*Автоматика и телемеханика, № 9, 2025*

# *Тематический выпуск (окончание)*[1]

**© 2025 г. С.С. АБЛАЕВ, канд. физ.-мат. наук (seydamet.ablaev@yandex.ru)**
**(Крымский федеральный университет им. В.И. Вернадского, Симферополь;**
**Московский физико-технический институт**
**(национальный исследовательский университет)),**
**Ф.С. СТОНЯКИН, д-р физ.-мат. наук (fedyor@mail.ru)**
**(Московский физико-технический институт**
**(национальный исследовательский университет);**
**Крымский федеральный университет им. В.И. Вернадского, Симферополь),**
**М.Н. ФЕДОТОВ (MaximFd-Nk@yandex.ru)**
**(Национальный исследовательский университет**
**«Высшая школа экономики», Москва),**
**М.С. АЛКУСА, канд. физ.-мат. наук (m.alkousa@innopolis.ru)**
**(Университет Иннополис),**
**О.С. САВЧУК (oleg.savchuk19@mail.ru)**
**(Крымский федеральный университет имени В.И. Вернадского, Симферополь;**
**Адыгейский государственный университет, Майкоп),**
**А.В. ГАСНИКОВ, д-р физ.-мат. наук (gasnikov@yandex.ru)**
**(Университет Иннополис, Долгопрудный;**
**Московский физико-технический институт**
**(национальный исследовательский университет))**

## ИССЛЕДОВАНИЕ ГРАДИЕНТНОГО МЕТОДА С НЕТОЧНОЙ ИНФОРМАЦИЕЙ О ГРАДИЕНТЕ НА НЕКОТОРЫХ КЛАССАХ $(L_0, L_1)$-ГЛАДКИХ НЕВЫПУКЛЫХ ЗАДАЧ[2]

Работа посвящена исследованию градиентного метода на классе $(L_0, L_1)$-гладких функций при условии, что на итерациях метода вместо точных значений градиента доступны лишь его приближенные значения, что соответствует ситуациям, возникающим при использовании зашумленных данных. Рассмотрено два класса задач: первый – квазар-выпуклые функции относительно всякого решения, удовлетворяющие условию градиентного доминирования Поляка–Лоясиевича, второй – квазар-выпуклые функции без требования выполнения условия Поляка–Лоясиевича, но с дополнительным ограничением на параметр квазар-выпуклости. Для квазар-выпуклых функций с PL-условием доказан результат о близкой к линейной скорости сходимости метода в окрестность точного решения. Если значения неточного градиента достаточно малы (что достигается за конечное число итераций), то метод сходится

[1] Статьи с 3 стр. по 63 стр. являются окончанием тематического выпуска № 8, 2025.

[2] Исследование выполнено при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания № 075-03-2024-074 по проекту «Исследование асимптотических характеристик колеблемости дифференциальных уравнений и систем, а также оптимизационных методов».

с близкой к линейной скоростью на классе задач с условием Поляка–Лоясиевича без дополнительного предположения о квазар-выпуклости. Для $(0, M)$-гладких квазар-выпуклых функций предложен адаптивный градиентный метод и получена оценка его скорости сходимости. Показано, что в случае использования точных значений градиента метод сходится с линейной скоростью.

*Ключевые слова*: градиентный метод, $\Delta$-неточный градиент, $(L_0, L_1)$-гладкая функция, $\rho$-квазар-выпуклая функция, условие Поляка–Лоясиевича, логистическая регрессия.

