ТРУДЫ ИНСТИТУТА МАТЕМАТИКИ И МЕХАНИКИ УрО РАН

Том 27 № 4 2021

УДК 519.85

# АДАПТИВНЫЕ МЕТОДЫ ГРАДИЕНТНОГО ТИПА ДЛЯ ЗАДАЧ ОПТИМИЗАЦИИ С ОТНОСИТЕЛЬНОЙ ТОЧНОСТЬЮ И ОСТРЫМ МИНИМУМОМ[1]

**Ф. С. Стонякин, С. С. Аблаев, И. В. Баран**

В статье исследован универсальный метод градиентного типа для задач выпуклой минимизации с относительной точностью, а также получены новые результаты о линейной скорости сходимости специальных вариантов субградиентного метода для задач с острым минимумом и некоторыми обобщениями выпуклости. Предложен новый подход к выбору параметров и правилу остановки адаптивного варианта метода подобных треугольников на классе задач минимизации выпуклых положительно однородных функционалов. Это позволило получить аналог универсального градиентного метода для задач с относительной точностью и доказать оптимальную оценку его скорости сходимости на выделенном классе задач. Приведен пример результатов численных экспериментов, иллюстрирующих возможность повышения качества выдаваемого методом приближенного решения по сравнению с теоретической оценкой за счет введенного адаптивного критерия остановки. Рассмотрен вариант субградиентного метода для задач минимизации слабо $\beta$-квазивыпуклых липшицевых функционалов в случае острого минимума, и доказана линейная скорость сходимости. Наконец, предложен вариант субградиентного метода, который сходится с линейной скоростью на классе задач минимизации квазивыпуклых гельдеровых функционалов с острым минимумом. Показана применимость этого результата для функционалов со слабым острым минимумом (гельдеровым ростом), и сформулировано следствие для задач с относительной точностью.

Ключевые слова: относительная точность, выпуклый положительно однородный функционал, слабо $\beta$-квазивыпуклый функционал, субградиентный метод, липшицев функционал, квазивыпуклый функционал.

**F. S. Stonyakin, S. S. Ablaev, I. V. Baran. Adaptive gradient-type methods for optimization problems with relative error and sharp minimum.**

A universal method of gradient type for convex minimization problems with relative error is studied, and new results on the linear convergence rate of specific versions of the subgradient method for problems with a sharp minimum and some generalizations of convexity are obtained. A new approach to the choice of parameters and the stopping rule of an adaptive version of the method of similar triangles on a class of minimization problems for convex positively homogeneous functionals is proposed. As a consequence, an analog of the universal gradient method for problems with relative error is obtained and an optimal estimate of its convergence rate on a chosen class of problems is proved. An example of the results of numerical experiments illustrating the possibility of improving the quality of the approximate solution produced by the method as compared to a theoretical estimate due to the introduced adaptive stopping rule is given. A version of the subgradient method for minimization problems with weakly $\beta$-quasiconvex Lipschitz-continuous functionals in the case of a sharp minimum is considered, and a linear convergence rate is proved. Finally, a version of the subgradient method is proposed that converges at a linear rate on a class of problems of minimizing quasiconvex Holder-continuous functionals with a sharp minimum. The applicability of this result for functionals with a weak sharp minimum (Holderian growth) is shown and a corollary of this result is formulated for problems with relative error.

Keywords: relative error, convex positively homogeneous functional, weak $\beta$-quasiconvex functional, subgradient method, Lipschitz-continuous functional, quasiconvex functional.

**MSC:** 90C25, 90C06, 49J52

**DOI:** 10.21538/0134-4889-2021-27-4-175-188

## 1. Введение

Предлагаемая работа посвящена исследованию некоторых методов градиентного типа для задач выпуклой минимизации с относительной точностью, а также субградиентных методов

[1]Исследования разд. 2 и 3 выполнены при поддержке гранта Президента Российской Федерации для государственной поддержки молодых российских ученых — кандидатов наук, код МК-15.2020.1. Результаты разд. 4 получены Ф. С. Стонякиным при поддержке гранта Российского научного фонда, код 21-71-30005.

для задач с острым минимумом. Первый рассматриваемый в настоящей работе класс задач имеет вид

$$f(x) \to \min_{x \in Q}, \tag{1.1}$$

где $Q$ — выпуклое замкнутое подмножество в $\mathbb{R}^n$, $f$ — выпуклая положительно однородная функция степени 1. Такая постановка задачи возникла в работах Ю. Е. Нестерова [3; 15; 16]. Будем говорить, что значение $f(\widehat{x})$, вычисленное в точке $\widehat{x}$ допустимого множества задачи $Q$, приближает минимальное значение $f^* > 0$ с относительной точностью $\gamma > 0$, если

$$f(\widehat{x}) \leq (1+\gamma) f^*.$$

Отметим, что в работах [3; 15; 16] обоснована целесообразность рассмотрения использования относительной точности для понимания качества приближенного решения некоторых прикладных задач (например, проектирование механических конструкций), а также адекватность предположения положительной однородности $f$. Вслед за упомянутыми выше работами Ю. Е. Нестерова (см., например, [3, гл. 6]) для задач с относительной точностью (1.1) сделаем следующее предположение:

$$\operatorname{dom} f \equiv \mathbb{R}^n, \quad 0 \in \operatorname{int} \partial f(0). \tag{1.2}$$

Иными словами, $f$ — опорная функция некоторого выпуклого компактного множества, у которого 0 — внутренняя точка. Как показано в работах [3;15;16], такое предположение естественно, и оно покрывает широкий класс задач, хотя в общем случае и нет возможности гарантировать, что $f$ удовлетворяет (1.2).

В разд. 2 описан аналог ускоренного универсального градиентного метода для задач выпуклой положительно однородной минимизации и получена оценка количества итераций (2.5), после которых гарантированно выполняется предложенный адаптивный критерий остановки. Выполнение этого критерия остановки заведомо обеспечивает достижение требуемой относительной точности для точки выхода алгоритма по целевому функционалу. Доказанная оценка (2.5) сублинейна, но за счет адаптивного критерия остановки потенциально возможно наблюдать более высокую скорость сходимости, что показано с помощью экспериментов в примере 1.

Однако естественный интерес представляет вопрос о возможности гарантировать линейную скорость сходимости для негладких задач, пусть и с дополнительными предположениями. Поэтому далее в разд. 3 и 4 были исследованы субградиентные методы для негладких задач с теоретическими результатами о линейной скорости сходимости. Хорошо известно, что это возможно при дополнительных предположениях о том, что минимум острый и доступно оптимальное значение $f^*$. В работе [6] при таких условиях предложен способ выбора шага субградиентного метода, который гарантирует линейную скорость сходимости по аргументу. Отметим, что условие острого минимума верно, например, для задачи проектирования точки на выпуклый компакт или задачи о нахождении общей точки системы множеств. В настоящей работе предлагаются варианты субградиентного метода, один из которых гарантирует линейную скорость сходимости на классе слабо $\beta$-квазивыпуклых липшицевых функционалов ($\beta \in (0;1]$), а другой — на классе обычных квазивыпуклых гельдеровых функционалов с острым минимумом. Напомним, что слабая $\beta$-квазивыпуклость введена несколько лет назад в [9]. Это понятие обобщает условие выпуклости. Например, невыпуклая функция одной переменной $f(x) = |x|(1 - e^{-|x|})$ слабо 1-квазивыпукла.

