КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2024 Т. 16 № 7 С. 1727–1746
DOI: 10.20537/2076-7633-2024-16-7-1727-1746

Ки&М

СПЕЦИАЛЬНЫЙ ВЫПУСК

УДК: 519.8

О некоторых методах зеркального спуска для задач сильно выпуклого программирования с липшицевыми функциональными ограничениями

О. С. Савчук[1,2,a], М. С. Алкуса[1,3,b], Ф. С. Стонякин[1,2,c]

[1]Московский физико-технический институт,
Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., д. 9
[2]Крымский федеральный университет им. В. И. Вернадского,
Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, д. 4
[3]Университет Иннополис,
Россия, 420500, г. Иннополис, ул. Университетская, д. 1

E-mail: [a] oleg.savchuk19@mail.ru, [b] mohammad.alkousa@phystech.edu, [c] fedyor@mail.ru

Получено 29.10.2024, после доработки — 13.11.2024
Принято к публикации 25.11.2024

Статья посвящена специальному подходу к субградиентным методам для задач сильно выпуклого программирования с несколькими функциональными ограничениями. Точнее говоря, рассматривается задача сильно выпуклой минимизации с несколькими сильно выпуклыми ограничениями-неравенствами и предлагаются оптимизационные методы первого порядка для такого класса задач. Особенность предложенных методов — возможность использования в теоретических оценках качества выдаваемого методом решения параметров сильной выпуклости именно тех функционалов ограничений, для которых нарушается условие продуктивности итерации. Основная задача — предложить для такой постановки субградиентный метод с адаптивными правилами подбора шагов и остановки метода. Ключевая идея предложенной в данной статье методики заключается в объединении двух подходов: схемы с переключениями по продуктивным и непродуктивным шагам и недавно предложенных модификаций зеркального спуска для задач выпуклого программирования, позволяющих игнорировать часть функциональных ограничений на непродуктивных шагах алгоритма. В статье описан субградиентний метод с переключением по продуктивным и непродуктивным шагам для задач сильно выпуклого программирования в случае, когда целевая функция и функциональные ограничения удовлетворяют условию Липшица. Также рассмотрен аналог этой схемы типа зеркального спуска для задач с относительно липшицевыми и относительно сильно выпуклыми целевой функцией и ограничениями. Для предлагаемых методов получены теоретические оценки качества выдаваемого решения, указывающие на оптимальность этих методов с точки зрения нижних оракульных оценок. Кроме того, поскольку во многих задачах операция нахождения точного вектора субградиента достаточно затратна, то для рассматриваемого класса задач исследованы аналоги указанных выше методов с заменой обычного субградиента на δ-субградиент целевого функционала или функциональных ограничений-неравенств. Отмеченный подход может позволить сэкономить вычислительные затраты метода за счет отказа от требования доступности точного значения субградиента в текущей точке. Показано, что оценки качества решения при этом изменяются на величину $O(\delta)$. Также приводятся результаты численных экспериментов, иллюстрирующие преимущество предлагаемых в статье методов в сравнении с некоторыми ранее известными.

Ключевые слова: субградиентный метод, зеркальный спуск, сильно выпуклая функция, липшицева функция, δ-субградиент, продуктивный шаг, непродуктивный шаг

Работа выполнена при поддержке гранта Российского научного фонда № 24-21-00210 (https://rscf.ru/project/24-21-00210/).

COMPUTER RESEARCH AND MODELING
2024 VOL. 16 NO. 7 P. 1727–1746
DOI: 10.20537/2076-7633-2024-16-7-1727-1746

SPECIAL ISSUE

UDC: 519.8

On some mirror descent methods for strongly convex programming problems with Lipschitz functional constraints

O. S. Savchuk[1,2,a], M. S. Alkousa[1,3,b], F. S. Stonyakin[1,2,c]

[1]Moscow Institute of Physics and Technology,
9 Institutskiy per., Dolgoprudny, Moscow region, 141701, Russia
[2]V. I. Vernadsky Crimean Federal University,
4 Academician Vernadsky ave., Simferopol, Republic of Crimea, 295007, Russia
[3]Innopolis University,
1 Universitetskaya st., Innopolis, 420500, Russia

E-mail: [a] oleg.savchuk19@mail.ru, [b] mohammad.alkousa@phystech.edu, [c] fedyor@mail.ru

Received 29.10.2024, after completion — 13.11.2024
Accepted for publication 25.11.2024

The paper is devoted to one approach to constructing subgradient methods for strongly convex programming problems with several functional constraints. More precisely, the strongly convex minimization problem with several strongly convex (inequality-type) constraints is considered, and first-order optimization methods for this class of problems are proposed. The special feature of the proposed methods is the possibility of using the strong convexity parameters of the violated functional constraints at nonproductive iterations, in theoretical estimates of the quality of the produced solution by the methods. The main task, to solve the considered problem, is to propose a subgradient method with adaptive rules for selecting steps and stopping rule of the method. The key idea of the proposed methods in this paper is to combine two approaches: a scheme with switching on productive and nonproductive steps and recently proposed modifications of mirror descent for convex programming problems, allowing to ignore some of the functional constraints on nonproductive steps of the algorithms. In the paper, it was described a subgradient method with switching by productive and nonproductive steps for strongly convex programming problems in the case where the objective function and functional constraints satisfy the Lipschitz condition. An analog of the proposed subgradient method, a mirror descent scheme for problems with relatively Lipschitz and relatively strongly convex objective functions and constraints is also considered. For the proposed methods, it obtained theoretical estimates of the quality of the solution, they indicate the optimality of these methods from the point of view of lower oracle estimates. In addition, since in many problems, the operation of finding the exact subgradient vector is quite expensive, then for the class of problems under consideration, analogs of the mentioned above methods with the replacement of the usual subgradient of the objective function or functional constraints by the δ-subgradient were investigated. The noted approach can save computational costs of the method by refusing to require the availability of the exact value of the subgradient at the current point. It is shown that the quality estimates of the solution change by $O(\delta)$. The results of numerical experiments illustrating the advantages of the proposed methods in comparison with some previously known ones are also presented.

Keywords: subgradient method, mirror descent, strongly convex function, Lipschitz-continuous function, δ-subgradient, productive step, nonproductive step

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 7, pp. 1727–1746 (Russian).
This work was supported by Russian Science Foundation, project 24-21-00210 (https://rscf.ru/project/24-21-00210/).

1. Введение

Задачи выпуклой минимизации с ограничениями-неравенствами часто встречаются в различных приложениях, в связи с чем им посвящаются всё новые исследования, в том числе и в области субградиентных методов (см., например, [Dvurechensky et al., 2020] и имеющиеся там ссылки). Первый субградиентный метод был предложен в [Shor, 1967] и обобщен для задач с ограничениями-неравенствами в [Поляк, 1967], где была предложена идея пошагового переключения между направлением субградиента целевого функционала и направлением субградиента функционала ограничения. Задачи минимизации негладкого функционала с ограничениями возникают в широких классах задач современной крупномасштабной оптимизации и ее приложений [Ben-Tal, Nemirovski, 1997; Shpirko, 2014]. Среди множества методов для таких задач можно отметить неевклидово обобщение стандартного субградиентного метода, так называемый метод зеркального спуска [Nemirovsky, 1979; Nemirovsky, Yudin, 1983; Beck, Teboulle, 2003]. Недавно в [Bayandina et al., 2018] были предложены алгоритмы зеркального спуска как с адаптивным выбором шага, так и с адаптивным критерием остановки. При этом, помимо случая липшицевых целевого функционала и функционального ограничения, был предложен оптимальный метод для класса условных задач выпуклой минимизации со специальными условиями роста (нестандартными) целевого функционала. Кроме того, в [Стонякин и др., 2018] также предложены адаптивные методы зеркального спуска для задач с липшицевым (вообще говоря, негладким) целевым функционалом и задач с липшицевым градиентом целевого функционала в случае, когда имеется несколько функциональных ограничений, а также рассматривается случай негладкого целевого функционала, равного максимуму нескольких гладких функционалов с липшицевым градиентом.