**DOI:** 10.31857/S0005231025090015, **EDN:** VMPJNF

## 1. Введение

В связи с активным развитием машинного обучения возрастает актуальность разработки и анализа численных методов оптимизации с эффективными скоростными гарантиями. Для достижения оптимальной скорости сходимости традиционных методов, таких как градиентный спуск и его модификации, ключевым предположением является гладкость целевой функции (т.е. липшицевость ее градиента). В контексте современных задач машинного обучения такое допущение может оказаться слишком жестким [1]. Часто достаточно естественные и простые с точки зрения реализации методы плохо работают для задач в глубоком обучении, где глобальное условие гладкости зачастую не выполняется. Например, методы редукции дисперсии [2–8], как известно, являются более быстрыми с точки зрения теории (для задач минимизации конечных сумм гладких функций), но на практике могут уступать методам, теоретически не снижающим дисперсию [9]. В экспериментах, приведенных в [1], показано, что при обучении нейронных сетей норма гессиана может коррелировать с нормой градиента функции потерь. Таким образом, вышеизложенные факторы стимулируют поиск и разработку альтернативных предположений, ослабляющих стандартное требование гладкости целевой функции.

Одним из таких предположений является так называемая обобщенная $(L_0, L_1)$-гладкость, впервые предложенная несколько лет назад в [1] для дважды дифференцируемых функций. Это предположение позволяет норме гессиана целевой функции расти линейно с ростом нормы градиента. В частности, $(L_0, L_1)$-гладкость может сохраняться даже для функций с полиномиально растущими градиентами – типичная ситуация для задач глубокого обучения. Более того, концепция $(L_0, L_1)$-гладкости охватывает широкий класс задач и допускает обобщение на дифференцируемые, но не обязательно дважды дифференцируемые функции [10, 11].

В последние годы во многих работах исследовались различные методы решения задач $(L_0, L_1)$-гладкой оптимизации. Однако, несмотря на интерес со стороны научного сообщества, существующие результаты относительно сходимости методов в ряде важных случаев остаются неоптимальными, а теоретический анализ зачастую является недостаточно полным. В то время как авторы [1], как и далее в [10–16], сосредоточились на максимально общем классе невыпуклых задач, допускающих лишь сходимость к нулю метрик типа нормы градиента, класс $(L_0, L_1)$-гладких выпуклых функций до совсем недавнего времени (конец 2024 – начало 2025 гг.) был изучен гораздо слабее. В частности, известные результаты о сходимости методов, таких как градиентный спуск с клиппингом [17] и градиентный спуск с шагом Поляка [18], в применении к $(L_0, L_1)$-гладким выпуклым задачам либо требуют дополнительного предположения об $L$-гладкости [19, 20], либо накладывают ограничения на размер шага, необходимые для обеспечения ограниченности траектории метода в области, где градиент удовлетворяет условию Липшица в силу $(L_0, L_1)$-гладкости целевой функции [21].

Приведем обзор нескольких работ последнего времени, посвященных оптимизации $(L_0, L_1)$-гладких функций.

В [22] авторы сосредоточились на исследовании класса сильно выпуклых $(L_0, L_1)$-гладких функций и получили новые гарантии скорости сходимости для нескольких существующих методов. В частности, были получены улучшенные оценки скорости сходимости для градиентного спуска с (сглаженным) клиппингом и для градиентного спуска с шагом Б.Т. Поляка. В отличие от уже существующих результатов эти оценки не основаны на стандартном предположении о гладкости и не имеют экспоненциальной зависимости от расстояния от начальной точки до решения.

Параллельно с [22] в [23] независимо были получены аналогичные результаты для градиентного метода с различными стратегиями выбора шага, включая шаг Поляка, а также получены новые оценки сложности для градиентного метода с нормированным шагом на классе $(L_0, L_1)$-гладких задач.

В [24] представлен более детальный анализ сходимости градиентного спуска и его вариантов в условиях обобщенной $(L_0, L_1)$-гладкости. В частности, показано, что в выпуклом случае скорость сходимости градиентного метода меняется в зависимости от нормы градиента: при $\|\nabla f(x_k)\| \geqslant \frac{L_0}{L_1}$ достигается линейная скорость, а при $\|\nabla f(x_k)\| < \frac{L_0}{L_1}$ наблюдается стандартная сублинейная скорость сходимости.