Опишем кратко структуру и новые результаты (вклад) статьи. В разд. 2 рассмотрен класс задач минимизации выпуклых положительно однородных функций с гельдеровым градиентом (субградиентом при $\nu = 0$). Предложен специальный подход к выбору параметров и критерию остановки адаптивного метода подобных треугольников на классе задач минимизации выпуклых положительно однородных функций. Это позволило обосновать применимость соответствующего аналога универсального градиентного метода для задач с относительной точностью и доказать оптимальную оценку его скорости сходимости на выделенном классе задач (теорема 3). Приведен пример результатов численных экспериментов (пример 1), иллюстрирующих возможность повышения скорости сходимости за счет предлагаемого адаптивного правила остановки (2.4). При этом сублинейная теоретическая оценка скорости сходимости алгоритма 1 побудила нас исследовать также и субградиентные методы в случаях, когда можно гарантировать линейную скорость сходимости. В разд. 3 представлен вариант субградиентного метода для задач минимизации слабо $\beta$-квазивыпуклых функционалов в случае острого минимума и доказан результат о линейной скорости сходимости для липшицевых функций указанного типа (теорема 4 и следствие 1). В разд. 4 описан вариант субградиентного метода, который сходится с линейной скоростью на классе задач минимизации квазивыпуклых гельдеровых функционалов с острым минимумом. Рассмотрено обобщение этого результата на класс задач со слабым острым минимумом (пример 3). Сформулированы следствия из теорем 4 и 5 для задач с относительной точностью на классе выпуклых положительно однородных функционалов степени 1 (следствия 2 и 3).

Обозначим через $\|\cdot\|_2$ стандартную евклидову норму в пространстве $\mathbb{R}^n$. Напомним следующий вспомогательный результат.

**Теорема 1** [3, теорема 6.1.1]. *Пусть для некоторого* $\gamma_0 > 0$ *верно*

$$f(x) \geq \gamma_0 \, \|x\|_2 \quad \forall \, x \in Q. \tag{1.3}$$

*Тогда для всякого точного решения* $x_*$ *задачи* (1.1) *и* $f^* = f(x_*)$ *верно неравенство*

$$\|x_0 - x_*\|_2 \leq \frac{2}{\gamma_0} \, f^* \leq \frac{2}{\gamma_0} \, f(x_0).$$

Заметим, что предположение (1.3) теоремы 1 есть следствие условия (1.2). Для результатов (теорема 3, следствия 2 и 3) о задачах с относительной точностью по аналогии с [3] условимся в качестве начальной точки $x_0$ использовать евклидову проекцию начала координат на допустимое множество задачи $Q$: $\|x_0\|_2 := \min\limits_{x \in Q}\{\|x\|_2\}$.

## 2. Универсальный метод для задач минимизации выпуклого положительно однородного функционала с относительной точностью

Предположим, что для градиента целевого функционала $f$ задачи (1.1) выполняется условие Гельдера, т. е. существует $\nu \in [0, 1]$ такое, что при некотором $L_\nu \in (0; +\infty)$ верно неравенство

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L_\nu \|x - y\|_2^\nu \quad \forall \, x, y \in Q. \tag{2.1}$$

Если $\nu = 0$, то под $\nabla f(x)$ в (2.1) можно понимать произвольный субградиент $f$ в точке $x$. При указанном допущении (2.1), как известно (см. [1; 17]), для произвольных $x, y \in Q$ верно неравенство

$$0 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{L(\delta)}{2} \, \|x - y\|_2 + \delta \tag{2.2}$$

при некотором фиксированном $\delta > 0$ и

$$L(\delta) = L_\nu \Big[ \frac{L_\nu}{2 \, \delta} \, \frac{1 - \nu}{1 + \nu} \Big]^{\frac{1-\nu}{1+\nu}}.$$

Опишем универсальный вариант градиентного метода, который работает на выделенном классе задач (1.1) с реализацией в ходе работы адаптивной настройки оценки качества выдаваемого решения с относительной точностью на параметр гладкости $L(\delta)$. Идея универсальных методов состоит в том, что используемый в оценке качества решения потенциально большой параметр гладкости $L = L(\delta)$ заменяется адаптивно подбираемым на итерациях множеством значений его локальных аналогов $L_{k+1}$ $(k \geq 0)$, что может повышать качество выдаваемого решения по сравнению с теоретической оценкой, содержащий глобальное значение $L$ (см. [17]). Мы отправляемся от следующего адаптивного варианта метода подобных треугольников (см. [7]). Сформулируем его вариант для задачи минимизации выпуклой функции $f$ с условием обобщенной гладкости (2.2).

А л г о р и т м 1: Адаптивный вариант метода подобных треугольников для задач с условием обобщенной гладкости.

**Require:** $x_0$ — начальная точка, $N$ — количество шагов, $\{\delta_{k+1}\}_{k=0}^{N-1}$ — последовательность положительных чисел и $L_0 > 0$;

1: **0-шаг:** $y_0 := x_0,\ u_0 := x_0,\ L_1 := L_0/2,\ \alpha_0 := 0,\ A_0 := \alpha_0$;
2: **for** $k = 1, \ldots$ **do**
3: Найти наибольший корень $\alpha_{k+1}$ квадратного уравнения

$$A_k + \alpha_{k+1} = L_{k+1}\alpha_{k+1}^2, \quad A_{k+1} := A_k + \alpha_{k+1}, \quad y_{k+1} := \frac{\alpha_{k+1}u_k + A_k x_k}{A_{k+1}};$$

$$u_{k+1} := \arg\min_{x \in Q}\left\{\frac{1}{2}\|x - u_k\|_2^2 + \alpha_{k+1}\langle \nabla f(y_{k+1}), x - y_{k+1}\rangle\right\}; \quad x_{k+1} := \frac{\alpha_{k+1}u_{k+1} + A_k x_k}{A_{k+1}};$$

4: **if** $f(x_{k+1}) \leq f(y_{k+1}) + \langle \nabla f(y_{k+1}), x_{k+1} - y_{k+1}\rangle + \frac{L_{k+1}}{2}\|x_{k+1} - y_{k+1}\|_2^2 + \delta_{k+1}$ **then**
5: $L_{k+2} := L_{k+1}/2$ и перейти к следующему шагу.
6: **else**
7: $L_{k+1} := 2L_{k+1}$ и повторить текущий шаг.
8: **end if**
9: **end for**

Для алгоритма 1 имеет место следующий результат, вытекающий из доказательства [7, теоремы 2], если рассматривать уже предположения настоящей статьи для алгоритма 1: евклидову прокс-структуру ($V(x,y) = 0.5\|x-y\|_2^2$), функцию модели $\psi_\delta(x,y) = \langle \nabla f(y), x - y\rangle$, а также рассматривать предположение о точном решении вспомогательных подзадач на итерациях алгоритма 1 согласно п. 3 листинга, т. е. $\widetilde{\delta} = 0$ (обозначения расхождения Брэгмана $V(x,y)$, функции-модели $\psi_\delta(x,y)$ и погрешности решения вспомогательных подзадач $\widetilde{\delta}$ соответствуют обозначениям [7]).