Данная работа посвящена некоторым модификациям субградиентных методов типа зеркального спуска для задач сильно выпуклого программирования с несколькими функциональными ограничениями. Точнее говоря, рассматривается задача сильно выпуклой минимизации с несколькими сильно выпуклыми ограничениями-неравенствами и описываются оптимизационные методы первого порядка для решения такого класса задач. Особенность предлагаемого в настоящей статье подхода заключается в возможности использования в теоретических оценках качества выдаваемого решения параметров сильной выпуклости именно тех функциональных ограничений, которые нарушаются на непродуктивных итерациях алгоритма. Основная цель данной статьи — предложить и исследовать для выделенного класса задач сильно выпуклого программирования субградиентные методы с адаптивным правилом выбора размера шага, а также адаптивным критерием остановки метода. Ключевая идея методов данной статьи — объединение двух подходов: схемы с переключениями по продуктивным и непродуктивным шагам для онлайн-оптимизации из [Savchuk et al., 2023] и недавно предложенных в [Стонякин и др., 2018] модификаций зеркального спуска для выпуклого программирования, позволяющих игнорировать часть функциональных ограничений на непродуктивных шагах алгоритма.

В настоящей статье предложен субградиентный метод с переключениями по продуктивным и непродуктивным шагам для задач сильно выпуклого программирования в случае, когда целевая функция и функциональные ограничения удовлетворяют условию Липшица. Также рассмотрен неевклидов аналог этой схемы типа зеркального спуска для задач с относительно липшицевыми [Nesterov, 2019; Lu, 2019] и относительно сильно выпуклыми [Lu, Freund, Nesterov, 2018] целевой функцией и функциональными ограничениями. Для рассматриваемых методов доказаны теоретические оценки качества выдаваемого решения, которые указывают на оптимальность предлагаемых методов с точки зрения нижних оракульных оценок. Кроме того, поскольку во многих задачах операция нахождения точного вектора субградиента может быть достаточно затратной, то для таких задач исследованы аналоги указанных выше методов на ситуацию, предполагающую доступность на итерациях метода вместо обычных субградиентов целевой функции или функциональных ограничений их δ-субградиентов (см., например, [Поляк, 1983]). Именно использование δ-субградиентов потенциально может сэкономить вычислительные затраты метода за счет отказа от требования доступности точного значения субградиента в текущей точке. Более того, показано, что оценки качества выдаваемого решения при этом изменяются на величину $O(\delta)$. Приведены результаты численных экспериментов, иллюстрирующие преимущества предлагаемых в статье методов в сравнении с ранее известными для некоторых примеров.

Работа состоит из введения, заключения и пяти основных параграфов. В § 2 приводится постановка задачи, а также основные определения и обозначения.

В § 3 исследована субградиентная схема с переключениями для задач сильно выпуклой минимизации с несколькими ограничениями-неравенствами при условии, что целевая функция и функциональные ограничения удовлетворяют условию Липшица. Доказаны теоретические оценки скорости сходимости предложенного метода.

В § 4 предложен субградиентный метод для задач сильно выпуклого программирования с использованием на итерациях алгоритма вместо обычных субградиентов неточных δ-субградиентов целевого функционала и функциональных ограничений.

В § 5 исследованы аналоги двух вышеуказанных алгоритмов — методы типа зеркального спуска для задач сильно выпуклого программирования с ограничениями с относительно липшицевыми и относительно сильно выпуклыми целевой функцией и ограничениями. Аналогично предыдущим разделам для рассмотренных методов доказаны теоретические оценки качества выдаваемого решения.

В § 6 приведены результаты некоторых проведенных численных экспериментов для иллюстрации эффективности предложенных в статье алгоритмов.

2. Основные определения и обозначения

Рассматривается следующая задача выпуклой минимизации с ограничениями:

$$\min\{f(x)\colon\ x \in Q,\ g_i(x) \leqslant 0,\ i = 1, 2, \ldots, m\}, \tag{1}$$

где $Q \subset \mathbb{R}^n$ — некоторое замкнутое выпуклое множество, $f\colon Q \to \mathbb{R}$ и $g_i\colon Q \to \mathbb{R}$, $i = 1, 2, \ldots, m$, — выпуклые функции, имеющие в каждой точке конечные субградиенты.

Пусть x^* — одно из точных решений задачи (1). Будем говорить, что точка $\widehat{x} \in Q$ является ε-решением задачи (1), если

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon, \quad g_p(\widehat{x}) \leqslant \varepsilon \quad \forall p = 1, 2, \ldots, m.$$

Определение 1. Будем говорить, что функция f удовлетворяет условию Липшица с константой $M > 0$ (или M-липшицева) на множестве Q, если справедливо неравенство

$$|f(x) - f(y)| \leqslant M\|x - y\|_2 \quad \forall x, y \in Q.$$

Здесь и всюду далее $\|\cdot\|_2$ — евклидова норма.

Определение 2. Функция f называется сильно выпуклой на множестве Q с константой $\mu > 0$ (или μ-сильно выпуклой), если справедливо неравенство

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{\lambda(1-\lambda)\mu}{2}\|x - y\|_2^2 \quad \forall x, y \in Q,\ \forall \lambda \in [0; 1].$$

В частности, эквивалентным условием μ-сильной выпуклости является следующее неравенство:

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|x - y\|_2^2 \quad \forall x, y \in Q,$$

справедливое для всякого субградиента $\nabla f(x)$ функции f в точке x.

Определение 3. Выпуклая функция $d\colon Q \to \mathbb{R}$ называется прокс-функцией, если d непрерывно дифференцируема в любой внутренней точке множества Q.

Определение 4. Дивергенцией (или расхождением) Брэгмана называют функцию

$$V(x, y) := d(x) - d(y) - \langle \nabla d(y), x - y\rangle \quad \forall x, y \in Q,$$

где d — произвольная прокс-функция.

Определение 5. Выпуклая функция $f\colon Q \to \mathbb{R}$ называется M-относительно липшицевой [Nesterov, 2019; Lu, 2019], если при некотором $M > 0$ справедливо неравенство

$$\langle \nabla f(x), y - x\rangle + M\sqrt{2V(y, x)} \geqslant 0 \quad \forall x, y \in Q.$$

Определение 6. Функция $f\colon Q \to \mathbb{R}$ называется μ-относительно сильно выпуклой [Lu, Freund, Nesterov, 2018] на выпуклом множестве Q, если при некотором $\mu > 0$ имеет место неравенство

$$f(x) \geqslant f(y) + \langle \nabla f(y), x - y\rangle + \mu V(x, y) \quad \forall x, y \in Q.$$

3. Субградиентная схема с переключениями для задач сильно выпуклого программирования с липшицевыми функционалами

В данном параграфе предложена схема с переключениями для задач сильно выпуклого программирования вида (1) в случае, когда целевая функция и функциональные ограничения удовлетворяют условию Липшица. Итак, рассмотрим следующий алгоритм.

Алгоритм 1. Субградиентный метод для задач сильно выпуклого программирования с липшицевыми функционалами

Require: $\varepsilon > 0$, $M = \max\limits_{1\leqslant p\leqslant m}\left\{M_f, M_{g_p}\right\} > 0$, $x_1 \in Q$.

$I := \varnothing$;
$N := 1$;
repeat
  if $g_p(x_N) \leqslant \varepsilon\ \forall p = 1, \ldots, m$ **then**
    $\eta_N^f = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^f\langle \nabla f(x_N), x\rangle + \frac{1}{2}\|x - x_N\|_2^2\right\}$; «продуктивный шаг»
    $N \to I$;
  else
    // $g_{p(N)}(x_N) > \varepsilon$ для некоторого $p(N) \in \{1, \ldots, m\}$
    $\eta_N^g = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^g\left\langle \nabla g_{p(N)}(x_N), x\right\rangle + \frac{1}{2}\|x - x_N\|_2^2\right\}$; «непродуктивный шаг»
  end if
  $N := N + 1$;
until $\dfrac{N}{\sum\limits_{k=1}^{N}\frac{1}{\mu_{1:k}}} \geqslant \frac{M^2}{2\varepsilon}$.

Ensure: $\widehat{x} = \arg\min\limits_{x_k,\, k\in I} f(x_k)$.

Обозначим множество продуктивных шагов ($g_p(x_k) \leqslant \varepsilon$, $p = 1, 2, \ldots, m$) через I, а множество непродуктивных шагов ($g_p(x_k) > \varepsilon$ для некоторого $1 \leqslant p \leqslant m$) — через J. Определим также $\mu_{1:k} := \sum\limits_{s=1}^{k} \mu_k$, где μ_k есть параметр сильной выпуклости функции f, если k — номер продуктивного шага. Пусть, кроме того, $\mu_{1:0} = 0$. Если же k — номер непродуктивного шага, для которого $g_p(x_k) > \varepsilon$, то тогда $\mu_k = \mu_{g_p}$, где μ_{g_p} — параметр сильной выпуклости функции g_p.