В отличие от рассмотренных выше работ в настоящем исследовании анализируется градиентный метод для $(L_0, L_1)$-гладких задач в условиях, когда на каждой итерации доступны лишь приближенные (аддитивно неточные) значения градиента. То есть получила развитие в новом направлении статья [25], где аналогичные проблемы исследовались для гладких задач в классическом понимании. При этом рассматривается несколько направлений. Во-первых, исследована сходимость метода для $\rho$-квазар-выпуклых относительно любого точного решения [26–28] $(L_0, L_1)$-гладких функций, удовлетворяющих условию Поляка–Лоясиевича

$$f(x) - f(x_*) \leqslant \frac{1}{2\mu}\|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^n,\ \mu > 0, \tag{1}$$

где $\|\cdot\|$ – евклидова норма, $x_*$ – точка минимума функции $f$. В работе показано, что при некоторых предположениях метод сходится с линейной скоростью в некоторую окрестность точного решения. Доказано, что начиная с некоторого номера итерации метод сходится на классе задач с условием Поляка–Лоясиевича (без требования квазар-выпуклости). Во-вторых, отдельно рассматривается класс $\rho$-квазар-выпуклых $(0, M)$-гладких функций, представляющий интерес в контексте задач машинного обучения, в частности при обучении моделей логистической регрессии. Для такого класса задач предложен адаптивный вариант градиентного метода с использованием неточных градиентов на итерациях, и получены оценки скорости сходимости при некотором ограничении на параметр квазар-выпуклости $\rho$. Эффективность предложенных методов подтверждена численными экспериментами на задачах логистической регрессии (удовлетворяющей условию Поляка–Лоясиевича на любом компактном множестве [29]) и на некоторой невыпуклой квазар-выпуклой задаче [27].

## 2. Постановка задачи и необходимые вспомогательные понятия

Будем рассматривать задачи минимизации вида

$$\min_{x\in\mathbb{R}^n} f(x), \tag{2}$$

где $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ – выпуклая и $(L_0, L_1)$-гладкая функция. Всюду в работе $x_*$ будет обозначать точку глобального минимума, а $f^* = f(x_*)$ – глобальное минимальное значение функции, норма $\|\cdot\|$ – евклидова.

Как уже было отмечено, условие $(L_0, L_1)$-гладкости $(L_0, L_1 > 0)$ было первоначально предложено в [1] для дважды дифференцируемых функций $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ следующим образом:

$$\|\nabla^2 f(x)\| \leqslant L_0 + L_1\|\nabla f(x)\| \quad \forall x \in \mathbb{R}^n, \tag{3}$$

где $\|\cdot\|$ обозначает евклидову норму для векторов и спектральную норму для матриц.

В [10] это понятие было распространено на класс дифференцируемых, но необязательно дважды дифференцируемых, функций.

*Определение 1. Функция* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *называется* $(L_0, L_1)$*-гладкой, если* $\forall\ x, y \in \mathbb{R}^n$ *при условии* $\|y - x\| \leqslant \frac{1}{L_1}$ *справедливо неравенство*

$$\|\nabla f(x) - \nabla f(y)\| \leqslant (L_0 + L_1\|\nabla f(y)\|)\,\|x - y\|. \tag{4}$$

Рассмотрим несколько примеров функций, которые удовлетворяют условию (4).

*Пример 1.* Функция $f(x) = \|x\|^{2m}$ является $(2m, 2m-1)$-гладкой. При этом $f$ не является $L$-гладкой (напомним, что $L$-гладкой называется функция, градиент которой удовлетворяет условию Липшица) при $m \geqslant 2$ для всех $L \geqslant 0$ (см. [22]).

*Пример 2.* Функция $f(x) = e^{ax}$, $a \in \mathbb{R}^n$ является $(0, \|a\|)$-гладкой. При этом $f(x)$ не является $L$-гладкой при $a \neq 0$ для всех $L \geqslant 0$ (см. [22]).

*Пример 3.* Логистическая функция вида $f(x) = \log(1 + \exp(-a^T x))$, где $a \in \mathbb{R}^n$, является $(L_0, L_1)$-гладкой с $L_0 = 0$ и $L_1 = \|a\|$ (см. [22]).