**Теорема 2.** *Пусть $\|x_* - x_0\|_2 \leq R\sqrt{2}$ для некоторой постоянной $R > 0$, где $x_0$ — начальная точка, $x_*$ — ближайшая точка минимума $f$ к $x_0$ по евклидовой норме. Тогда для алгоритма* 1 *выполнено следующее неравенство*:

$$f(x_N) - f(x_*) \leq \frac{R^2}{A_N} + \frac{2\sum\limits_{k=0}^{N-1}\delta_{k+1}A_{k+1}}{A_N}, \tag{2.3}$$

*где* $A_N = \sum\limits_{k=0}^{N-1}\frac{1}{L_{k+1}}$.

Для получения оценки скорости сходимости алгоритма 1 для задач оптимизации с относительной точностью выберем критерий остановки вида

$$A_N \geq \frac{R}{\varepsilon}. \tag{2.4}$$

Покажем, как на базе предыдущей теоремы можно выбрать параметры алгоритма 1, чтобы получить аналог универсального градиентного метода для задач с относительной точностью вида (1.1).

**Теорема 3.** *Пусть $f$ — положительная однородная выпуклая функция, причем $f(x) \geq \gamma_0\|x\|_2$ для всякого $x \in Q$, а также $\|x_* - x_0\|_2 \leq R\sqrt{2} \leq \|x_* - x_0\|_2\sqrt{2}$ для некоторой постоянной $R > 0$, где $x_0\colon \|x_0\|_2 = \min\{\|x\|_2 \mid x \in Q\}$ — начальная точка, $x_*$ — ближайшая точка минимума $f$ к $x_0$ по евклидовой норме. Тогда после выполнения критерия остановки* (2.4) *в случае $\varepsilon = \dfrac{\gamma\gamma_0}{3}$ и $\delta_{k+1} \leq R\varepsilon\dfrac{\alpha_{k+1}}{4A_{k+1}}$ для всякого $k \geq 0$ будет гарантированно достигнута $\gamma$-относительная точность приближенного решения $x_N\colon f(x_N) \leq (1+\gamma)f^*$. При этом критерий остановки* (2.4) *будет выполнен не более, чем после*

$$N \geq \left(R\varepsilon^{-\frac{2}{1+\nu}}\, 2^{\frac{2+4\nu}{1+\nu}}\, L_\nu^{\frac{2}{1+\nu}}\right)^{\frac{1+\nu}{1+3\nu}} \tag{2.5}$$

*итераций алгоритма* 1.

Доказательство. Выбор в алгоритме 1 параметров $\delta_{k+1} \leq R\varepsilon\dfrac{\alpha_{k+1}}{4A_{k+1}}$ $\forall k \geq 0$ означает, что из теоремы 2 вытекает неравенство

$$f(x_N) - f(x_*) \leq \frac{R^2}{A_N} + \frac{2\sum\limits_{k=0}^{N-1} A_{k+1}\delta_{k+1}}{A_N}.$$

Далее, имеет место неравенство (см. [7, п. 4.3; 17]) $A_N \geq \dfrac{N^{\frac{1+3\nu}{1+\nu}}\varepsilon^{\frac{1-\nu}{1+\nu}}}{2^{\frac{2+4\nu}{1+\nu}}L_\nu^{\frac{2}{1+\nu}}}$. Критерий остановки (2.4) алгоритма 1 заведомо выполнен, если $\dfrac{N^{\frac{1+3\nu}{1+\nu}}\varepsilon^{\frac{1-\nu}{1+\nu}}}{2^{\frac{2+4\nu}{1+\nu}}L_\nu^{\frac{2}{1+\nu}}} \geq \dfrac{R}{\varepsilon}$. Следовательно,

$$N \geq \varepsilon^{-\frac{2}{1+3\nu}}\left(R\, 2^{\frac{2+4\nu}{1+\nu}}\, L_\nu^{\frac{2}{1+\nu}}\right)^{\frac{1+\nu}{1+3\nu}}.$$

Пусть $\varepsilon = \dfrac{\gamma\gamma_0}{3}$. Тогда с учетом сделанного предположения о выборе параметров $\delta_{k+1}$ после выполнения критерия остановки (2.4) алгоритма 1 имеем неравенство $f(x_N) - f^* \leq \dfrac{3\varepsilon R}{2}$. По теореме 1 верно $R \leq \dfrac{2}{\gamma_0}f^*$, откуда имеем неравенство

$$f(x_N) \leq f^* + \frac{3\,\varepsilon}{\gamma_0}f^*.$$

Наконец, ввиду $\varepsilon = \dfrac{\gamma\gamma_0}{3}$ получаем $f(x_N) \leq (1+\gamma)f^*$, что и требовалось. □

Доказанная в теореме 2 оценка скорости сходимости алгоритма 1, как известно (см. [2]), оптимальна на выделенном классе задач минимизации выпуклых функционалов с гельдеровым градиентом с точностью до умножения на постоянную величину. Однако за счет универсальности метода (адаптивной настройки на параметр и уровень гладкости задачи $\nu \in [0;1]$ в процессе работы) возможно ожидать на практике более высокую скорость повышения качества выдаваемого решения с ростом количества итераций. Для иллюстрации этого факта приведем пример с результатами некоторых вычислительных экспериментов.

Пример 1. $f(x) = \|x\|_2$, $x \in \mathbb{R}^n$, $n = 10^6$, $Q = B_1(2p)$ — евклидов шар единичного радиуса с центром в точке $2p$, где $p = \dfrac{(1,\ldots,1)}{\|(1,\ldots,1)\|_2} \in \mathbb{R}^{10^6}$. Очевидно, что в таком случае $f^* > 0$ и $\gamma_0 = 1$. Отметим, что полученная в теореме 3 оценка скорости сходимости (2.5) сублинейна. В лучшем случае ($\nu = 1$) оценка (2.5) может гарантировать оценку сложности $O(\sqrt{\gamma^{-1}})$, которая неулучшаема на классе задач минимизации выпуклых функционалов с липшицевым градиентом. Однако, как видно из таблицы ниже, адаптивный подбор параметров $L_{k+1}$ ($k = 0,1,\ldots$) приводит к повышению скорости сходимости.