Для алгоритма 1 имеет место следующий результат.

Теорема 1. *Предположим, что функция $f(x)$ M_f-липшицева и μ_f-сильно выпукла, а все функциональные ограничения $g_i(x)$ M_{g_i}-липшицевы и μ_{g_i}-сильно выпуклы для произвольного $i = 1, 2, \ldots, m$. Тогда для любого $\varepsilon > 0$ после N итераций алгоритма 1 будет верно неравенство*

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} - \varepsilon|J|\right).$$

Если выполнен критерий остановки алгоритма 1:

$$\frac{N}{\sum\limits_{k=1}^{N} \frac{1}{\mu_{1:k}}} \geqslant \frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{2\varepsilon},$$

то множество продуктивных шагов не пусто и

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon, \quad g_p(\widehat{x}) \leqslant \varepsilon \quad \forall p = 1, 2, \ldots, m.$$

При этом алгоритм 1 *работает не более*

$$N = O\left(\frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{\min\limits_{1\leqslant p\leqslant m}\left\{\mu_f, \mu_{g_p}\right\}\varepsilon}\right) \tag{2}$$

итераций.

Доказательство. Пусть $[N] := \{k \in \overline{1, N}\}$, $J = [N] \setminus I$. Учитывая, что f M_f-липшицева, для каждого продуктивного шага ($k \in I$) имеем

$$\begin{aligned}\eta_k^f\left(f(x_k) - f(x^*)\right) &\leqslant \eta_k^f\left(\left\langle\nabla f(x_k), x_k - x^*\right\rangle - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2\right) \leqslant \\ &\leqslant \frac{1}{2}\left(\left(\eta_k^f\right)^2\|\nabla f(x_k)\|_2^2 + \left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2 - \eta_k^f\mu_k\left\|x^* - x_k\right\|_2^2\right).\end{aligned}$$

Отсюда после деления обеих частей приведенного выше неравенства на η_k^f получим

$$\begin{aligned}f(x_k) - f(x^*) &\leqslant \frac{1}{2}\left(\eta_k^f\|\nabla f(x_k)\|_2^2 + \frac{\left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2}{\eta_k^f} - \mu_k\left\|x^* - x_k\right\|_2^2\right) = \\ &= \frac{1}{2}\left(\frac{1}{\mu_{1:k}}\|\nabla f(x_k)\|_2^2 + \mu_{1:k}\|x^* - x_k\|_2^2 - \mu_k\|x^* - x_k\|_2^2 - \mu_{1:k}\left\|x^* - x_{k+1}\right\|_2^2\right) \leqslant \\ &\leqslant \frac{M_f^2}{2\mu_{1:k}} + \frac{\mu_{1:k}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 = \\ &= \frac{M_f^2}{2\mu_{1:k}} + \frac{\mu_{1:k-1}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2.\end{aligned} \tag{3}$$

Аналогично, учитывая, что каждая из функций-ограничений g_i, $i = 1, \ldots, m$, M_{g_i}-липшицева, для всякого непродуктивного шага ($k \in J$) будем иметь

$$\eta_k^g \left(g_{p(k)}(x_k) - g_{p(k)}(x^*)\right) \leqslant \eta_k^g \left(\left\langle \nabla g_{p(k)}(x_k), x_k - x^* \right\rangle - \frac{\mu_k}{2} \left\|x^* - x_k\right\|_2^2\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\left(\eta_k^g\right)^2 \|\nabla g_{p(k)}(x_k)\|_2^2 + \left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2 - \eta_k^g \mu_k \left\|x^* - x_k\right\|_2^2\right).$$

Поделив обе части последнего неравенства на η_k^g, получим

$$\begin{aligned} g_{p(k)}(x_k) - g_{p(k)}(x^*) &\leqslant \frac{1}{2}\left(\eta_k^g \|\nabla g_{p(k)}(x_k)\|_2^2 + \frac{\left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2}{\eta_k^g} - \mu_k \left\|x^* - x_k\right\|_2^2\right) = \\ &= \frac{1}{2}\left(\frac{1}{\mu_{1:k}} \|\nabla g_{p(k)}(x_k)\|_2^2 + \mu_{1:k} \left\|x^* - x_k\right\|_2^2 - \mu_k \left\|x^* - x_k\right\|_2^2 - \mu_{1:k} \left\|x^* - x_{k+1}\right\|_2^2\right) \leqslant \\ &\leqslant \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \frac{\mu_{1:k}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 = \\ &= \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \frac{\mu_{1:k-1}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2. \end{aligned} \tag{4}$$

Просуммируем неравенства (3) и (4) по продуктивным и непродуктивным шагам. Обозначив $M = \max\limits_{1 \leqslant p \leqslant m} \left\{M_f, M_{g_p}\right\}$, в итоге получим, что

$$\sum_{k \in I}\left(f(x_k) - f(x^*)\right) + \sum_{k \in J}\left(g_{p(k)}(x_k) - g(x^*)\right) \leqslant \frac{1}{2}\sum_{k=1}^{N}\left(\frac{M^2}{\mu_{1:k}} + \mu_{1:k-1}\left\|x^* - x_k\right\|_2^2 - \mu_{1:k}\left\|x^* - x_{k+1}\right\|_2^2\right) =$$
$$= \frac{1}{2}\sum_{k=1}^{N}\frac{M^2}{\mu_{1:k}} - \frac{\mu_{1:N}}{2}\left\|x^* - x_{N+1}\right\|_2^2 \leqslant \frac{1}{2}\sum_{k=1}^{N}\frac{M^2}{\mu_{1:k}}.$$

Используя тот факт, что для всякого непродуктивного шага

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \geqslant g_{p(k)}(x_k) > \varepsilon,$$

получим следующую оценку:

$$\sum_{k \in I}\left(f(x_k) - f(x^*)\right) \leqslant \frac{1}{2}\sum_{k=1}^{N}\frac{M^2}{\mu_{1:k}} - \sum_{k \in J}\left(g_{p(k)}(x_t) - g_{p(k)}(x^*)\right) \leqslant$$
$$\leqslant \frac{1}{2}\sum_{k=1}^{N}\frac{M^2}{\mu_{1:k}} - \sum_{k \in J}\varepsilon = \frac{1}{2}\sum_{k=1}^{N}\frac{M^2}{\mu_{1:k}} - \varepsilon|J|.$$

Поскольку

$$\sum_{k \in I}\left(f(x_k) - f(x^*)\right) \geqslant |I|(f(\widehat{x}) - f(x^*)),$$

то

$$|I|\left(f(\widehat{x}) - f(x^*)\right) \leqslant \sum_{k \in I}\left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} - \varepsilon|J|.$$

Тогда после выполнения критерия остановки алгоритма 1 имеем

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon, \quad g_p(\widehat{x}) \leqslant \varepsilon \quad \forall p = 1, 2, \ldots, m.$$

Оценим теперь достаточное для достижения приемлемого качества решения количество итераций алгоритма 1. Предположим, что $\mu_k \geqslant \mu > 0$ для всех $1 \leqslant k \leqslant N$. Тогда

$$\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} - \varepsilon|J| \leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu k} - \varepsilon|J| \leqslant \frac{M^2}{2\mu}(1 + \ln N) - \varepsilon|J|.$$

Таким образом, в этом случае

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\frac{M^2}{2\mu}(1 + \ln N) - \varepsilon|J|\right).$$

Следовательно, $f(\widehat{x}) - f(x^*) \leqslant \varepsilon$ при

$$\frac{M^2}{2\mu}(1 + \ln N) - \varepsilon|J| \leqslant |I|\varepsilon,$$

или

$$\frac{M^2}{2\mu\varepsilon} \leqslant \frac{N}{1 + \ln N}. \tag{5}$$

Пусть число итераций N ограничено сверху некоторой константой, например $N < e^{10}$. Тогда $\ln N < 10$ и $\frac{11M^2}{2\mu\varepsilon} < N < e^{10}$. Таким образом, для достижения заданной точности $\varepsilon > 0$ потребуется

$$N = \left\lceil \frac{11 \max\limits_{1 \leqslant p \leqslant m} \left\{M_f^2, M_{g_p}^2\right\}}{2 \min\limits_{1 \leqslant p \leqslant m} \left\{\mu_f, \mu_{g_p}\right\} \varepsilon} \right\rceil$$

итераций алгоритма 1.

