathbb{R}^n$ be a fixed point and $h \in \mathbb{R}^n$ be a fixed direction. Denote

$$f^{\uparrow}_{Cl}(x_0; h) = \lim_{x' \to x_0} \sup_{\alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha h) - f(x')).$$

Value $f^{\uparrow}_{Cl}(x_0; h)$ is called the Clarke subdifferential of functional f at the point x_0 in the direction h. This function is subadditive and positively homogeneous, thus we can define the subdifferential of the function f at the point x_0 as follows:

$$\partial_{Cl} f(x_0) := \left\{ v \in \mathbb{R} \mid f^{\uparrow}_{Cl}(x_0; g) \geqslant vg \ \forall g \in \mathbb{R} \right\}.$$

According to this,

$$f^{\uparrow}_{Cl}(x_0; h) = \max_{v \in \partial_{Cl} f(x_0)} \langle v, h \rangle.$$

Note, that from now on we will understand any element (vector) of the Clarke subdifferential as the subgradient of the quasi-convex (locally Lipschitz) functional f. For convex functional g, we understand the concept of a subgradient in the standard way.

Lemma 2. *Let $f\colon X \to \mathbb{R}$. For any $y \in Q$, vector $p_y \in E^*$ and $h > 0$ define $z = Mirr_y(h \cdot p_y)$. Then for any $x \in Q$ the next inequality holds:*

$$h\langle p_y, y - x\rangle \leqslant \frac{h^2}{2}\|p_y\|_*^2 + V(y, x) - V(z, x).$$

Note, that for convex subdifferentiable functional f and subgradient $p_y = \nabla f(y)$ this inequality is modified as follows:

$$h(f(y) - f(x)) \leqslant \langle \nabla f(y), y - x\rangle \leqslant \frac{h^2}{2}\|\nabla f(y)\|_*^2 + V(y, x) - V(z, x).$$

Note that for quasi-convex objective functional f and constraint g instead of (sub)gradient $\nabla f(y)$ in $v_f(y, x_*)$ (see (7)) we can consider some element of the following set

$$\widehat{D}f(x) = \{p \mid \langle p, x - y\rangle \geqslant 0 \quad \forall y \in X : \ f(y) < f(x)\}.$$

Generally, this set is a non-empty, closed and convex cone. Following [Nesterov, 1989], we assume that $\widehat{D}f(x) \neq \{0\}$ for $x \neq x_*$. Hereinafter denote $Df(x)$ as one arbitrary vector from $\widehat{D}f(x)$:

$$Df(x) \in \widehat{D}f(x).$$

However, the (sub)gradient or the Clarke subdifferential of f or g can be used, if they are finite and nonzero (we assume that $\nabla f(y) \neq 0$ for $y \neq x_*$).

Theorem 3. *Let f be a quasi-convex functional and g be a convex functional. Then for Algorithm 2 after $N = \left\lceil \frac{2\Theta_0^2}{\varepsilon^2} \right\rceil$ steps the following inequalities hold:*

$$\min_{k \in I} v_f(x^k, x_*) \leqslant \varepsilon, \qquad \max_{k \in I} g(x^k) \leqslant \varepsilon M_g.$$

Now let us consider the case when both objective f and functional constraint g are quasi-convex.

Lemma 3. *Lemma 1 is valid for $v_g(y, x_*)$ in the case of a quasi-convex objective functional and functional constraint.*

Proof. Let us note that for any non-productive point x the following inequality holds:

$$g(x) \geq g(x_*).$$

Set $\Omega_x = \{y \in Q \,|\, g(y) = x\}$. For some $\lambda \geq 0$ we define $y_* = x_* + \lambda Dg(x)$. As $\langle y_* - x, Dg(x) \rangle = 0$, we have

$$\langle x_* + \lambda Dg(x) - x, Dg(x) \rangle = 0 \text{ and } \langle x_* - x, Dg(x) \rangle = -\lambda \|Dg(x)\|^2.$$

It means that

$$\lambda \|Dg(x)\| = \left\langle \frac{Dg(x)}{\|Dg(x)\|}, x - x_* \right\rangle, \quad \|y_* - x_*\| = v_g(x, x_*),$$

and $g(x) - g(x_*) \leq g(y_*) - g(x_*) \leq M_g \|y_* - x_*\| = M_g v_g(x, x_*)$. □

Consider the following modification of Algorithm 2 ($N = 0, 1, 2, \ldots$).