**Скорость изменения качества решения по мере роста числа итераций**

| Кол-во итераций | Теор. оценка качества решения согл. (2.3) | Время работы (с) |
|---|---|---|
| 10 | 9.996e-05 | 0:00:02.91 |
| 15 | 3.061e-06 | 0:00:04.46 |
| 20 | 9.540e-08 | 0:00:05.91 |
| 25 | 2.980e-09 | 0:00:07.37 |
| 30 | 9.313e-11 | 0:00:08.84 |
| 35 | 2.910e-12 | 0:00:10.33 |
| 40 | 9.094e-14 | 0:00:11.84 |
| 45 | 2.842e-15 | 0:00:18.93 |
| 50 | 8.881e-17 | 0:00:20.38 |
| 55 | 2.775e-18 | 0:00:22.27 |
| 60 | 8.673e-20 | 0:00:23.78 |
| 65 | 2.710e-21 | 0:00:25.31 |
| 70 | 8.470e-23 | 0:00:26.90 |
| 75 | 2.646e-24 | 0:00:28.39 |
| 80 | 8.271e-26 | 0:00:29.98 |
| 85 | 2.584e-27 | 0:00:31.49 |
| 90 | 8.077e-29 | 0:00:32.99 |
| 95 | 2.524e-30 | 0:00:34.47 |
| 100 | 7.888e-32 | 0:00:35.90 |

Вычисления были произведены с помощью Python 3.4 на компьютере с Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 1992 Mhz, 4 Core(s). ОЗУ компьютера составляла 8 Гб.

Прокомментируем совместимость предположения о гельдеровости градиента (2.1) целевой функции и естественного для задач с относительной точностью требования положительной однородности $f$ (см. [3;15;16]). Во-первых, если 0 не лежит в допустимом множестве задачи $Q$, то на $Q$ целевая функция $f$ может иметь как гельдеров, так и липшицев градиент, что верно, к примеру, для задачи из примера 1. Более того, сама постановка задачи с относительной точностью подразумевает, что $f^* > 0$, и поэтому 0 не должен лежать в допустимом множестве задачи. Во-вторых, липшицева функция $f$ удовлетворяет (2.1) при $\nu = 0$, а адаптивность алгоритма 1 потенциально позволяет улучшить оценку скорости сходимости по сравнению с оптимальной при $\nu = 0$ оценкой вида $O(\gamma^{-2})$, что как раз и показывают результаты экспериментов из примера 1.

## 3. Адаптивный субградиентный метод для минимизации слабо $\beta$-квазивыпуклых функционалов с острым минимумом

Сублинейность оценки скорости сходимости в теореме 3 приводит к идее исследовать подклассы задач, для которых все же возможна линейная скорость сходимости субградиентного метода. Это, в частности, возможно в случае предположений о том, что минимум острый и доступно $f^*$ (см. [6]). Поэтому данный и следующий разделы работы посвящены исследованию методов с линейной скоростью сходимости для задач с острым минимумом на классах задач с обобщениями выпуклости (слабая $\beta$-квазивыпуклость и обычная квазивыпуклость). Отметим, что значение $f^*$ бывает известно в геометрических задачах (проекция точки на множество или нахождение общей точки системы множеств). Также можно упомянуть задачу типа $f(x) = \|Ax - b\|_2 \to \min$, где $A$ — матрица размера $n \times n$, $b \in \mathbb{R}^n$. Она разрешима, если существует $x_*$ такое, что $Ax_* = b$. Будем рассматривать задачи вида

$$f(x) \to \min_{x \in Q}, \tag{3.1}$$

где $Q$ — выпуклое замкнутое подмножество $\mathbb{R}^n$, $f^* = f(x_*) = \min\limits_{x \in Q} f(x)$, $f$ — слабо $\beta$-квазивыпуклый функционал при некотором $\beta \in (0;1]$. Напомним (см. [9]), что $f$ называется *слабо*

*$\beta$-квазивыпуклым* относительно точки минимума $x_*$ задачи (3.1) на множестве $Q$, если для произвольного $x \in Q$ выполнено неравенство

$$f(x_*) \geqslant f(x) + \frac{1}{\beta}\langle \nabla f(x), x_* - x\rangle, \tag{3.2}$$

где $\nabla f(x)$ — произвольный субградиент $f$ в точке $x$. Под субградиентом здесь и всюду далее понимаем элемент субдифференциала Кларка $f$ в точке $x$ и предполагаем его существование. Если $f$ дифференцируем в точке $x$, то под $\nabla f$ понимаем обычный градиент. Это вполне естественно для липшицевых функционалов (для существования субдифференциала Кларка в точке достаточно локальной липшицевости $f$ в окрестности этой точки). Если функционал $f$ выпуклый, то субдифференциал Кларка совпадает с обычным субдифференциалом в смысле выпуклого анализа, и в таком случае условие слабой $\beta$-квазивыпуклости (3.2) верно при $\beta = 1$.

Ясно, что если неравенство (3.2) верно для некоторого $\beta = \beta_0 \in (0; 1]$, то оно верно и при $\beta \in (0; \beta_0]$. Примеры функционалов, для которых возможно проверить свойство слабой $\beta$-квазивыпуклости и оценить параметр $\beta$, приведены в [10]. В частности, это верно для невыпуклой функции $f(x) = |x|(1 - e^{-|x|})$ при $\beta = 1$ (см. [9]). Отметим также, что субградиент $f$ может быть нулевым только в точке минимума: равенство $\nabla f(x) = 0$ влечет $f(x) \leqslant f(x_*)$, что автоматически означает $f(x) = f^*$.

Предположим, что также верно условие острого минимума для некоторого $\alpha > 0$

$$f(x) - f^* \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2, \tag{3.3}$$

где $X_*$ — множество точек минимума функции $f$ на множестве $Q$. В частности, условие (3.3) верно для задачи евклидова проектирования точки $x$ на выпуклый компакт $X_* \subset Q$, причем $f^* = 0$. Всюду далее будем понимать под $\mathrm{Pr}_Q$ оператор евклидова проектирования на множество $Q$.

Предложим следующий вариант субградиентного метода с шагом Б. Т. Поляка (см. [6]) для задачи минимизации слабо $\beta$-квазивыпуклого функционала:

$$x_{k+1} = \mathrm{Pr}_Q\{x_k - h_k \nabla f(x_k)\}, \tag{3.4}$$

где при всяком $k \geqslant 0$ верно $\|\nabla f(x_k)\|_2 \neq 0$ (иначе решение уже найдено), а также

$$h_k = \frac{\beta(f(x_k) - f^*)}{\|\nabla f(x_k)\|_2^2}.$$

Справедлива следующая теорема.

**Теорема 4.** *Пусть $f$ — слабо $\beta$-квазивыпуклая функция и для задачи минимизации $f$* (3.1) *с острым минимумом используется алгоритм* (3.4) *с шагом $h_k = \dfrac{\beta(f(x_k) - f^*)}{\|\nabla f(x_k)\|_2^2}$. Тогда после $k$ итераций алгоритма* (3.4) *верно неравенство*

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq \prod_{i=0}^{k}\Big(1 - \frac{\alpha^2\beta^2}{\|\nabla f(x_i)\|_2^2}\Big) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2.$$

Доказательство. Как известно (см. [6, теорема 1, соотношение (3)]), для всякого $x_* \in X_*$ верны неравенства

$$2\beta h_k(f(x_k) - f(x_*)) \leq 2h_k\langle \nabla f(x_k), x_k - x_*\rangle \leq h_k^2\|\nabla f(x_k)\|_2^2 + \min_{x_* \in X_*} \|x_k - x_*\|_2^2 - \min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2.$$