Кроме того, с увеличением N неравенство (5) не нарушается. Действительно,

$$\lim_{N\to\infty} \frac{N}{1 + \ln N} = \lim_{N\to\infty} \frac{1}{\frac{1}{N} + \frac{\ln N}{N}} = \infty.$$

Таким образом, с ростом значения N правая часть неравенства (5) также будет увеличиваться, а потому даже при выходе N за выбранную верхнюю границу $N < e^{10}$ оценка (5) сохраняется.

Покажем также, что наш алгоритм обязательно будет совершать хотя бы один продуктивный шаг. Действительно, предположим, что число продуктивных шагов равно нулю. Тогда

$$\varepsilon|J| \leqslant \sum_{k\in J} \left(g_{p(k)}(x_k) - g_{p(k)}(x^*)\right) \leqslant \frac{M^2}{2\mu}(1 + \ln|J|).$$

Очевидно, что при достаточно большом $|J|$ приведенное выше неравенство не выполняется. Таким образом, для достаточно большого числа непродуктивных шагов обязательно будет по крайней мере один продуктивный. $\square$

4. Субградиентный метод для задач сильно выпуклого программирования с использованием на итерациях δ-субградиентов

Во многих задачах операция нахождения точного вектора субградиента целевого функционала может быть достаточно затратной. Одним из подходов к указанной проблеме может быть использование вместо субградиентов функции f δ-субградиентов $\nabla_\delta f(x)$ [Поляк, 1983] в текущей точке x:

$$f(y) - f(x) \geqslant \langle \nabla_\delta f(x), y - x\rangle - \delta$$

для всякого $y \in Q$ при некотором фиксированном $\delta \geqslant 0$ (очевидно, что $\nabla_\delta f(x) = \nabla f(x)$ — обычный субградиент при $\delta = 0$).

Например [Поляк, 1983], для функции $f(x) = \max\limits_{y\in P} \varphi(x, y)$, где φ выпукла по x и непрерывна по $y \in P$ (P — некоторое компактное множество), $\nabla_\delta f(x)$ есть обычный субградиент по переменной x функции $\varphi(x, y_\delta)$, такой, что

$$\max_{y\in P} \varphi(x, y) - \varphi(x, y_\delta) \leqslant \delta.$$

В качестве примера задачи, для которой можно использовать вместо обычного субградиента δ-субградиент, стоит также упомянуть задачу проектирования точки на выпуклый компакт X_*. Действительно, если точка x далека от множества X_*, то при нахождении субградиента функции расстояния от точки до множества можно пренебречь точностью направления единичного вектора субградиента и использовать δ-субградиент вместо обычного субградиента.

Допустим теперь, что нам недоступна информация о точном значении субградиентов ∇f и ∇g_i ($\forall i = 1, \ldots, m$), используемая на итерациях, а доступны лишь δ-субградиенты $\nabla_\delta f$ и $\nabla_\delta g_i$ ($\forall i = 1, \ldots, m$). Изначально оговорим, что здесь и далее в статье будут рассматриваться $\nabla_\delta f$ и $\nabla_\delta g_i$ ($\forall i = 1, \ldots, m$), удовлетворяющие следующим условиям:

$$\begin{gathered} f(y) - f(x) \geqslant \langle \nabla f(x), y - x\rangle \geqslant \langle \nabla_\delta f(x), y - x\rangle - \delta, \\ g_i(y) - g_i(x) \geqslant \langle \nabla g_i(x), y - x\rangle \geqslant \langle \nabla_\delta g_i(x), y - x\rangle - \delta \quad \forall i = 1, \ldots, m, \end{gathered}$$

и с учетом этого предложим алгоритм 2.

Для предложенного алгоритма 2 справедлива следующая теорема.

Теорема 2. *Пусть функция $f(x)$ M_f-липшицева и μ_f-сильно выпукла, а функциональные ограничения $g_i(x)$ M_{g_i}-липшицевы и μ_{g_i}-сильно выпуклы для каждого $i = 1, 2, \ldots, m$. Предположим также, что $\|\nabla_\delta f(x)\|_2 \leqslant M_f$ и $\|\nabla_\delta g_i(x)\|_2 \leqslant M_{g_i}$ $\forall x \in Q$, где $i = 1, \ldots, m$. Тогда для любого $\varepsilon > 0$ после N итераций алгоритма 2 будет верно неравенство*

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J|\right).$$

Если выполнен критерий остановки алгоритма 2:

$$\frac{N}{\sum\limits_{k=1}^{N} \frac{1}{\mu_{1:k}}} \geqslant \frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{2\varepsilon},$$

то множество продуктивных шагов не пусто и

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon + \delta, \quad g_p(\widehat{x}) \leqslant \varepsilon + \delta \quad \forall p = 1, 2, \ldots, m.$$

Алгоритм 2. Субградиентный метод для задач сильно выпуклого программирования с использованием на итерациях δ-субградиентов

Require: $\varepsilon > 0$, $M = \max\limits_{1\leqslant p\leqslant m}\left\{M_f, M_{g_p}\right\} > 0$, $x_1 \in Q$.

$I := \varnothing$;
$N := 1$;
repeat
  if $g_p(x_N) \leqslant \varepsilon + \delta$ $\forall p = 1, 2, \ldots, m$ **then**
    $\eta_N^f = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^f\langle\nabla_\delta f(x_N), x\rangle + \frac{1}{2}\|x - x_N\|_2^2\right\}$;   «продуктивный шаг»
    $N \to I$;
    $N := N + 1$;
  else
    // $g_{p(N)}(x_N) > \varepsilon + \delta$ для некоторого $p(N) \in \{1, \ldots, m\}$
    $\eta_N^g = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^g\left\langle\nabla_\delta g_{p(N)}(x_N), x\right\rangle + \frac{1}{2}\|x - x_N\|_2^2\right\}$;   «непродуктивный шаг»
    $N := N + 1$;
  end if
until $\frac{N}{\sum_{k=1}^{N}\frac{1}{\mu_{1:k}}} \geqslant \frac{M^2}{2\varepsilon}$.

Ensure: $\widehat{x} = \arg\min\limits_{x_k,\, k\in I} f(x_k)$.

При этом алгоритм 2 *работает не более*

$$N = O\left(\frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{\min\limits_{1\leqslant p\leqslant m}\left\{\mu_f, \mu_{g_p}\right\}\varepsilon}\right)$$

итераций.

Доказательство. Для всякого продуктивного шага ($k \in I$) имеем

$$\eta_k^f\left(f(x_k) - f(x^*)\right) \leqslant \eta_k^f\left(\left\langle\nabla_\delta f(x_k), x_k - x^*\right\rangle - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 + \delta\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\left(\eta_k^f\right)^2\|\nabla_\delta f(x_k)\|_2^2 + \left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2 - \eta_k^f\mu_k\left\|x^* - x_k\right\|_2^2\right) + \eta_k^f\delta.$$

Отсюда после деления обеих частей приведенного выше неравенства на η_k^f получается

$$f(x_k) - f(x^*) \leqslant \frac{1}{2}\left(\eta_k^f\|\nabla_\delta f(x_k)\|_2^2 + \frac{\left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2}{\eta_k^f} - \mu_k\left\|x^* - x_k\right\|_2^2\right) + \delta =$$
$$= \frac{1}{2}\left(\frac{1}{\mu_{1:k}}\|\nabla_\delta f(x_k)\|_2^2 + \mu_{1:k}\left\|x^* - x_k\right\|_2^2 - \mu_k\left\|x^* - x_k\right\|_2^2 - \mu_{1:k}\left\|x^* - x_{k+1}\right\|_2^2\right) + \delta \leqslant$$
$$\leqslant \frac{M_f^2}{2\mu_{1:k}} + \frac{\mu_{1:k}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 + \delta =$$
$$= \frac{M_f^2}{2\mu_{1:k}} + \frac{\mu_{1:k-1}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 + \delta. \qquad (6)$$

Аналогично для каждого непродуктивного шага ($k \in J$) будем иметь

$$\eta_k^g\left(g_{p(k)}(x_k) - g_{p(k)}(x^*)\right) \leqslant \eta_k^g\left(\left\langle \nabla_\delta g_{p(k)}(x_k),\ x_k - x^*\right\rangle - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 + \delta\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\left(\eta_k^g\right)^2\|\nabla_\delta g_{p(k)}(x_k)\|_2^2 + \left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2 - \eta_k^g\mu_k\left\|x^* - x_k\right\|_2^2\right) + \eta_k^g\delta.$$