Algorithm 3. Adaptive Mirror Descent, the Case of Quasi-Convex Functional Constraint

1: **if** $g(x^N) \leqslant \varepsilon \cdot M_g$ **then**
2: $\quad x^{N+1} = Mirr_{x^N}(h_N^f Df(x^N))$ // *"productive steps"*
3: **else**
4: $\quad x^{N+1} = Mirr_{x^N}(h_N^g Dg(x^N))$ // *"productive steps"*
5: **end if**

Let us choose the step-sizes as follows:

$$h_k^f = \frac{C_f}{\|Df(x^k)\|_*}, \qquad h_k^g = \frac{C_g}{\|Dg(x^k)\|_*}.$$

Denote N_I and N_J as the number of productive and non-productive steps during the work of Algorithm 3 respectively.

Similarly to [Bayandina, 2018a] (see also the proof of Theorem 2), the next inequality holds:

$$C_f N_I \min_{k \in I} v_f(x^k, x_*) \leqslant \frac{1}{2} \sum_{k \in I} (h_k^f)^2 \|Df(x^k)\|_2^2 - C_g \sum_{k \in J} v_g(x^k, x_*) + \frac{1}{2} \sum_{k \in J} (h_k^g)^2 \|Dg(x^k)\|_2^2 + \Theta_0^2.$$

Let $C_g = C_f = \varepsilon$, $N \geqslant \frac{2\Theta_0^2}{\varepsilon^2}$. As $g(x^k) \geqslant M_g \varepsilon$, $k \in J$, and using Lemma 3 for constraint $g(x)$ with Lipschitz constant M_g we get

$$-v_g(x^k, x_*) \leqslant \frac{g(x_*) - g(x^k)}{M_g} \leqslant -\frac{g(x^k)}{M_g} \leqslant -\varepsilon.$$

Remark 1. Algorithm 3 unlike Algorithm 2 solves the problem under the assumption of quasi-convexity of the functional constraint. Note, that we have to know the Lipschitz constant M_g which appears in the transition criteria of Algorithm 2.

Theorem 4. *Let f be quasi-convex, g be quasi-convex with Lipschitz constant M_g. Then for Algorithm 3 after* $N = \left\lceil \frac{2\Theta_0^2}{\varepsilon^2} \right\rceil$ *steps the following inequalities hold:*

$$\min_{k \in I} v_f(x^k, x_*) \leqslant \varepsilon, \qquad \max_{k \in I} g(x^k) \leqslant \varepsilon M_g.$$

Now we will show the optimality of the proposed algorithmic procedures for the case of Hölder-continuous objective functionals.

REMARK 2. Let f satisfy the Hölder condition ($\nu \in [0; 1)$)

$$|f(x) - f(y)| \leqslant M_{f,\nu} \|x - y\|^{\nu} \quad \forall x, y \in Q. \tag{15}$$

For example, $f(x) = \sqrt{x}$ and $f(x) = \sqrt[4]{x}$.

Let us recall the following inequality ([Gasnikov, 2018, Remark 5.1])

$$M_\nu a^\nu \leqslant M_\nu \left(\frac{M_\nu}{\delta}\right)^{\frac{1-\nu}{1+\nu}} \frac{a^2}{2} + \delta, \tag{16}$$

which is true for each $\delta > 0$. Then by (15) we have

$$|f(x) - f(y)| \leqslant \frac{M_\nu^{\frac{2}{1+\nu}}}{2\delta^{\frac{1-\nu}{1+\nu}}} \|x - y\|^2 + \delta.$$

Set $\delta = \varepsilon$. Then

$$|f(x) - f(y)| \leqslant \underbrace{\frac{M_\nu^{\frac{2}{1+\nu}}}{2\varepsilon^{\frac{1-\nu}{1+\nu}}}}_{M} \|x - y\|^2 + \varepsilon. \tag{17}$$

Then by Lemma 1 after the stopping of Algorithm 2, $\min\limits_{k \in I} v_f(x^k, x_*) < \varepsilon$ means that the following inequality holds:

$$f(\widehat{x}) - f^* \leqslant \frac{M_\nu^{\frac{2}{1+\nu}}}{2\varepsilon^{\frac{1-\nu}{1+\nu}}} \varepsilon^2 + \varepsilon = \frac{M_\nu^{\frac{2}{1+\nu}}}{2} \varepsilon^{1+\frac{2\nu}{1+\nu}} + \varepsilon. \tag{18}$$