Поэтому

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq h_k^2\|\nabla f(x_k)\|_2^2 - 2\,\beta^2 h_k(f(x_k) - f(x_*)) + \min_{x_* \in X_*} \|x_k - x_*\|_2^2$$

$$= \frac{\beta^2(f(x_k)-f(x_*))^2}{\|\nabla f(x_k)\|_2^2} - \frac{2\,\beta^2(f(x_k)-f(x_*))^2}{\|\nabla f(x_k)\|_2^2} + \min_{x_*\in X_*}\|x_k - x_*\|_2^2$$

$$= -\frac{\beta^2(f(x_k)-f(x_*))^2}{\|\nabla f(x_k)\|_2^2} + \min_{x_*\in X_*}\|x_k - x_*\|_2^2.$$

Согласно условию острого минимума имеем

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le -\frac{\alpha^2\beta^2}{\|\nabla f(x_k)\|_2^2}\min_{x_*\in X_*}\|x_k-x_*\|_2^2 + \min_{x_*\in X_*}\|x_k-x_*\|_2^2$$

$$= \left(1 - \frac{\alpha^2\beta^2}{\|\nabla f(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k-x_*\|_2^2.$$

Далее, получаем цепочку неравенств

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le \left(1 - \frac{\alpha^2\beta^2}{\|\nabla f(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k-x_*\|_2^2$$

$$\le \left(1-\frac{\alpha^2\beta^2}{\|\nabla f(x_k)\|_2^2}\right)\left(1-\frac{\alpha^2\beta^2}{\|\nabla f(x_{k-1})\|_2^2}\right)\min_{x_*\in X_*}\|x_k-x_*\|_2^2 \le \ldots \le \prod_{i=0}^{k}\left(1-\frac{\alpha^2\beta^2}{\|\nabla f(x_i)\|_2^2}\right)\min_{x_*\in X_*}\|x_0-x_*\|_2^2.$$

□

**Следствие 1.** *Если в условиях теоремы* 4 *допустить, что* $f$ *удовлетворяет условию Липшица с константой* $M > 0$, *то можно утверждать сходимость алгоритма* (3.4) *со скоростью геометрической прогрессии*

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le \left(1 - \frac{\alpha^2\beta^2}{M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2.$$

Отметим, что теорема 4 применима к функционалам, не удовлетворяющим условию Липшица. Это возможно ввиду адаптивности оценки скорости сходимости из теоремы 4 (использование норм субградиентов $\|\nabla f(x_i)\|_2^2$ вместо $M^2$). Но если нет уверенности в выполнении условия Липшица, то в данной общей ситуации невозможно гарантировать линейную скорость сходимости.

Покажем, как можно получить оценку скорости сходимости алгоритма (3.4) для достижения $\gamma$-относительной точности по целевому функционалу $f(x_{k+1}) \le (1+\gamma)f^*$ при $\gamma > 0$. Следуя работам [3; 15; 16], будем рассматривать задачу с относительной точностью на классе выпуклых ($\beta = 1$) положительно однородных функционалов $f\colon Q \to \mathbb{R}\colon f^* > 0$, а также с условием $f(x) \ge \gamma_0\|x\|_2$ для всякого $x \in Q$ при некотором фиксированном $\gamma_0 > 0$. Начальная точка $x_0$ рассматриваемого алгоритма (3.4) выбирается так, чтобы $\|x_0\|_2 := \min_{x\in Q}\{\|x\|_2\}$.

По следствию 1 имеем неравенства

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\| \le \left(1 - \frac{\alpha^2}{M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2.$$

С учетом теоремы 1 получаем

$$\|x_{k+1}-x_*\| \le \frac{2\,f^*}{\gamma_0}\sqrt{\left(1-\frac{\alpha^2}{M^2}\right)^{k+1}} \le \gamma f^*, \quad f(x_{k+1}) - f^* \le \frac{2\,Mf^*}{\gamma_0}\sqrt{\left(1-\frac{\alpha^2}{M^2}\right)^{k+1}} \le \gamma f^*,$$

откуда $f(x_{k+1}) \le f^* + \gamma f^* = (1+\gamma)f^*$. Потребуем, чтобы $2\,M\sqrt{\left(1-\frac{\alpha^2}{M^2}\right)^{k+1}} \le \gamma\,\gamma_0$. Это означает, что

$$k \ge \log_{\left(1-\frac{\alpha^2}{M^2}\right)}\frac{\gamma^2\,\gamma_0^2}{4\,M^2} - 1 = \frac{2\,(\ln 2M - \ln\gamma\,\gamma_0)}{\ln M^2 - \ln\,(M^2-\alpha^2)} - 1.$$

Таким образом, из следствия 1 при $\beta = 1$ (в выпуклом случае) вытекает

**Следствие 2.** *Пусть $f$ — выпуклый $M$-липшицев ($M > 0$) положительно однородный функционал, причем для всякого $x \in Q$ верно неравенство $\gamma_0\|x\|_2 \leq f(x)$. Тогда для всякого достаточно малого $\gamma > 0$ при условии $\|x_0\|_2 = \min\limits_{x\in Q}\|x\|_2$ после*

$$k \geq \frac{2\left(\ln 2M - \ln\gamma\gamma_0\right)}{\ln M^2} - \ln\left(M^2 - \alpha^2\right) - 1$$

*итераций алгоритма* (3.4) *имеем неравенство* $f(x_{k+1}) \leq (1+\gamma)f^*$.

## 4. Об универсальности одного субградиентного метода для задач минимизации гельдеровых квазивыпуклых функционалов с острым минимумом

Теперь рассмотрим класс задач (3.1) минимизации квазивыпуклых функционалов $f$:

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

в предположении о том, что $f$ удовлетворяет условию Гельдера

$$|f(x) - f(y)| \leq M_\nu\|x-y\|_2^\nu \quad \forall\, x, y \in Q \tag{4.1}$$

при некотором фиксированном $\nu \in [0;1]$, $0 \leq M_\nu < +\infty$.

Введем (следуя [4; 5]) вспомогательную величину

$$\upsilon_f(x, x_*) := \begin{cases} \left\langle \dfrac{\nabla f(x)}{\|\nabla f(x)\|_2}, x - x_* \right\rangle, & \text{если} \quad x \neq x_*; \\ 0 & \text{при} \quad x = x_*. \end{cases}$$

Если $\nabla f(x) = 0$ при $x \neq x_*$, то вместо $\nabla f(x)$ можно использовать ненулевой вектор нормали $\mathcal{D}f(x)$ ко множеству уровня функционала $f$ в точке $x$ [4]. Но для упрощения изложения материала далее сделаем допущение, что $\nabla f(x) \neq 0$ при $x \neq x_*$. Справедлива следующая лемма.

**Лемма.** *Если функционал $f$ квазивыпуклый и удовлетворяет* (4.1)*, то для всякого $x \in Q$*

$$f(x) - f(x_*) \leq M\upsilon_f(x, x_*), \tag{4.2}$$

*где* $M = M(\nu) = \max\Big\{M_\nu, \dfrac{M_\nu^{\frac{2}{1+\nu}} + 1}{2}\Big\}$.