Разделив обе части последнего неравенства на η_k^g, получаем

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \leqslant \frac{1}{2}\left(\eta_k^g\|\nabla_\delta g_{p(k)}(x_k)\|_2^2 + \frac{\left\|x^* - x_k\right\|_2^2 - \left\|x^* - x_{k+1}\right\|_2^2}{\eta_k^g} - \mu_k\left\|x^* - x_k\right\|_2^2\right) + \delta =$$
$$= \frac{1}{2}\left(\frac{1}{\mu_{1:k}}\|\nabla_\delta g_{p(k)}(x_k)\|_2^2 + \mu_{1:k}\left\|x^* - x_k\right\|_2^2 - \mu_k\left\|x^* - x_k\right\|_2^2 - \mu_{1:k}\left\|x^* - x_{k+1}\right\|_2^2\right) + \delta \leqslant$$
$$\leqslant \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \frac{\mu_{1:k}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_k}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 + \delta =$$
$$= \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \frac{\mu_{1:k-1}}{2}\left\|x^* - x_k\right\|_2^2 - \frac{\mu_{1:k}}{2}\left\|x^* - x_{k+1}\right\|_2^2 + \delta. \tag{7}$$

Просуммируем неравенства (6) и (7) по продуктивным и непродуктивным шагам. Обозначив через $M = \max\limits_{1\leqslant p\leqslant m}\left\{M_f,\ M_{g_p}\right\}$, в итоге получим, что

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) + \sum_{k\in J}\left(g_{p(k)}(x_k) - g(x^*)\right) \leqslant \frac{1}{2}\sum_{k=1}^{N}\left(\frac{M^2}{\mu_{1:k}} + \mu_{1:k-1}\left\|x^* - x_k\right\|_2^2 - \mu_{1:k}\left\|x^* - x_{k+1}\right\|_2^2 + \delta\right) \leqslant$$
$$\leqslant \frac{1}{2}\sum_{k=1}^{N}\left(\frac{M^2}{\mu_{1:k}} - \mu_{1:N}\left\|x^* - x_N\right\|_2^2 + \delta\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} + N\delta.$$

Используя тот факт, что для непродуктивных шагов

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \geqslant g_{p(k)}(x_k) > \varepsilon + \delta,$$

получим следующую оценку:

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} + N\delta - \sum_{k\in J}\left(g_{p(k)}(x_t) - g_{p(k)}(x^*)\right) \leqslant$$
$$\leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} + N\delta - \sum_{k\in J}(\varepsilon + \delta) = \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J|.$$

Поскольку

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) \geqslant |I|\left(f(\widehat{x}) - f(x^*)\right),$$

то

$$|I|\left(f(\widehat{x}) - f(x^*)\right) \leqslant \sum_{k\in I}\left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J|.$$

После выполнения критерия остановки алгоритма 2 будут верны неравенства

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon + \delta, \quad g_p(\widehat{x}) \leqslant \varepsilon + \delta \quad \forall p = 1, 2, \ldots, m.$$

Аналогично сделаем оценку на количество итераций, требуемое для достижения приемлемого качества решения. Пусть $\mu_k \geqslant \mu > 0$ для любого $1 \leqslant k \leqslant N$. Тогда

$$\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J| \leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu k} + \delta|I| - \varepsilon|J| \leqslant \frac{M^2}{2\mu}(1 + \ln N) + \delta|I| - \varepsilon|J|.$$

Таким образом, в этом случае

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\frac{M^2}{2\mu}(1 + \ln N) + \delta|I| - \varepsilon|J|\right).$$

Следовательно, $f(\widehat{x}) - f(x^*) \leqslant \varepsilon + \delta$ при

$$\frac{M^2}{2\mu}(1 + \ln N) + \delta|I| - \varepsilon|J| \leqslant |I|\varepsilon + \delta|I|,$$

или

$$\frac{M^2}{2\mu\varepsilon} \leqslant \frac{N}{1 + \ln N}.$$

Дальнейшие рассуждения аналогичны тем, что были представлены в доказательстве теоремы 1. □

5. Зеркальный спуск для задач сильно выпуклой минимизации с ограничениями. Обобщенная относительная M-липшицевость и μ-сильная выпуклость

В данном параграфе предлагаются методы зеркального спуска для задач выпуклого программирования с относительно липшицевыми и относительно сильно выпуклыми функционалами как неевклидовы обобщения методов 1 и 2 для негладких задач. Указанные методы описаны далее как алгоритмы 3 и 4.

Для алгоритма 3 справедлива следующая теорема.

Теорема 3. *Предположим, что $f(x)$ — M_f-относительно липшицева и μ_f-относительно сильно выпуклая функция, а все функции ограничений $g_i(x)$ M_{g_i}-относительно липшицевы и μ_{g_i}-относительно сильно выпуклы для каждого $i = 1, 2, \ldots, m$. Тогда для любого $\varepsilon > 0$ после N итераций алгоритма 3 будет верно неравенство*

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} - \varepsilon|J|\right).$$

Если выполнен критерий остановки алгоритма 3:

$$\frac{N}{\sum\limits_{k=1}^{N} \frac{1}{\mu_{1:k}}} \geqslant \frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2,\, M_{g_p}^2\right\}}{2\varepsilon},$$

то множество продуктивных шагов не пусто и

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon, \quad g_p(\widehat{x}) \leqslant \varepsilon \quad \forall p = 1, 2, \ldots, m.$$

Алгоритм 3. Зеркальный спуск для задач выпуклого программирования с относительно липшицевыми и относительно сильно выпуклыми функционалами

Require: $\varepsilon > 0$, $M = \max\limits_{1\leqslant p\leqslant m}\left\{M_f, M_{g_p}\right\} > 0$, $x_1 \in Q$.

$I := \varnothing$;
$N := 1$;
repeat
  if $g_p(x_N) \leqslant \varepsilon\ \forall p = 1, 2, \ldots, m$ **then**
    $\eta_N^f = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^f\langle\nabla f(x_N), x\rangle + V(x, x_N)\right\}$;   «продуктивный шаг»
    $N \to I$;
    $N := N + 1$;
  else
    // $g_{p(N)}(x_N) > \varepsilon$ для некоторого $p(N) \in \{1, \ldots, m\}$
    $\eta_N^g = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x\in Q}\left\{\eta_N^g\langle\nabla g_{p(N)}(x_N), x\rangle + V(x, x_N)\right\}$;   «непродуктивный шаг»
    $N := N + 1$;
  end if
until $\frac{N}{\sum_{k=1}^{N}\frac{1}{\mu_{1:k}}} \geqslant \frac{M^2}{2\varepsilon}$.

Ensure: $\widehat{x} = \arg\min\limits_{x_k,\, k\in I} f(x_k)$.

При этом алгоритм 3 *работает не более*

$$N = O\left(\frac{\max\limits_{1\leqslant p\leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{\min\limits_{1\leqslant p\leqslant m}\left\{\mu_f, \mu_{g_p}\right\}\varepsilon}\right)$$

итераций.

Доказательство. Пусть $[N] := \{k \in \overline{1, N}\}$, $J = [N] \setminus I$. Учитывая, что f M_f-относительно липшицева, для каждого продуктивного шага ($k \in I$) имеем

$$\eta_k^f\left(f(x_k) - f(x^*)\right) \leqslant \eta_k^f\left(\left\langle\nabla f(x_k), x_k - x^*\right\rangle - \mu_k V\left(x^*, x_k\right)\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\eta_k^f\right)^2 M_f^2 + V\left(x^*, x_k\right) - V\left(x^*, x_{k+1}\right) - \eta_k^f\mu_k V\left(x^*, x_k\right).$$

Отсюда после деления обеих частей приведенного выше неравенства на η_k^f получается

$$\begin{aligned} f(x_k) - f(x^*) &\leqslant \frac{1}{2}\eta_k^f M_f^2 + \frac{V\left(x^*, x_k\right) - V\left(x^*, x_{k+1}\right)}{\eta_k^f} - \mu_k V\left(x^*, x_k\right) = \\ &= \frac{1}{2\mu_{1:k}}M_f^2 + \mu_{1:k}V\left(x^*, x_k\right) - \mu_k V\left(x^*, x_k\right) - \mu_{1:k}V\left(x^*, x_{k+1}\right) = \\ &= \frac{M_f^2}{2\mu_{1:k}} + \mu_{1:k-1}V\left(x^*, x_k\right) - \mu_{1:k}V\left(x^*, x_{k+1}\right). \end{aligned} \tag{8}$$