Понятие $(L_0, L_1)$-гладкости было обобщено в [11] следующим образом: $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ называется $\alpha$-симметрично обобщенно гладкой функцией, если при некотором $\alpha \in [0, 1]$ $\forall x, y \in \mathbb{R}^n$ выполняется условие

$$(5) \qquad \|\nabla f(y) - \nabla f(x)\| \leqslant \left(L_0 + L_1 \max_{\theta \in [0,1]} \|\nabla f(\theta y + (1-\theta)x)\|^{\alpha}\right) \|y - x\|.$$

В [11] показано, что условие (5) является более общим, чем (3) и (4): из выполнения (3) или (4) следует выполнение (5) (при $\alpha = 1$). При этом для дважды дифференцируемых функций условия (3) и (5) эквивалентны при $\alpha = 1$ [11].

Приведем определение квазар-выпуклой функции, которое является одним из ключевых в данной работе.

*Определение 2. Пусть $\rho \in (0, 1]$ и $\mu > 0$. Функция $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ называется $(\rho, \mu)$-квазар сильно выпуклой, если для всех $x \in \mathbb{R}^n$ справедливо неравенство*

$$(6) \qquad f(x_*) \geqslant f(x) + \frac{1}{\rho}\langle \nabla f(x), x_* - x\rangle + \frac{\mu}{2}\|x_* - x\|^2,$$

*где $x_*$ – точка глобального минимума функции $f$. При $\mu = 0$ функция $f$ называется $\rho$-квазар выпуклой функцией.*

Условие Поляка–Лоясиевича можно рассматривать не только как релаксацию обычной сильной выпуклости, но и сильной квазар-выпуклости [27]: если $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ есть $(\rho, \mu)$-квазар сильно выпуклая функция, то она удовлетворяет условию Поляка–Лоясиевича (1) с константой $\mu\rho^2$.

*Пример 4* [27]. Пусть $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ определяется как

$$f(x) = \|x\|^2 g\left(\frac{x}{\|x\|}\right), \quad \forall x \in \mathbb{R}^n,$$

где

$$g(x) = 1 + \sum_{i=1}^{n} a_i \sin(b_i x_i)^2, \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Тогда функция $f$ является $(1, 2)$-квазар сильно выпуклой, т.е. удовлетворяет PL-условию (1) с константой 2.

Вопросы сходимости градиентного метода для выпуклых задач с обобщенным условием гладкости (4) активно исследовались в последнее время, например, в [22–24]. В этих работах получены результаты о сходимости нормированного градиентного метода, градиентного метода с шагом Поляка и ускоренного градиентного метода исходя из предположения обобщенной гладкости. Также было показано, что градиентный метод сходится с линейной скоростью сходимости до тех пор, пока значения норм градиентов на итерациях остаются достаточно большими. Данная статья посвящена некоторому развитию этих работ.

Здесь сосредоточимся на исследовании градиентного спуска для задач минимизации вида (2) в условиях, когда вместо точных значений градиента на каждой итерации метода доступен лишь градиент, содержащий аддитивную погрешность. Напомним, что вектор $\widetilde{\nabla} f(x)$ называется неточным градиентом в точке $x$ при $\Delta > 0$, если

$$\|\nabla f(x) - \widetilde{\nabla} f(x)\| \leqslant \Delta.$$

В случае малых значений норм градиента на итерациях метода становится проблематично получить теоретические оценки о сходимости. Это обусловлено тем, что в отличие от случая использования точных градиентов невозможно гарантировать монотонное убывание нормы градиента на каждой итерации. Эту проблему можно обойти, если дополнительно потребовать выполнение условия Поляка–Лоясиевича для целевой функции $f$. Как показано в [25], для задач с неточными градиентами из условия Поляка–Лоясиевича следует неравенство

$$\|\widetilde{\nabla} f(x)\|^2 \geqslant \mu(f(x) - f^*) - \Delta^2 \quad \forall x \in \mathbb{R}^n,\ \mu > 0. \tag{7}$$