Д о к а з а т е л ь с т в о. Неравенство (4.1) означает, что

$$f(x) - f(x_*) \leq M_\nu\, \upsilon_f^\nu(x, x_*) \quad \forall x \in Q \tag{4.3}$$

(см. монографию [5, лемма 3.2.1]). Если в (4.3) $\upsilon_f(x, x_*) \geq 1$ для некоторого $x \in Q$, то

$$f(x) - f(x_*) \leq M_\nu\, \upsilon_f\,(x, x_*).$$

Если же $\upsilon_f(x, x_*) < 1$ , то применим неравенство из [1, замечание 5.1]

$$M_\nu a^\nu \leq \frac{M_\nu^{\frac{2}{1+\nu}}}{2}\,\frac{a^2}{\delta^{\frac{1-\nu}{1+\nu}}} + \frac{\delta}{2},$$

которое верно $\forall\, a,\ \delta > 0,\ M_\nu > 0$ и при всяком $\nu \in [0;1]$. В таком случае (4.3) означает, что

$$f(x) - f(x_*) \leq \frac{M_\nu^{\frac{2}{1+\nu}}}{2\delta^{\frac{1-\nu}{1+\nu}}}\,\upsilon_f^2(x, x_*) + \frac{\delta}{2}.$$

Если выбрать $\delta = \upsilon_f(x,x_*) > 0$, то получим

$$f(x) - f(x_*) \le \frac{M_\nu^{\frac{2}{1+\nu}}}{2}\upsilon_f^{2-\frac{1-\nu}{1+\nu}}(x,x_*) + \frac{\upsilon_f(x,x_*)}{2} \le \frac{M_\nu^{\frac{2}{1+\nu}}+1}{2}\upsilon_f(x,x_*)$$

ввиду $\upsilon_f(x,x_*) < 1$. Случай $\upsilon_f(x,x_*) = 0$ очевиден. □

Рассмотрим метод ($k = 0,1,2,\ldots$)

$$x_{k+1} = Pr_Q\{x_k - h_k\nabla f(x_k)\}, \quad \text{где} \quad h_k = \frac{f(x_k)-f(x_*)}{M\,\|\nabla f(x_k)\|_2}, \tag{4.4}$$

$M$ удовлетворяет (4.2). Докажем, что данный алгоритм универсален в том смысле, что сходится со скоростью геометрической прогрессии для задачи минимизации квазивыпуклого гельдерова функционала с острым минимумом и заранее известным точным значением $f^*$ при всех $\nu \in [0;1]$.

**Теорема 5.** *Пусть верно* (4.2) *при некотором* $0 < M < +\infty$ *и* $f$ *имеет* $\alpha$*-острый минимум* (3.3). *Тогда метод* (4.4) *сходится со скоростью геометрической прогрессии*:

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le \left(1-\frac{\alpha^2}{M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2. \tag{4.5}$$

Доказательство. Действительно, для метода (4.4) верны неравенства (см. [6, теорема 1, соотношения (3)])

$$2\,h_k\langle\nabla f(x_k), x_k - x_*\rangle \le h_k^2\,\|\nabla f(x_k)\|_2^2 + \|x_k-x_*\|_2^2 - \|x_{k+1}-x_*\|_2^2 \quad \forall\, k\ge 0.$$

Поэтому имеем

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le \min_{x_*\in X_*}\|x_k-x_*\|_2^2 + h_k^2\|\nabla f(x_k)\|_2^2 - 2h_k\langle\nabla f(x_k), x_k-x_*\rangle$$

$$= \min_{x_*\in X_*}\|x_k-x_*\|_2^2 + \frac{(f(x_k)-f(x_*))^2}{M^2} - \frac{2\,(f(x_k)-f(x_*))}{M}\left\langle\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2}, x_k-x_*\right\rangle$$

$$= \min_{x_*\in X_*}\|x_k-x_*\|_2^2 + \left(\frac{f(x_k)-f(x_*)}{M}\right)^2 - \frac{2(f(x_k)-f(x_*))}{M}\upsilon_f(x_k,x_*)$$

$$\le \min_{x_*\in X_*}\|x_k-x_*\|_2^2 - \left(\frac{f(x_k)-f(x_*)}{M}\right)^2 \le \min_{x_*\in X_*}\|x_k-x_*\|_2^2\left(1-\frac{\alpha^2}{M^2}\right)$$

ввиду (3.3), так как $\upsilon_f(x_k,x_*) \ge \dfrac{f(x_k)-f(x_*)}{M}$. Ясно, что

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \le \left(1-\frac{\alpha^2}{M^2}\right)\min_{x_*\in X_*}\|x_k-x_*\|_2^2$$

$$\le \left(1-\frac{\alpha^2}{M^2}\right)\min_{x_*\in X_*}\|x_{k-1}-x_*\|_2^2 \le \ldots \le \left(1-\frac{\alpha^2}{M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2,$$

что и требовалось. □

Сформулируем теперь следствие из теоремы 5 об оценке скорости сходимости алгоритма (4.4) для задач выпуклой положительно однородной минимизации с относительной точностью по целевому функционалу.

**Следствие 3.** *Пусть $f$ — выпуклый $M$-липшицев ($M > 0$) положительно однородный функционал, причем для всякого $x \in Q$ верно неравенство $\gamma_0 \|x\|_2 \leq f(x)$. Тогда для всякого достаточно малого $\gamma > 0$ при условии $\|x_0\|_2 = \min\limits_{x \in Q} \|x\|_2$ после*

$$k \geq \frac{2\,(\ln 2M - \ln \gamma\,\gamma_0)}{\ln M^2 - \ln\,(M^2 - \alpha^2)} - 1$$

*итераций алгоритма* (4.4) *имеем неравенство* $f(x_{k+1}) \leq (1+\gamma) f^*$.

З а м е ч а н и е. Некоторым недостатком алгоритма (4.4) по сравнению с алгоритмом (3.4) может считаться требование знать $M$ из неравенства леммы при организации шагов. Однако эта же особенность позволяет предложить вариацию субградиентного метода для задач с некоторым $\Delta$-обобщением острого минимума вида

$$f(x) - \overline{f} \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2 - \Delta \quad \forall\, x \in Q$$

при фиксированном значении $\Delta > 0$ и заданном $\overline{f} \geq f^*$. В частности, данное условие логично использовать при отсутствии точной информации о $f^*$. Если выбрать в алгоритме (4.4) шаги вида

$$h_k = \frac{f(x_k) - \overline{f}}{M \|\nabla f(x_k)\|_2},$$

то получим соотношения (аналогично рассуждениям из доказательства теоремы 5)

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq \min_{x_* \in X_*} \|x_k - x_*\|_2^2 - \frac{1}{M^2}\,(f(x_k) - \overline{f})^2$$

(применим неравенство $a^2 \geq (1/2)(a+b)^2 - b^2$, справедливое для всяких $a, b \in \mathbb{R}$)

$$\leq \min_{x_* \in X_*} \|x_k - x_*\|_2^2 \Big(1 - \frac{\alpha^2}{2M^2}\Big) + \frac{\Delta^2}{M^2}$$

$$\leq \min_{x_* \in X_*} \|x_0 - x_*\|_2^2 \Big(1 - \frac{\alpha^2}{2\,M^2}\Big)^{k+1} + \frac{\Delta^2}{M^2}\Big(1 + \Big(1 - \frac{\alpha^2}{2\,M^2}\Big) + \ldots + \Big(1 - \frac{\alpha^2}{2\,M^2}\Big)^k\Big)$$

$$\leq \min_{x_* \in X_*} \|x_0 - x_*\|_2^2 \Big(1 - \frac{\alpha^2}{2\,M^2}\Big)^{k+1} + \frac{2\,\Delta^2}{\alpha^2}.$$

Таким образом, наличие неточности $\Delta$ приводит к дополнительному слагаемому вида $O(\Delta^2)$ в оценке невязки $\min\limits_{x_* \in X_*} \|x_{k+1} - x_*\|_2$. По-видимому, в случае малых значений $\|\nabla f(x_k)\|_2$ для алгоритма (3.4) такой вывод сделать уже нельзя.