Аналогично, учитывая, что каждая из функций-ограничений g_i, $i = 1, \ldots, m$, M_{g_i}-липшицева, для всякого непродуктивного шага ($k \in J$) будем иметь

$$\eta_k^g \left(g_{p(k)}(x_k) - g_{p(k)}(x^*)\right) \leqslant \eta_k^g \left(\left\langle \nabla g_{p(k)}(x_k),\, x_k - x^*\right\rangle - \mu_k V\left(x^*,\, x_k\right)\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\eta_k^g\right)^2 M_{g_{p(k)}}^2 + V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right) - \eta_k^g \mu_k V\left(x^*,\, x_k\right).$$

Поделив обе части последнего неравенства на η_k^g, получаем

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \leqslant \frac{1}{2}\eta_k^g M_{g_{p(k)}}^2 + \frac{V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right)}{\eta_k^g} - \mu_k V\left(x^*,\, x_k\right) =$$
$$= \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \mu_{1:k} V\left(x^*,\, x_k\right) - \mu_k V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) =$$
$$= \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \mu_{1:k-1} V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right). \tag{9}$$

Просуммируем неравенства (8), (9) по продуктивным и непродуктивным шагам, при этом $M = \max\limits_{1\leqslant p\leqslant m}\left\{M_f,\, M_{g_p}\right\}$. В итоге получим следующее:

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) + \sum_{k\in J}\left(g_{p(k)}(x_k) - g(x^*)\right) \leqslant \sum_{k=1}^{N}\left(\frac{M^2}{2\mu_{1:k}} + \mu_{1:k-1} V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right)\right) =$$
$$= \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} - \mu_{1:N} V\left(x^*,\, x_{N+1}\right) \leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}}.$$

Используя тот факт, что для всякого непродуктивного шага

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \geqslant g_{p(k)}(x_k) > \varepsilon,$$

получим следующую оценку:

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} - \sum_{k\in J}\left(g_{p(k)}(x_t) - g_{p(k)}(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} - \sum_{k\in J}\varepsilon = \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} - \varepsilon|J|.$$

Поскольку

$$\sum_{k\in I}\left(f(x_k) - f(x^*)\right) \geqslant |I|\,(f(\widehat{x}) - f(x^*)),$$

то

$$|I|\,(f(\widehat{x}) - f(x^*)) \leqslant \sum_{k\in I}\left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}} - \varepsilon|J|.$$

После выполнения критерия остановки алгоритма 3 будем иметь

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon, \quad g_p(\widehat{x}) \leqslant \varepsilon \quad \forall p = 1, 2, \ldots, m.$$

Оценка на количество итераций, достаточное для достижения заданной точности решения, полностью совпадает с оценкой из доказательства теоремы 1. □

Предложим теперь аналог алгоритма 3 для задач с δ-субградиентами целевого функционала и функциональных ограничений. Предложенный алгоритм указан ниже как алгоритм 4.

Для алгоритма 4 имеет место следующий результат.

Алгоритм 4. Зеркальный спуск для задач выпуклого программирования с относительно сильно выпуклыми функционалами и с использованием δ-субградиентов

Require: $\varepsilon > 0$, $M = \max\limits_{1 \leqslant p \leqslant m} \left\{M_f, M_{g_p}\right\} > 0$, $x_1 \in Q$.

$I := \varnothing$;
$N := 1$;
repeat
  if $g_p(x_N) \leqslant \varepsilon + \delta \ \forall p = 1, 2, \ldots, m$ **then**
    $\eta_N^f = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x \in Q} \left\{\eta_N^f \langle \nabla_\delta f(x_N), x\rangle + V(x, x_N)\right\}$; «продуктивный шаг»
    $N \to I$;
    $N := N + 1$;
  else
    // $g_{p(N)}(x_N) > \varepsilon + \delta$ для некоторого $p(N) \in \{1, \ldots, m\}$
    $\eta_N^g = \frac{1}{\mu_{1:N}}$;
    $x_{N+1} := \arg\min\limits_{x \in Q} \left\{\eta_N^g \langle \nabla_\delta g_{p(N)}(x_N), x\rangle + V(x, x_N)\right\}$; «непродуктивный шаг»
    $N := N + 1$;
  end if
until $\frac{N}{\sum\limits_{k=1}^{N} \frac{1}{\mu_{1:k}}} \geqslant \frac{M^2}{2\varepsilon}$.

Ensure: $\widehat{x} = \arg\min\limits_{x_k,\, k \in I} f(x_k)$.

Теорема 4. *Пусть $f(x)$ — M_f-относительно липшицева и μ_f-относительно сильно выпуклая функция, а функции ограничений $g_i(x)$ M_{g_i}-относительно липшицевы и μ_{g_i}-относительно сильно выпуклы для каждого $i = 1, 2, \ldots, m$. Предположим также, что $\|\nabla_\delta f(x)\|_2 \leqslant M_f$ и $\|\nabla_\delta g_i(x)\|_2 \leqslant M_{g_i}$ $\forall x \in Q$, где $i = 1, \ldots, m$. Тогда для любого $\varepsilon > 0$ после N итераций алгоритма 4 будет верно неравенство*

$$f(\widehat{x}) - f(x^*) \leqslant \frac{1}{|I|}\left(\sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J|\right).$$

Если выполнен критерий остановки алгоритма 4:

$$\frac{N}{\sum\limits_{k=1}^{N} \frac{1}{\mu_{1:k}}} \geqslant \frac{\max\limits_{1 \leqslant p \leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{2\varepsilon},$$

тогда множество продуктивных шагов не пусто и

$$f(\widehat{x}) - f(x^*) \leqslant \varepsilon + \delta, \quad g_p(\widehat{x}) \leqslant \varepsilon + \delta \quad \forall p = 1, 2, \ldots, m.$$

При этом алгоритм 4 работает не более

$$N = O\left(\frac{\max\limits_{1 \leqslant p \leqslant m}\left\{M_f^2, M_{g_p}^2\right\}}{\min\limits_{1 \leqslant p \leqslant m}\left\{\mu_f, \mu_{g_p}\right\}\varepsilon}\right)$$

итераций.

Доказательство. Для каждого продуктивного шага ($k \in I$) имеем

$$\eta_k^f \left(f(x_k) - f(x^*)\right) \leqslant \eta_k^f \left(\left\langle \nabla_\delta f(x_k),\, x_k - x^* \right\rangle - \mu_k V\left(x^*,\, x_k\right) + \delta\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\eta_k^f\right)^2 \|\nabla_\delta f(x_k)\|_2^2 + V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right) - \eta_k^f \mu_k V\left(x^*,\, x_k\right) + \eta_k^f \delta.$$

Отсюда после деления обеих частей приведенного выше неравенства на η_k^f получается

$$f(x_k) - f(x^*) \leqslant \frac{1}{2}\eta_k^f \|\nabla_\delta f(x_k)\|_2^2 + \frac{V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right)}{\eta_k^f} - \mu_k V\left(x^*,\, x_k\right) + \delta =$$
$$= \frac{1}{2\mu_{1:k}} \|\nabla_\delta f(x_k)\|_2^2 + \mu_{1:k} V\left(x^*,\, x_k\right) - \mu_k V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) + \delta \leqslant$$
$$\leqslant \frac{M_f^2}{2\mu_{1:k}} + \mu_{1:k-1} V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) + \delta. \tag{10}$$

Аналогично для каждого непродуктивного шага ($k \in J$) будем иметь

$$\eta_k^g \left(g_{p(k)}(x_k) - g_{p(k)}(x^*)\right) \leqslant \eta_k^g \left(\left\langle \nabla_\delta g_{p(k)}(x_k),\, x_k - x^* \right\rangle - \mu_k V\left(x^*,\, x_k\right) + \delta\right) \leqslant$$
$$\leqslant \frac{1}{2}\left(\eta_k^g\right)^2 \|\nabla_\delta g_{p(k)}(x_k)\|_2^2 + V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right) - \eta_k^g \mu_k V\left(x^*,\, x_k\right) + \eta_k^g \delta.$$

Разделив обе части последнего неравенства на η_k^g, получаем

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \leqslant \frac{1}{2}\eta_k^g \|\nabla_\delta g_{p(k)}(x_k)\|_2^2 + \frac{V\left(x^*,\, x_k\right) - V\left(x^*,\, x_{k+1}\right)}{\eta_k^g} - \mu_k V\left(x^*,\, x_k\right) + \delta \leqslant$$
$$\leqslant \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \mu_{1:k} V\left(x^*,\, x_k\right) - \mu_k V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) + \delta =$$
$$= \frac{M_{g_{p(k)}}^2}{2\mu_{1:k}} + \mu_{1:k-1} V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) + \delta. \tag{11}$$