В данном разделе приведены постановка задачи и необходимые предварительные сведения. В разделе 3 исследован градиентный метод для $(L_0, L_1)$-гладких задач при наличии аддитивной погрешности в градиенте на каждой итерации. Получена оценка скорости сходимости метода на классе $\rho$-квазар-выпуклых функций, удовлетворяющих условию Поляка–Лоясиевича. Показано, что до определенного уровня значений градиента скорость сходимости метода линейная. В разделе 4 исследован адаптивный вариант градиентного метода для $(0, M)$-гладких $\rho$-квазар-выпуклых задач. В разделе 5 приведены результаты вычислительных экспериментов для задачи логистической регрессии (удовлетворяющей на всяком компакте условию Поляка–Лоясиевича [29]) и для некоторой квазар-выпуклой задачи [27], которая не выпукла.

## 3. Градиентный метод для $(L_0, L_1)$-гладких $\rho$-квазар выпуклых функций

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ – $(L_0, L_1)$-гладкая (4) $\rho$-квазар выпуклая относительно любого решения $x_*$ функция (см. определение 2, $\mu = 0$):

$$f(x_*) \geqslant f(x) + \frac{1}{\rho}\langle \nabla f(x), x_* - x\rangle, \quad \rho \in (0, 1]. \tag{8}$$

Если $f$ сильно квазар-выпукла (см. определение 2, $\mu > 0$), то решение $x_*$ единственно, и тогда $f$ сильно квазар-выпукла относительно любого решения. Отметим, что задачи с невыпуклыми гладкими (а значит, и $(L_0, L_1)$-гладкими) $\rho$-квазар-выпуклыми функциями нередко встречаются в приложениях (см. пример 4) [27].

Для решения задачи (2) исследуем градиентный метод вида

$$x_{k+1} = x_k - \eta_k \widetilde{\nabla} f(x_k) \tag{9}$$

с шагом

$$\eta_k = \frac{\alpha}{L_0 + L_1 \left( \|\widetilde{\nabla} f(x_k)\| + \Delta \right)}, \tag{10}$$

где $\alpha \in (0, 1]$, а $\widetilde{\nabla} f(x_k)$ обозначает неточный градиент функции $f$ в точке $x_k$.

Важно отметить, что, несмотря на использование приближенных значений градиента в алгоритме, в теоретическом анализе предполагается существование точного градиента.

Докажем несколько лемм, необходимых для обоснования оценок скорости сходимости метода.

*Лемма 1* [10, Lemma A3]. *Пусть функция $f(x)$ $(L_0, L_1)$-гладкая (4). Тогда $\forall x, y \in \mathbb{R}^n$ при условии $\|x - y\| \leqslant \frac{1}{L_1}$ справедливо неравенство*

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{L_0 + L_1 \|\nabla f(x)\|}{2} \|y - x\|^2. \tag{11}$$

*Доказательство.* Пусть $g(t) = f(x + t(y - x))$ при $t \in [0, 1]$. Тогда

$$f(y) - f(x) = g(1) - g(0) = \int\limits_0^1 g'(t) dt = \int\limits_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle \, dt.$$

Добавим и вычтем $\langle \nabla f(x), y - x \rangle$ под интегралом:

$$f(y) - f(x) = \int\limits_0^1 \langle \nabla f(x), y - x \rangle \, dt + \int\limits_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \, dt.$$

Вычисляя первый интеграл и применяя неравенство Коши–Буняковского для выражения под вторым интегралом, получим следующее неравенство:

$$f(y) - f(x) \leqslant \langle \nabla f(x), y - x \rangle + \int\limits_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\| \|y - x\| dt.$$

Учитывая, что

$$\|\nabla f(y) - \nabla f(x)\| \leqslant (L_0 + L_1 \|\nabla f(x)\|) \, \|y - x\|,$$