Note that for $\varepsilon < 1$ the inequality (18) means

$$f(\widehat{x}) - f^* \leqslant \widehat{M} \varepsilon$$

for some $\widehat{M} > 0$. So, for problems with a (quasi)convex Hölder-continuous differentiable (or $\widehat{D}f(x) \neq \{0\}$ for $x \in x_*$) objective functional and convex Lipschitz-continuous functional constraints we can achieve an ε-solution after

$$O\left(\frac{1}{\varepsilon^2}\right)$$

iterations of the Mirror Descent method. This estimate is optimal due to its optimality on a significantly narrower class of problems with Lipschitz-continuous objective functionals [Nemirovsky, 1983].

Optimal methods for Mirror Descent on the class of non-smooth strongly convex problems

Consider the optimization problem under the assumption of strong convexity of the objective function and functional constraint with the parameter μ.

$$f(x) \to \min, \quad g(x) \leqslant 0, \quad x \in Q, \tag{19}$$

where X is a closed convex set.

Let the prox function $d(x)$ be bounded on the unit sphere with respect to the chosen norm $\|\cdot\|$:

$$d(x) \leqslant \Omega^2 \quad \forall x \in Q \colon \|x\| \leqslant 1. \tag{20}$$

Let $x^0 \in Q$ and there exist $R_0 > 0$, such that $\|x^0 - x_*\|^2 \leqslant R_0^2$.

We will propose some methods which can guarantee an *ε-solution* $\widehat{x}$ of the problem (19):

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

The main idea is using the restart technique of Algorithm 2. Consider one well-known statement (see [Bayandina, 2018b]).

Lemma 4. *Let f and g be μ-strongly convex functionals with respect to the norm $\|\cdot\|$ on Q, $x_* = \arg\min_{x\in Q} f(x)$, $g(x) \leqslant 0$ ($\forall x \in Q$) and for some $\varepsilon_f > 0$ and $\varepsilon_g > 0$ the next inequalities hold:*

$$f(x) - f(x_*) \leqslant \varepsilon_f, \qquad g(x) \leqslant \varepsilon_g. \tag{21}$$

Then

$$\frac{\mu}{2}\|x - x_*\|^2 \leqslant \max\{\varepsilon_f, \varepsilon_g\}. \tag{22}$$

Let us consider an analogue of Algorithm 2 for strongly convex problems. We must emphasize that for Algorithm 2 one can obtain effective estimates of the rate of convergence for the objective functionals with any level of smoothness. Consider, in particular, the following example.

Let $f(x) = \max_{i=\overline{1,m}} f_i(x)$, where f_i are differentiable at any $x \in Q$ and their gradients are Lipschitz-continuous:

$$\|\nabla f_i(x) - \nabla f_i(y)\|_* \leqslant L_i\|x - y\| \quad \forall x, y \in Q \quad \forall i = \overline{1, m}. \tag{23}$$

Consider function $\tau \colon \mathbb{R}^+ \to \mathbb{R}^+$:

$$\tau(\delta) = \max\left\{\delta\|\nabla f(x_*)\|_* + \frac{\delta^2 L}{2},\ \delta\right\}, \tag{24}$$

where $L := \max_{i=\overline{1,m}}\{L_i\}$.

It is obvious that τ decreases, $\tau(0) = 0$, so for any $\varepsilon > 0$ there exists

$$\widehat{\varphi}(\varepsilon) > 0 \colon \tau(\widehat{\varphi}(\varepsilon)) = \varepsilon.$$

Algorithm 4. Restart procedure for Algorithm 2

Require: accuracy $\varepsilon > 0$; initial point x^0; Ω s.t. $d(x) \leqslant \Omega^2 \quad \forall x \in Q : \|x\| \leqslant 1$;
strong convexity parameter μ; R_0 such that $\|x^0 - x_*\|^2 \leqslant R_0^2$