На первый взгляд, условие (3.3) представляется довольно ограничительным. Однако это не совсем так, что показано в следующих примерах.

П р и м е р 2. Пусть $f$ имеет *квадратичный рост*:

$$f(x) - f(x_*) \geq \mu\, \|x - x_*\|_2^2 \quad \forall\, x \in Q \tag{4.6}$$

при некотором $\mu > 0$. Например, условие (4.6) заведомо верно для всякого $\mu$-сильно выпуклого функционала $f$. Условие (4.6) означает, что $F(x) := \sqrt{f(x) - f(x_*)} \geq \sqrt{\mu}\,\|x - x_*\|_2$ при всяком $x \in Q$ и $F$ удовлетворяет условию острого минимума (3.3) при $\alpha = \sqrt{\mu}$. Если функционал $f$ квазивыпуклый, то и $F$ квазивыпуклый. Если $f$ $M_f$-липшицев, то неравенство $|\sqrt{a} - \sqrt{b}| \leq \sqrt{|a-b|}$ $(a, b \in \mathbb{R})$ влечет гельдеровость $F$ при $\nu = 1/2$ (аналогичный вывод можно сделать и для гельдеровых $f$). Это приводит к заключению: оценка (4.5) верна для минимизации $M_f$-липшицева квазивыпуклого функционала $f$ с квадратичным ростом в предположении доступности $f^* > 0$. При этом $\alpha^2/M^2 = O\,(\mu/M_f)$.

П р и м е р 3. Во многих прикладных задачах возникает также условие так называемого *слабого острого минимума* (в западной литературе для этого понятия можно встретить термин *условие гельдерова роста*; см., например, [12–14])

$$f(x) - f^* \geq \mu \min_{x_* \in X_*} \|x - x_*\|_2^p \quad \forall\, x \in Q$$

для некоторого фиксированного $p \in [1; +\infty)$ и некоторой постоянной $\mu > 0$. Ясно, что это условие обобщает как обычное условие острого минимума (3.3), так и условие квадратичного роста (4.6). В этом случае снова возможно применить схему рассуждений предыдущего примера. Действительно, если функционал $f$ квазивыпуклый, то

$$F(x) := (f(x) - f^*)^{1/p}$$

также квазивыпуклый и $\min\limits_{x \in Q} F(x) = 0$, т. е. $F$ имеет $(\mu^{1/p})$-острый минимум. В случае, если $f$ удовлетворяет условию Гельдера с показателем $\nu \in [0; 1]$, то $F$ также удовлетворяет условию Гельдера с показателем $\nu/p$, причем $\nu/p \in [0; 1]$. Таким образом, к задаче минимизации $f$ применима оценка (4.5) теоремы 5, означающая линейную скорость сходимости. Это указывает на некоторую универсальность алгоритма (4.4) по параметру $p \in [1; +\infty)$. Отметим, что ранее для задач со слабым острым минимумом при $\nu \neq p > 1$ были известны результаты только о сублинейной скорости сходимости субградиентных методов, и вообще не рассматривался случай $\nu = 0$ (см. [11] и имеющиеся в этой работе ссылки).

## Заключение

В настоящей статье описаны новые результаты о скорости сходимости методов градиентного типа для двух вариантов постановки задачи. Получена оценка скорости сходимости адаптивных методов для задач выпуклой однородной минимизации с гарантией достижения заданной относительной точности по целевому функционалу. Вторая часть статьи (разд. 3 и 4) посвящена результатам о линейной скорости сходимости субградиентных методов для задач с острым минимумом. Наиболее тонкий, на наш взгляд, результат получен в разд. 4 для класса задач минимизации квазивыпуклых гельдеровых функционалов с острым минимумом. Как следствие обоснована возможность построения субградиентного метода с линейной скоростью сходимости по аргументу для задач со слабым острым минимумом (гельдеровым ростом) в предположении доступности информации о минимальном значении $f^*$. Последнее условие представляется довольно ограничительным, особенно для выводов по задачам с относительной точностью (следствия 2 и 3). Поэтому в качестве возможного развития настоящей работы представляется интересным исследование вариаций предложенных методов с неточной информацией как об $f^*$ (см. замечание), так и о значениях функции $f$, субградиентов $\nabla f$ в произвольных запрашиваемых точках допустимого множества решаемой задачи, т. е. для задач с какими-то подходящими аналогами $(\delta, L)$-оракула (см. [8]).

**Благодарности**. Авторы благодарят А. В. Гасникова и Ю. Е. Нестерова за полезные обсуждения.

## СПИСОК ЛИТЕРАТУРЫ

1. **Гасников А. В.** Современные численные методы оптимизации. Метод универсального градиентного спуска. М.: МЦНМО, 2021. 272 с.
2. **Немировский А. С., Юдин Д. Б.** Сложность задач и эффективность методов оптимизации. М.: Наука, 1979. 384 с.
3. **Нестеров Ю. Е.** Алгоритмическая выпуклая оптимизация: дис. . . . докт. физ.-мат. наук. М.: Моск. физ.-техн. ин-т, 2013. 367 с.
4. **Нестеров Ю. Е.** Эффективные методы нелинейного программирования. М.: Радио и связь, 1989. 301 с.