Просуммируем неравенства (10), (11) по продуктивным и непродуктивным шагам, причем $M = \max\limits_{1 \leqslant p \leqslant m} \left\{M_f,\, M_{g_p}\right\}$. В итоге получим

$$\sum_{k \in I} \left(f(x_k) - f(x^*)\right) + \sum_{k \in J} \left(g_{p(k)}(x_k) - g(x^*)\right) \leqslant \sum_{k=1}^{N} \left(\frac{M^2}{2\mu_{1:k}} + \mu_{1:k-1} V\left(x^*,\, x_k\right) - \mu_{1:k} V\left(x^*,\, x_{k+1}\right) + \delta\right) \leqslant$$
$$\leqslant \sum_{k=1}^{N} \left(\frac{M^2}{2\mu_{1:k}} - \mu_{1:N} V\left(x^*,\, x_N\right) + \delta\right) \leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + N\delta.$$

Используя тот факт, что для непродуктивных шагов

$$g_{p(k)}(x_k) - g_{p(k)}(x^*) \geqslant g_{p(k)}(x_k) > \varepsilon + \delta,$$

получим следующую оценку:

$$\sum_{k \in I} \left(f(x_k) - f(x^*)\right) \leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + N\delta - \sum_{k \in J} \left(g_{p(k)}(x_t) - g_{p(k)}(x^*)\right) \leqslant$$
$$\leqslant \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + N\delta - \sum_{k \in J} (\varepsilon + \delta) = \sum_{k=1}^{N} \frac{M^2}{2\mu_{1:k}} + \delta|I| - \varepsilon|J|.$$

Поскольку

$$\sum_{k\in I}\Big(f(x_k)-f(x^*)\Big)\geqslant |I|\,(f(\widehat{x})-f(x^*)),$$

то

$$|I|\,(f(\widehat{x})-f(x^*))\leqslant\sum_{k\in I}\Big(f(x_k)-f(x^*)\Big)\leqslant\sum_{k=1}^{N}\frac{M^2}{2\mu_{1:k}}+\delta|I|-\varepsilon|J|.$$

После выполнения критерия остановки алгоритма 4 будут верны неравенства

$$f(\widehat{x})-f(x^*)\leqslant\varepsilon+\delta,\quad g_p(\widehat{x})\leqslant\varepsilon+\delta\quad\forall p=1,2,\ldots,m.$$

Оценка на количество итераций, достаточное для достижения заданной точности решения, аналогична оценке из доказательства теоремы 1. □

6. Результаты вычислительных экспериментов

В данном параграфе для иллюстрации эффективности алгоритма 1 описаны результаты вычислительных экспериментов для некоторых задач типа (1) по сравнению его работы с работой адаптивного алгоритма 3 (Mod.AMD-L), предложенного в [Стонякин и др., 2018]. Все эксперименты проводились на Python 3.4 на компьютере с Intel(R) Core(TM) i7-8550U CPU (1,80 ГГц, 4 ядра, 8 потоков). Оперативная память компьютера составляла 8 ГБ.

Пример 1. В этом примере рассматривается следующая целевая функция:

$$f(x)=\max_{1\leqslant i\leqslant T}\left\{f_i(x)=\langle a_i,x\rangle+\frac{\mu}{2}\|x\|_2^2\right\},\tag{12}$$

где $a_i\in\mathbb{R}^n$ и $\mu>0$.

Пример 2. В этом примере рассматривается следующая целевая функция:

$$f(x)=\frac{1}{T}\sum_{i=1}^{T}\|x-A_i\|_2+\frac{\mu}{2}\|x\|_2^2,\tag{13}$$

где A_i — некоторые точки в $\mathbb{R}^n$ и $\mu>0$.

Функциональные ограничения в обоих примерах 1 и 2 имеют следующий вид:

$$g_i(x)=\langle\alpha_i,x\rangle+\beta_i+\frac{\mu_i}{2}\|x\|_2^2,\quad i=1,2,\ldots,m,\tag{14}$$

где $\alpha_i\in\mathbb{R}^n$, $\beta_i\in\mathbb{R}$ и $\mu_i>0$ $\forall i=1,2,\ldots,m$.

Множество Q было выбрано в виде единичного шара в $\mathbb{R}^n$. На этом множестве функция f в примере 1 M_f-липшицева, где $M_f=\max\limits_{1\leqslant i\leqslant T}\{\|a_i\|_2+\mu\}$, и μ-сильно выпуклая. А в примере 2 функция f есть M_f-липшицева, где $M_f=1+\mu$, и μ-сильно выпуклая. Кроме того, функционалы g_i ($\forall i=1,\ldots,m$) Mg_i-липшицевы, где $Mg_i=\|\alpha_i\|_2+\mu_i$, и μ_i-сильно выпуклые.

Коэффициенты a_i, $\alpha_j\in\mathbb{R}^n$, точки A_i и константы $\beta_j\in\mathbb{R}$ $\forall i=1,\ldots,T$, $\forall j=1,\ldots,m$, в (12), (13) и (14) генерируются случайным образом с равномерным распределением на [0; 1). Параметры сильной выпуклости μ_i также выбираются случайно на интервале (0; 1).

Были проведены запуски алгоритмов 1 и 3 (Mod.AMD-L) из [Стонякин и др., 2018] при $\mu=$ $=1$, $n=500$, $m=50$, $T=10$ в (12) и $\mu=1$, $n=200$, $m=50$, $T=10$ в (13), при различных значениях $\varepsilon=\left\{\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{32},\frac{1}{64}\right\}$.

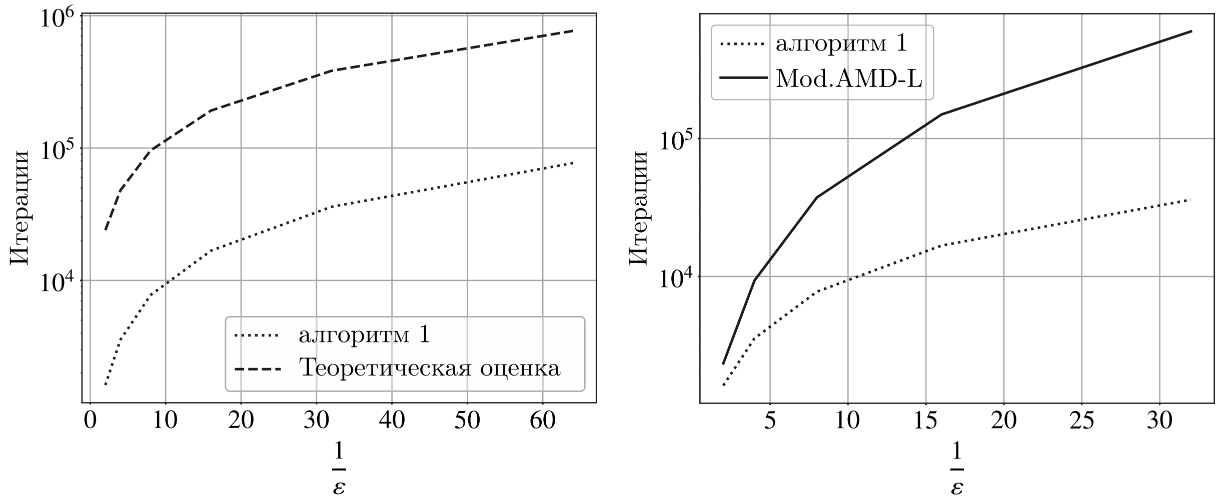

Рис. 1. Результаты алгоритма 1 и Mod.AMD-L (алгоритм 3 из [Стонякин и др., 2018]) для задачи (1) с (12) и (14), $n = 500$, $m = 50$, $T = 10$. На этих рисунках показаны динамика итераций алгоритма 1 и теоретическая оценка (2) (слева) и динамика итераций алгоритмов 1, Mod.AMD-L (справа)