1: Set $d_0(x) = d\left(\frac{x - x^0}{R_0}\right)$
2: Set $p = 1$
3: **repeat**
4: Set $R_p^2 = R_0^2 \cdot 2^{-p}$
5: Set $\varepsilon_p = \frac{\mu R_p^2}{2}$
6: Set x^p as the output of Algorithm 2 with accuracy $\widehat{\varphi}(\varepsilon_p)$, prox function $d_{p-1}(\cdot)$ and Ω^2
7: $d_p(x) \leftarrow d\left(\frac{x - x^p}{R_p}\right)$
8: Set $p = p + 1$
9: **until** $p > \log_2 \frac{\mu R_0^2}{2\varepsilon}$

Theorem 5. *Let ∇f be Lipschitz-continuous, f and g be μ-strongly convex on $Q \subset \mathbb{R}^n$ and $d(x) \leqslant \Omega^2$ for all $x \in Q$, such that $\|x\| \leqslant 1$. Let initial point $x^0 \in Q$ and $R_0 > 0$ satisfy*

$$\|x^0 - x_*\|^2 \leqslant R_0^2.$$

Then for $\widehat{p} = \left\lceil \log_2 \frac{\mu R_0^2}{2\varepsilon} \right\rceil$ output $x^{\widehat{p}}$ is an ε-solution of the problem (19), also the following inequalities hold:

$$f(x^{\widehat{p}}) - f(x_*) \leqslant \varepsilon, \quad g(x^{\widehat{p}}) \leqslant M_g \varepsilon,$$
$$\|x^{\widehat{p}} - x_*\|^2 \leqslant \frac{2\varepsilon}{\mu} \max\{1, M_g\}.$$

Proof. Function $d_p(x) = d\left(\frac{x - x^p}{R_p}\right)$, defined in Algorithm 4, is 1-strongly convex with respect to the norm $\frac{\|.\|}{R_p}$ for all $p \geqslant 0$. It is also easy to prove the following inequality

$$\|x^p - x_*\|^2 \leqslant R_p^2 \quad \forall p \geqslant 0.$$

If $p = 0$, the statement holds due to the choice of x^0 and R_0. Suppose that $\|x^p - x_*\|^2 \leqslant R_p^2$ for some p. Let us prove that $\|x^{p+1} - x_*\|^2 \leqslant R_{p+1}^2$. As $d_p(x_*) \leqslant \Omega^2$, on the restart number $(p + 1)$ after no more than

$$N_{p+1} = \left\lceil \frac{2\Omega^2 R_p^2}{\widehat{\varphi}^2(\varepsilon_{p+1})} \right\rceil$$

iterations of Algorithm 2, for $x^{p+1} = \bar{x}^{N_{p+1}}$ the next inequalities hold:

$$f(x^{p+1}) - f(x_*) \leqslant \varepsilon_{p+1}, \quad g(x^{p+1}) \leqslant \varepsilon_{p+1} M_g, \quad \text{if} \quad \varepsilon_{p+1} = \frac{\mu R_{p+1}^2}{2}.$$

According to Lemma 4,

$$\|x^{p+1} - x_*\|^2 \leqslant \frac{2\varepsilon_{p+1}}{\mu} \max\{1, M_g\} = R_{p+1}^2 \max\{1, M_g\}.$$

So, for any $p \geqslant 0$ we have proved that

$$\|x^p - x_*\|^2 \leqslant R_p^2 \max\{1, M_g\} = \frac{R_0^2}{2^p} \max\{1, M_g\},$$
$$f(x^p) - f(x_*) \leqslant \frac{\mu R_0^2}{2^{p+1}}, \quad g(x^p) \leqslant \frac{\mu R_0^2 M_g}{2^{p+1}}.$$

Consequently, $p = \widehat{p} = \left\lceil \log_2 \frac{\mu R_0^2}{2\varepsilon} \right\rceil$, output x^p is an ε-solution of the problem (19) and next inequalities hold:

$$\|x^p - x_*\|^2 \leqslant R_p^2 \max\{1, M_g\} = \frac{R_0^2}{2^p} \max\{1, M_g\} \leqslant \frac{2\varepsilon}{\mu} \max\{1, M_g\}.$$