имеем неравенство

$$f(y)-f(x)\leqslant\langle\nabla f(x),y-x\rangle+\|y-x\|\int\limits_0^1\left(L_0+L_1\|\nabla f(x)\|\right)\|y-x\|tdt,$$

откуда следует

$$f(y)\leqslant f(x)+\langle\nabla f(x),y-x\rangle+\frac{L_0+L_1\|\nabla f(x)\|}{2}\|y-x\|^2.$$

*Замечание 1.* При наличии аддитивной неточности в градиенте неравенство (11) принимает вид

$$\begin{aligned}f(y)\leqslant f(x)&+\langle\widetilde{\nabla}f(x),y-x\rangle+\\&+\frac{L_0+L_1(\|\widetilde{\nabla}f(x)\|+\Delta)}{2}\|y-x\|^2+\Delta\|y-x\|.\end{aligned}\tag{12}$$

*Лемма 2. Пусть $f(x)$ – $(L_0,L_1)$-гладкая функция (4). Предположим, что $\|\nabla f(x_k)\|>\Delta$. Тогда на итерациях алгоритма (9) с шагом (10) значения функции монотонно убывают $f(x_{k+1})\leqslant f(x_k)$, причем справедливо неравенство*

$$f(x_k)-f(x_{k+1})\geqslant\frac{\alpha(2-\alpha)}{2}\frac{(\|\nabla f(x_k)\|-\Delta)^2}{L_0+L_1\|\nabla f(x_k)\|+2L_1\Delta}.\tag{13}$$

*Доказательство.* Положим в неравенстве (11) $y=x_{k+1}$, $x=x_k$. Отметим, что $\|x_{k+1}-x_k\|\leqslant\frac{1}{L_1}$. Тогда

$$\begin{aligned}&f(x_{k+1})-f(x_k)\leqslant\\&\leqslant\frac{-\alpha\langle\nabla f(x_k),\widetilde{\nabla}f(x_k)\rangle}{L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)}+\frac{L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)}{2\left(L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)\right)^2}\alpha^2\|\widetilde{\nabla}f(x_k)\|^2=\\&=\frac{\alpha^2\|\widetilde{\nabla}f(x_k)\|^2-2\alpha\langle\nabla f(x_k),\widetilde{\nabla}f(x_k)\rangle}{2\left(L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)\right)}=\\&=\frac{\alpha^2\|\widetilde{\nabla}f(x_k)\|^2-2\alpha\langle\nabla f(x_k),\widetilde{\nabla}f(x_k)\rangle+\|\nabla f(x_k)\|^2-\|\nabla f(x_k)\|^2}{2\left(L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)\right)}=\\&=\frac{\|\alpha\widetilde{\nabla}f(x_k)-\nabla f(x_k)\|^2-\|\nabla f(x_k)\|^2}{2\left(L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)\right)}.\end{aligned}$$

Следовательно,

$$f(x_k)-f(x_{k+1})\geqslant\frac{\|\nabla f(x_k)\|^2-\|\alpha\widetilde{\nabla}f(x_k)-\nabla f(x_k)\|^2}{2\left(L_0+L_1\left(\|\widetilde{\nabla}f(x_k)\|+\Delta\right)\right)}.$$

Воспользуемся следующим соотношением для $\alpha \in (0;1]$, $a, b \in \mathbb{R}^n$:

$$\|\alpha a - b\| = \|\alpha(a-b) + (\alpha - 1)b\| \leqslant \alpha\|a-b\| + (1-\alpha)\|b\|.$$