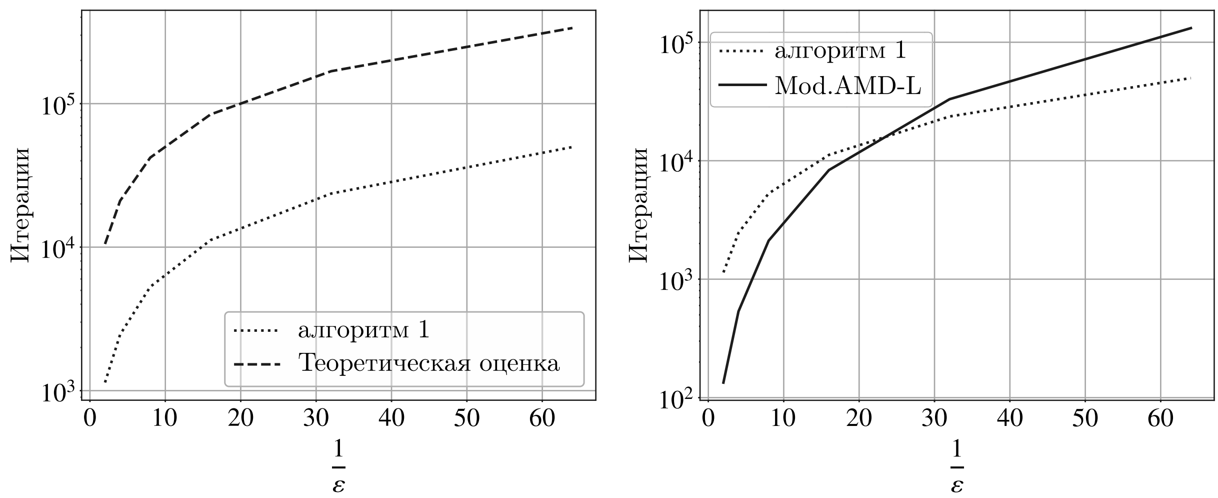

Рис. 2. Результаты алгоритма 1 и Mod.AMD-L (алгоритм 3 из [Стонякин и др., 2018]) для задачи (1) с (13) и (14), $n = 200$, $m = 50$, $T = 10$. На этих рисунках показаны динамика итераций алгоритма 1 и теоретическая оценка (2) (слева) и динамика итераций алгоритмов 1, Mod.AMD-L (справа)

Результаты работы этих алгоритмов представлены на рис. 1 и рис. 2. Эти результаты демонстрируют количество итераций, осуществляемых алгоритмом 1, включая теоретическую оценку количества итераций (2) для данного алгоритма (см. график слева на рис. 1). Также они демонстрируют сравнение количества итераций, осуществляемых алгоритмом 1 и Mod.AMD-L (см. график справа на рис. 1).

Исходя из результатов выполненных численных экспериментов, можно видеть, что алгоритм 1 работает быстрее, чем Mod.AMD-L для всех значений ε относительно примера 1, но что касается примера 2, то алгоритм 1 работает быстрее для достаточно малых значений ε. Кроме того, видно, что алгоритм 1 может достичь решения задачи с любой желаемой точностью за количество итераций, значительно меньшее, чем указано в теоретической оценке (2), доказанной в теореме 1.

7. Заключение

В данной статье исследованы субградиентные методы для задач сильно выпуклого программирования с несколькими функциональными ограничениями-неравенствами. Предложенные в работе методы основаны на объединении схемы с переключениями по продуктивным и непродуктивным шагам и недавно предложенных модификаций методов зеркального спуска для выпуклого программирования, позволяющих игнорировать часть функциональных ограничений на непродуктивных шагах алгоритма.

Предложена субградиентная схема с переключениями для задач сильно выпуклого программирования с несколькими ограничениями-неравенствами в случае, когда целевая функция и функциональные ограничения удовлетворяют условию Липшица. Исследовано обобщение описанного подхода на ситуацию с предположением о доступности на итерациях алгоритма δ-субградиентов целевой функции или функциональных ограничений, что потенциально может сэкономить вычислительные затраты метода за счет отказа от требования доступности точного значения субградиента в текущей точке. Также предложены методы типа зеркального спуска для задач сильно выпуклой минимизации с относительно липшицевыми и относительно сильно выпуклыми целевой функцией и ограничениями. Для всех рассматриваемых методов доказаны теоретические оценки качества выдаваемого решения, указывающие на оптимальность предлагаемых алгоритмов с точки зрения нижних оракульных оценок.

Ключевая идея предложенного подхода заключается в возможности использования в теоретических оценках параметров сильной выпуклости только тех функциональных ограничений, которые нарушаются на непродуктивных шагах алгоритма. Действительно, на каждом шаге метода учитывается параметр сильной выпуклости целевой функции (на продуктивных шагах) или первого нарушенного ограничения (на непродуктивных шагах). Это весьма существенно, поскольку неизвестно заранее, какое из ограничений будет нарушено в процессе работы метода. За счет этого и достигаются адаптивный подбор шага алгоритма и критерий остановки. Таким образом, предлагаемые модификации субградиентных методов позволяют сэкономить время работы алгоритма за счет рассмотрения не всех функциональных ограничений на непродуктивных шагах.

Результаты проведенных численных экспериментов иллюстрируют преимущество предлагаемого нами подхода в сравнении с адаптивным алгоритмом 3 (Mod.AMD-L) из [Стонякин и др., 2018]. Также, исходя из этих результатов, можно видеть, что алгоритм 1 может достичь решения задачи с любой желаемой точностью за значительно меньшее количество итераций, чем указано в теоретических оценках.

Список литературы (References)

Поляк Б. Т. Введение в оптимизацию. — М.: Наука, 1983.

Polyak B. T. Vvedenie v optimizatsiyu [Introduction to optimization]. — Moscow: Nauka, 1983 (in Russian).

Поляк Б. Т. Один общий метод решения экстремальных задач // Докл. АН СССР. — 1967. — Т. 174, № 1. — С. 33–36.

Polyak B. T. Odin obshchii metod resheniya ehkstremal'nykh zadach [A general method of solving extremum problems] // Dokl. AN SSSR. — 1967. — Vol. 174, No. 1. — P. 33–36 (in Russian).

Стонякин Ф. С., Алкуса М. С., Степанов А. Н., Баринов М. А. Адаптивные алгоритмы зеркального спуска в задачах выпуклого программирования с липшицевыми ограничениями // Труды института математики и механики УрО РАН. — 2018. — Т. 24, № 2. — С. 266–279.

Stonyakin F. S., Alkousa M. S., Stepanov A. N., Barinov M. A. Adaptivnye algoritmy zerkal'nogo spuska v zadachakh vypuklogo programmirovaniya s lipshitsevymi ogranicheniyami [Adaptive mirror descent algorithms in convex programming problems with Lipschitz constraints] // Trudy instituta matematiki i mekhaniki URO RAN. — 2018. — Vol. 24, No. 2. — P. 266–279 (in Russian).

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

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

Ben-Tal A., Nemirovski A. Robust truss topology design via semidefinite programming // SIAM J. Optim. — 1997. — Vol. 7, No. 4. — P. 991–1016.

Dvurechensky P., Gasnikov A., Nurminsky E., Stonyakin F. Advances in low-memory subgradient optimization // Numerical nonsmooth optimization. State of the art algorithms / A. M. Bagirov et al. (eds.). — Springer Nature Switzerland AG, 2020. — P. 19–59. — https://arxiv.org/abs/1902.01572v1

Lu H. Relative continuity for non-Lipschitz nonsmooth convex optimization using stochastic (or deterministic) mirror descent // INFORMS Journal on Optimization. — 2019. — Vol. 1, No. 4. — P. 288–303.

Lu H., Freund R., Nesterov Yu. Relatively smooth convex optimization by first-order methods, and applications // SIAM Journal on Optimization. — 2018. — Vol. 28, No. 1. — P. 333–354.

Nemirovsky A. Efficient methods for large-scale convex optimization problems // Ekonomika i matematicheskie metody. — 1979. — Vol. 15 (in Russian).

Nemirovsky A., Yudin D. Problem complexity and method efficiency in optimization. — New York: J. Wiley & Sons, 1983.

Nesterov Yu. Relative smoothness: new paradigm in convex optimization // Conference report, EUSIPCO-2019, A Coruna, Spain. — 2019. — Vol. 4.

Savchuk O. S., Stonyakin F. S., Alkousa M. S., Zabirova R. R., Titov A. A., Gasnikov A. V. Online optimization problems with functional constraints under relative Lipschitz continuity and relative strong convexity conditions // Mathematical Optimization Theory and Operations Research: Recent Trends. — 2023. — P. 29–43.

Shor N. Z. Generalized gradient descent with application to block programming // Kibernetika. — 1967. — Vol. 3, No. 3. — P. 53–55.

Shpirko S., Nesterov Yu. Primal-dual subgradient methods for huge-scale linear conic problem // SIAM J. Optim. — 2014. — Vol. 24, No. 3. — P. 1444–1457.