Let K be the number of iterations of Algorithm 2 during the work of Algorithm 4, N_p be the total number of iterations of Algorithm 2 on the restart number p. As function $\tau\colon \mathbb{R}^+ \to \mathbb{R}^+$ increases and for any $\varepsilon > 0$ there exists $\widehat{\varphi}(\varepsilon) > 0$: $\tau(\widehat{\varphi}(\varepsilon)) = \varepsilon$, it means that

$$K = \sum_{p=1}^{\widehat{p}} N_p = \sum_{p=1}^{\widehat{p}} \left\lceil \frac{2\Omega^2 R_p^2}{\widehat{\varphi}^2(\varepsilon_p)} \right\rceil \leq \widehat{p} + \sum_{p=1}^{\widehat{p}} \frac{2\Omega^2 R_p^2}{\widehat{\varphi}^2(\varepsilon_p)}.$$

The number of iterations of Algorithm 2 during the work of Algorithm 4 will not exceed

$$\widehat{p} + \sum_{p=1}^{\widehat{p}} \frac{2\Omega^2}{\widehat{\varphi}^2(\varepsilon_p)}, \quad \text{where} \ \ \varepsilon_p = \frac{\mu R_0^2}{2^{p+1}}.$$

□

REMARK 3. The estimate of the number of iterations of Algorithm 2 can be detailed in the case of $\varepsilon < 1$. For any $\delta < 1$ there is such constant C, that $\tau(\delta) \leqslant C\delta$ for some constant C. So, we can suppose that $\widehat{\varphi}(\varepsilon) = \widehat{C} \cdot \varepsilon$ for the corresponding constant $\widehat{C} > 0$. On the restart number $p + 1$ of Algorithm 2 after no more than

$$k_{p+1} = \left\lceil \frac{2\Omega^2 R_p^2}{\varepsilon_{p+1}^2} \right\rceil \tag{25}$$

iterations of Algorithm 2, the output x^{p+1} satisfies the following inequality:

$$f(x^{p+1}) - f(x_*) \leqslant \widehat{C} \cdot \varepsilon_{p+1}, \quad g(x^{p+1}) \leqslant \varepsilon_{p+1},$$

where $\varepsilon_{p+1} = \dfrac{\mu R_{p+1}^2}{2}$. According to Lemma 4,

$$\|x^{p+1} - x_*\|^2 \leqslant \frac{2\max\{1, \widehat{C}\}\varepsilon_{p+1}}{\mu} = \max\{1, \widehat{C}\} \cdot R_{p+1}^2.$$

So, for all $p \geqslant 0$,

$$\|x^p - x_*\|^2 \leqslant \max\{1, \widehat{C}\} \cdot R_p^2 = \max\{1, \widehat{C}\} \cdot R_0^2 \cdot 2^{-p}.$$

Note, that for all $p \geqslant 1$ the following inequalities hold:

$$f(x^p) - f(x_*) \leqslant \max\{1, \widehat{C}\} \cdot \frac{\mu R_0^2}{2} \cdot 2^{-p}, \quad g(x_p) \leqslant \max\{1, \widehat{C}\} \cdot \frac{\mu R_0^2}{2} \cdot 2^{-p}.$$

Thereby, if $p > \log_2 \dfrac{\mu R_0^2}{2\varepsilon}$, then x^p will be $\left(\max\{1, \widehat{C}\}\varepsilon\right)$-solution to the problem, moreover:

$$\|x^p - x_*\|^2 \leqslant \max\{1, \widehat{C}\} \cdot R_0^2 \cdot 2^{-p} \leqslant \frac{2\varepsilon}{\mu}.$$

Let us evaluate the total number of iterations N of Algorithm 2. Let $\widehat{p} = \left\lceil \log_2 \dfrac{\mu R_0^2}{2\varepsilon} \right\rceil$. According to (25), up to multiplication by a constant we have:

$$N = \sum_{p=1}^{\widehat{p}} k_p \leqslant \sum_{p=1}^{\widehat{p}} \left(1 + \frac{2\Omega^2 R_p^2}{\varepsilon_{p+1}^2}\right) = \sum_{p=1}^{\widehat{p}} \left(1 + \frac{32\Omega^2 2^p}{\mu^2 R_0^2}\right) \leqslant \widehat{p} + \frac{64\Omega^2 2^{\widehat{p}}}{\mu^2 R_0^2} \leqslant \widehat{p} + \frac{64\Omega^2}{\mu\varepsilon}.$$

Note, that the method can be applied to solve the problem (1) in the case of a strongly quasi-convex objective functional.