5. **Нестеров Ю. Е.** Методы выпуклой оптимизации. М.: МЦНМО, 2010. 281 с.
6. **Поляк Б. Т.** Минимизация негладких функционалов // Журн. вычисл. математики и мат. физики. 1969. Т. 9, № 3. С. 509–521.
7. **Тюрин А. И., Гасников А. В.** Быстрый градиентный спуск для задач выпуклой минимизации с оракулом, выдающим $(\delta, L)$-модель функции в запрошенной точке // Журн. вычисл. математики и мат. физики. 2019. Т. 59, № 7. С. 1137–1150. doi: 10.1134/S0044466919070081 .
8. **Devolder O., Glineur F., Nesterov Yu.** First-order methods of smooth convex optimization with inexact oracle // Math. Programming. 2014. Vol. 146, no. 1. P. 37–75.
9. **Hardt M., Ma T., Recht B.** Gradient descent learns linear dynamical systems // J. of Machine Learning Research. 2018. Vol. 19, no. 29. P. 1–44.
10. **Hinder O., Sidford A., Sohoni N. S.** Near-optimal methods for minimizing star-convex functions and beyond // Proceedings of Machine Learning Research. 2020. Vol. 125. P. 1894–1938.
11. **Hu Y., Li J., Yu C. K. W.** Convergence rates of subgradient methods for quasi-convex optimization problems. 2019. 28 p. URL: arXiv:1910.10879 [math.OC] .
12. **Jiang R., Li X.** Holderian error bounds and Kurdyka–Lojasiewicz inequality for the trust region subproblem. 2020. 30 p. URL: arXiv:1911.11955 [math.OC] .
13. **Johnstone P. R., Moulin P.** Faster subgradient methods for functions with Holderian growth. 2018. 50 p. URL: arXiv:1704.00196 [math.OC] .
14. **Liu M., Yang T.** Adaptive accelerated gradient converging methods under Holderian error bound condition // Proc. 31st Intern. Conf. on Neural Information Processing Systems (NIPS'17). 2017. P. 3107–3117. ISBN: 9781510860964.
15. **Nesterov Yu.** Rounding of convex sets and efficient gradient methods for linear programming problems // Optimization Methods and Software. 2008. Vol. 23, no. 1. P. 109–128. doi: 10.1080/10556780701550059 .
16. **Nesterov Yu.** Unconstrained convex minimization in relative scale // Math. Oper. Res. 2009. Vol. 34, no. 1. P. 180–193. doi: 10.1287/moor.1080.0348 .
17. **Nesterov Yu.** Universal gradient methods for convex optimization problems // Math. Programming. 2015. Vol. 152, no. 1-2(A). P. 381–404. doi: 10.1007/s10107-014-0790-0 .

Поступила 19.07.2021
После доработки 1.09.2021
Принята к публикации 6.09.2021

Стонякин Федор Сергеевич
д-р физ.-мат. наук
доцент
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: fedyor@mail.ru

Аблаев Сейдамет Серверович
аспирант
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: seydamet.ablaev@yandex.ru

Баран Инна Викторовна
канд. физ.-мат. наук
старший преподаватель
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: matemain@mail.ru

## REFERENCES

1. Gasnikov A.V. *Sovremennaya vypuklaya optimizatsiya. Method universalnogo gradientnogo spuska* [Modern numerical optimization methods. The universal gradient descent method]. Moscow: MCCME Publ., 2021, 272 p. ISBN: 978-5-4439-1614-9 .

2. Nemirovsky A.S., Yudin D.B. *Problem complexity and method efficiency in optimization.* Chichester; NY: Wiley, 1983, 388 p. ISBN: 0471103454 . Original Russian text published in Nemirovsky A.S., Yudin D.B. *Slozhnost' zadach i effektivnost' metodov optimizatsii*, Moscow: Nauka Publ., 1979, 384 p.
3. Nesterov Yu. E. *Algoritmicheskaya vypuklaya optimizatsiya* [Algorithmic convex optimization]. Dr. Phys.-Math. Sci. Dissertation. Moscow: Mosk. Phys.-Techn. Inst. (State University), 2013. 367 p.
4. Nesterov Yu. *Effectivnye metody v nelineinom programmirovanii* [Effective methods in nonlinear programming]. Moscow: Radio i Svyaz' Publ., 1989. ISBN: 5-256-00524-3 .
5. Nesterov Yu.E. *Metody vypukloy optimizatsii* [Methods of convex optimization]. Moscow: Publishing House MCNMO, 2010. ISBN: 978-5-94057-623-5 .
6. Polyak B.T. Minimization of nonsmooth functionals. *U.S.S.R. Comput. Math. Math. Phys.*, 1969, vol. 9, no. 3, pp. 14–29. doi: 10.1016/0041-5553(69)90061-5 .
7. Gasnikov A.V., Turin A.I. Fast gradient descent for convex minimization problems with an oracle producing a $(\delta, L)$-model of a function in a requested point. *Comput. Math. Math. Phys.*, 2019, vol. 59, no. 7, pp. 1085–1097. doi: 10.1134/S0965542519070078 .
8. Devolder O., Glineur F., Nesterov Yu. First-order methods of smooth convex optimization with inexact oracle. *Math. Program.*, 2014, vol. 146, no. 1-2 (A), pp. 37–75. doi: 10.1007/s10107-013-0677-5 .
9. Hardt M., Ma T., Recht B. Gradient descent learns linear dynamical systems. *J. Mach. Learn. Res.*, 2018, vol. 19, art. no. 29, 44 p.
10. Hinder O., Sidford A., Sohoni N.S. Near-optimal methods for minimizing star-convex functions and beyond. *Proceedings of Machine Learning Research*, 2020, vol. 125, pp. 1894–1938.
11. Hu Y., Li J., Yu C.K.W. Convergence rates of subgradient methods for quasi-convex optimization problems. 2019. 28 p. Available on: arXiv:1910.10879 [math.OC] .
12. Jiang R., Li X. Holderian error bounds and Kurdyka–Lojasiewicz inequality for the trust region subproblem. 2020. 30 p. Available on: arXiv:1911.11955 [math.OC] .
13. Johnstone P.R., Moulin P. Faster subgradient methods for functions with Holderian growth. 2018. 50 p. Available on: arXiv:1704.00196 [math.OC] .
14. Liu M., Yang T. Adaptive accelerated gradient converging methods under Holderian error bound condition. In: *Proc. 31st Intern. Conf. on Neural Information Processing Systems* (*NIPS'17*), 2017, pp. 3107–3117. ISBN: 9781510860964.
15. Nesterov Yu. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optim. Methods Softw.*, 2008, vol. 23, no. 1, pp. 109–128. doi: 10.1080/10556780701550059 .
16. Nesterov Yu. Unconstrained convex minimization in relative scale. *Math. Oper. Res.*, 2009, vol. 34, no. 1, pp. 180–193. doi: 10.1287/moor.1080.0348 .
17. Nesterov Yu. Universal gradient methods for convex optimization problems. *Math. Program.*, 2015, vol. 152, no. 1-2 (A), pp. 381–404. doi: 10.1007/s10107-014-0790-0 .



Received July 19, 2021
Revised September 1, 2021
Accepted September 6, 2021

**Funding Agency**: The research in Sects. 2 and 3 was supported by the Russian President's Grant for Young Russian Scientists no. MK-15.2020.1. Stonyakin's results in Sect. 4 were supported by the Russian Science Foundation grant (project no. 27-31-30005).



*Fedor Sergeevich Stonyakin*, Cand. Sci. (Phys.-Math.), V. I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: fedyor@mail.ru .

*Seydamet Serverovich Ablaev*, doctoral student, V. I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: seydamet.ablaev@yandex.ru .

*Inna Viktorovna Baran*, Cand. Sci. (Phys.-Math.), V. I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: matemain@mail.ru .



Cite this article as: F. S. Stonyakin, S. S. Ablaev, I. V. Baran. Adaptive gradient-type methods for optimization problems with relative error and sharp minimum, *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2021, vol. 27, no. 4, pp. 175–188 .