REMARK 4. Function $f \colon Q \to \mathbb{R}$ is called strongly quasi-convex [Necoara, 2019], if for each $x \in Q$

$$f(x_*) - f(x) \geqslant \langle \nabla f(x), x - x_* \rangle + \frac{\mu}{2}\|x_* - x\|^2,$$

where x_* is the nearest (by Bregman divergence V) to x solution of the optimization problem.

Numerical Experiments

All calculations were performed in CPython 3.7 on a computer fitted with a 3-core AMD Athlon II X3 450 processor with a clock frequency of 3.2 GHz. The computer's RAM was 8 GB. As a rule, we indicate the operating time of the algorithms in seconds.

An analogue of the Fermat–Torricelli–Steiner problem

EXAMPLE 1. Input data: $n = 1000$, point coordinates $A_k = (a_{1k}, a_{2k}, \dots, a_{nk})$ $(k = 1, 2, \dots, 5)$ are represented by integers from the interval $[-10, 10]$, objective functional $(M_f = 1)$

$$f(x) = \frac{1}{5}\sum_{k=1}^{5} \sqrt{(x_1 - a_{1k})^2 + (x_2 - a_{2k})^2 + \ldots + (x_n - a_{nk})^2},$$

$x^0 = \dfrac{(0.1, \dots, 0.1)}{\|(0.1, \dots, 0.1)\|}$, $Q = \{x = (x_1, \dots, x_n) \mid \sum\limits_{i=1}^{n} x_i^2 \leqslant 1\}$, $\Theta_0^2 = 2$ and functional constraint

$$\begin{gathered} g(x) = \max_{m=1,2,3,\ldots,20} \{g_m(x)\} \leqslant 0, \\ g_1(x) = \alpha_{11}|x_1| + \alpha_{12}|x_2| + \ldots + \alpha_{1n}|x_n| - 1, \\ g_2(x) = \alpha_{21}|x_1| + \alpha_{22}|x_2| + \ldots + \alpha_{2n}|x_n| - 1, \\ \ldots \\ g_m(x) = \alpha_{m1}|x_1| + \alpha_{m2}|x_2| + \ldots + \alpha_{mn}|x_n| - 1, \end{gathered} \tag{26}$$

where the coefficients $\alpha_{11}, \alpha_{12}, \dots, \alpha_{mn}$ are represented by the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 2 & 2 & 2 & \ldots & 2 & 2 \\ 1 & 3 & 3 & 3 & \ldots & 3 & 3 \\ 1 & 2 & 3 & 4 & \ldots & 999 & 1000 \\ 1 & 3 & 4 & 5 & \ldots & 1000 & 1001 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 18 & 19 & 20 & \ldots & 1015 & 1016 \end{pmatrix}. \tag{27}$$

The averaged results of 10 experiments with a random selection of points A_k for Example 1 are presented in Table 1. As one can observe, Algorithm 2 works faster than Algorithm 1.

Table 1. Comparison of the results of the algorithms, Example 1

ε	Iterations	Time, s	Iterations	Time, s
	Algorithm 1		Algorithm 2	
1/2	—	>300	16	0.068
1/4	—	>300	64	0.264
1/6	—	>300	144	0.526
1/8	—	>300	256	0.920

An analogue of the problem of the smallest covering ball

EXAMPLE 2. Input data: $n = 1000$, point coordinates $A_k = (a_{1k}, a_{2k}, \dots, a_{nk})$ $(k = 1, 2, \dots, 5)$ are represented by integers from the interval $[-10, 10]$, objective functional $(M_f = 1)$

$$f(x) = \max_{k=\overline{1,5}}\left(\sqrt{(x_1 - a_{1k})^2 + (x_2 - a_{2k})^2 + \ldots + (x_n - a_{nk})^2}\right),$$

$x^0 = \dfrac{(0.1, \dots, 0.1)}{\|(0.1, \dots, 0.1)\|}$, $Q = \{x = (x_1, \dots, x_n) \mid \sum\limits_{i=1}^{n} x_i^2 \leqslant 1\}$ $\Theta_0^2 = 2$ and functional constraint (26), where the coefficients $\alpha_{11}, \alpha_{12}, \dots, \alpha_{mn}$ are represented by the matrix (27).

The averaged results of 10 experiments with a random selection of points A_k for Example 2 are presented in Table 2. As one can observe, Algorithm 2 works faster than Algorithm 1.

Table 2. Comparison of the results of the algorithms, Example 2

ε	Iterations	Time, s	Iterations	Time, s
	Algorithm 1		Algorithm 2	
1/2	—	>300	16	0.071
1/4	—	>300	64	0.259
1/6	—	>300	144	0.575
1/8	—	>300	256	1

An example of a concave objective functional satisfying the Hölder condition

EXAMPLE 3. Input data: $n = 1000$, objective functional ($M_{f,1/2} = 1$)

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{x_i},$$

$x^0 = \dfrac{(0.1,\ldots,0.1)}{\|(0.1,\ldots,0.1)\|}$, $Q = \{x = (x_1,\ldots,x_n) \mid x_i \geqslant 0\ \forall i\ \sum_{i=1}^{n} x_i^2 \leqslant 1\}$, $\Theta_0^2 = 2$ and functional constraint

$$\begin{aligned} g(x) &= \max_{m=1,2,3,\ldots,20}\{g_m(x)\},\\ g_1(x) &= \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n - 1 \leqslant 0,\\ g_2(x) &= \alpha_{21}x_1 + \alpha_{22}x_2 + \ldots + \alpha_{2n}x_n - 1 \leqslant 0,\\ &\ldots\\ g_m(x) &= \alpha_{m1}x_1 + \alpha_{m2}x_2 + \ldots + \alpha_{mn}x_n - 1 \leqslant 0, \end{aligned} \tag{28}$$

where the coefficients $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{mn}$ are represented by the matrix (27).

The results of Example 3 are presented in Table 3. As one can observe, Algorithm 2 works faster than Algorithm 1.

Examples with large dimensions

Table 4 presents the results of Algorithm 2 for the dimension $n = 3 \cdot 10^5$. Because of the large dimensionality it is impossible to obtain the results for Algorithm 1 in a relatively short period of time.

Table 3. Comparison of the results of the algorithms, Example 3

ε	Iterations	Time, s	Iterations	Time, s
	Algorithm 1		Algorithm 2	
1/2	—	>300	16	0.158
1/4	—	>300	64	0.575
1/6	—	>300	144	1.089
1/8	—	>300	256	1.848

Table 4. Some results of Algorithm 2 for $n = 3 \cdot 10^5$

ε	Iterations	Time, s	Iterations	Time, s	Iterations	Time, s
	Example 1		Example 2		Example 3	
1/2	16	34	16	30	16	35
1/4	64	123	64	118	64	141
1/6	144	278	144	272	144	326

An example of a geometrical problem of a quasi-convex objective functional

EXAMPLE 4. Suppose we are given several points A_k (the centers of the balls ω_k). It is necessary to find the ball of the smallest radius R that covers these points. In other words, it is necessary to find the center of such a ball that the maximum distance from the center to these points is the shortest possible. At the same time, we assume that the point (center) X can lie on some set which is defined by functional constraint (28), where the coefficients $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{mn}$ are represented by the matrix (27). The distance from X to each of the fixed points A_k is determined as follows:

$$d(X, A_k) = \begin{cases} XA_k + (\rho - 1)r_k, & \text{if } |XA_k| > r_k \ (r_k \text{ — radius of } \omega_k, \rho > 1), \\ \rho XA_k, & \text{otherwise,} \end{cases}$$

where $d(X, A_k) =: f(x)$ is a concave function ($M_f = \rho$). Note that $d(X, A_k)$ is non-smooth at points X: $|XA_k| = r_k$. For points of non-smoothness we use some element of Clarke subdifferential as an analogue of subgradient.

Other input data: $n = 1000$, $\rho = 2$, $x^0 = \dfrac{(0.1, \ldots, 0.1)}{\|(0.1, \ldots, 0.1)\|}$, $\Theta_0^2 = 2$. The coordinates of the points A_k are chosen in such a way that $\|A_k\| \in [1; 2]$, the number of points A_k is equal to 1000 and $r_k = 1$ for all $k = \overline{1, 100}$.

The averaged results of 10 experiments with a random selection of points A_k for Example 4 are presented in Table 5. As one can observe, Algorithm 2 works faster than Algorithm 1. Note that we can use Algorithm 2 for problems with quasi-convex objective functional and convex functional constraints (see Theorem 3). Another approach (Algorithm 3) assumes knowledge of the Lipschitz constant of the constraint M_g.

Table 5. Comparison of the results of the algorithms, Example 4

ε	Iterations	Time, s	Iterations	Time, s
	Algorithm 1		Algorithm 2	
1/2	32680	199	16	0.095
1/4	65392	392	64	0.391
1/6	98135	587	144	0.862
1/8	—	>1000	256	1
1/10	—	>1000	400	2
1/12	—	>1000	576	3

Conclusion

Summing up, let us remark the conclusions of the article. There was proposed an analogue of adaptive Mirror Descent [Bayandina, 2018a, Section 3.3] for convex programming problems with another step-size strategy. The estimates of the rate of its convergence were proved. Optimality in terms of lower bounds was stated. Moreover, it was shown that the proposed methods can be used to minimize quasi-convex objective functionals with different levels of smoothness. Also, using the restart technique an optimal method was proposed to solve optimization problems with strongly convex objective functionals. Some numerical experiments were carried out to solve geometrical problems with convex constraints. The advantages of the proposed methods were demonstrated during these experiments. Numerical examples for the minimization of quasi-convex functionals were given. They illustrate that the proposed methods work faster than [Bayandina, 2018a, Section 3.3], since in Algorithm 2 the number of iterations is fixed for a given ε due to the stopping criterion. However, functional constraint evaluation, generally, can deteriorate: $g(\bar{x}) < M_g\varepsilon$ instead of $g(\bar{x}) < \varepsilon$ in [Bayandina, 2018a].

The authors are very grateful to Y. Nesterov for fruitful discussions.

References

Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A. Mirror descent and convex optimization problems with non-smooth inequality constraints // Large-Scale and Distributed Optimization. Lecture Notes in Mathematics. — 2018. — Vol. 2227. — P. 181–213.

Bayandina A., Gasnikov A., Gasnikova E., Matsievsky S. Primal-dual method of Mirror Descent for stochastic constrained optimization problems // Journal of Computational Mathematics and Mathematical Physics. — 2018. — Vol. 58, No. 11. — P. 1728–1736.

Beck A., Teboulle M. Mirror descent and nonlinear projected subgradient methods for convex optimization // Operations Research Letters. — 2003. — Vol. 31, No. 3. — P. 167–175.

Ben-Tal A., Nemirovski A. Lectures on Modern Convex Optimization. — Philadelphia: Society for Industrial and Applied Mathematics, 2001.

Ben-Tal A., Nemirovski A. Robust Truss Topology Design via Semidefinite Programming // SIAM Journal on Optimization. — 1997. — Vol. 7, No. 4. — P. 991–1016.

Clarke F. Optimization and non-smooth analysis. — New York: John Wiley and Sons, 1983.

Gasnikov A. Modern numerical optimization methods. The method of universal gradient descent. — Moscow: MIPT, 2018.

Hazan E., Levy K., Shalev-Shwartz S. Beyond convexity: Stochastic quasi-convex optimization // Advances in Neural Information Processing Systems. — 2015. — P. 1594–1602.

Juditsky A., Nemirovski A. First order methods for nonsmooth convex large-scale optimization, I: general purpose methods // Optimization for Machine Learning. — 2010. — P. 121–148.

Konnov I. On convergence properties of a subgradient method // Optimization Methods and Software. — 2003. — Vol. 18, No. 1. — P. 53–62.

Necoara I., Nesterov Y. Glineur F. Linear convergence of first order methods for non-strongly convex optimization // Math. Program. — 2019. — Vol. 175. — P. 69–107.

Nemirovsky A., Yudin D. Problem Complexity and Method Efficiency in Optimization. — J. Wiley & Sons, New York, 1983.

Nesterov Yu. Effective methods in nonlinear programming. — Moscow: Radio and Communication, 1989.

Nesterov Yu. Introductory Lectures on Convex Optimization: a basic course. — Massachusetts: Kluwer Academic Publishers, 2004.

Nesterov Yu. Minimization methods for nonsmooth convex and quasiconvex functions // Math. Ekon. — 1984. — Vol. 29. — P. 519–531.

Nesterov Yu. Subgradient methods for Huge-Scale Optimization Problems // Math. Prog. — 2015. — Vol. 146, No. 1–2. — P. 275–297.

Shpirko S., Nesterov Yu. Primal-dual Subgradient Methods for Huge-scale Linear Conic Problem // SIAM Journal on Optimization. — 2014. — Vol. 24, No. 3. — P. 1444–1457.