$$\begin{aligned}
& f(x_k) - f(x_{k+1}) \geqslant \\
& \geqslant \frac{\|\nabla f(x_k)\|^2 - \left(\alpha\|\widetilde{\nabla} f(x_k) - \nabla f(x_k)\| + (1-\alpha)\|\nabla f(x_k)\|\right)^2}{2\left(L_0 + L_1\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)\right)} \geqslant \\
& \geqslant \frac{\|\nabla f(x_k)\|^2 - (\alpha\Delta + (1-\alpha)\|\nabla f(x_k)\|)^2}{2\left(L_0 + L_1\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)\right)} = \\
& = \frac{\alpha\left(\|\nabla f(x_k)\| - \Delta\right)\left((2-\alpha)(\|\nabla f(x_k)\| - \Delta) + 2\Delta\right)}{2\left(L_0 + L_1\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)\right)} > \\
& > \frac{\alpha(2-\alpha)}{2} \frac{(\|\nabla f(x_k)\| - \Delta)^2}{L_0 + L_1\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} \geqslant \\
& \geqslant \frac{\alpha(2-\alpha)}{2} \frac{(\|\nabla f(x_k)\| - \Delta)^2}{L_0 + L_1\|\nabla f(x_k)\| + 2L_1\Delta}.
\end{aligned}$$

*Лемма 3. Пусть $f$ – $(L_0, L_1)$-гладкая функция (4), удовлетворяющая условию Поляка–Лоясиевича (1), а $x_*$ – ближайшее к $x_{k+1}$ точное решение задачи (2). Тогда для градиентного метода вида (9) с шагом (10) справедливо неравенство*

$$\|x_{k+1} - x_*\| \leqslant \sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}. \tag{14}$$

*Доказательство.* Как известно ([29, Appendix A]), из условия Поляка–Лоясиевича следует (для ближайшего точного решения $x_*$), что функция имеет квадратичный рост:

$$\frac{\mu}{2}\|x_{k+1} - x_*\|^2 \leqslant f(x_{k+1}) - f^*.$$

С учетом монотонности функции $f$ неравенство принимает вид

$$\frac{\mu}{2}\|x_{k+1} - x_*\|^2 \leqslant f(x_0) - f^*,$$

откуда вытекает искомое неравенство

$$\|x_{k+1} - x_*\| \leqslant \sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}.$$

Введем обозначение $g_k = \|\nabla f(x_k)\| - \Delta$. Тогда из леммы 2 вытекает неравенство

$$f(x_k) - f(x_{k+1}) > \frac{\alpha(2-\alpha)}{2} \frac{g_k^2}{L_0 + L_1 g_k + 3L_1\Delta}. \tag{15}$$

Предположим, что $g_k \geqslant 3\Delta$. Будем исследовать сходимость метода в зависимости от величины $g_k$. Возможны два случая: $g_k > \frac{L_0}{L_1}$ и $g_k \leqslant \frac{L_0}{L_1}$.

При $g_k > \frac{L_0}{L_1}$ из (15) имеем

$$\begin{aligned} f(x_k) - f(x_{k+1}) &> \frac{\alpha(2-\alpha)}{2} \frac{g_k^2}{L_1 g_k + L_1 g_k + L_1 g_k} > \\ &> \frac{\alpha(2-\alpha)}{2} \frac{g_k}{3L_1} = \frac{\alpha(2-\alpha)}{6L_1} g_k. \end{aligned} \tag{16}$$

Если же $g_k \leqslant \frac{L_0}{L_1}$, то из (15) следует

$$\begin{aligned} f(x_k) - f(x_{k+1}) &> \frac{\alpha(2-\alpha)}{2} \frac{g_k^2}{L_0 + L_0 + L_0} > \\ &> \frac{\alpha(2-\alpha)}{2} \frac{g_k^2}{3L_0} = \frac{\alpha(2-\alpha)}{6L_0} g_k^2. \end{aligned} \tag{17}$$

Для такого класса задач справедлива следующая теорема.

Т е о р е м а 1. *Пусть $f$ – $\rho$-квазар-выпуклая относительно всякого точного решения $(L_0, L_1)$-гладкая функция, удовлетворяющая условию Поляка–Лоясиевича (1). Пусть также справедливы неравенства $\|\widetilde{\nabla} f(x_k)\| \geqslant 5\Delta$ и $\mu < \min\left\{\frac{6L_0}{\alpha(2-\alpha)}; \frac{72L_1^2(f(x_0)-f^*)}{\alpha^2(2-\alpha)^2\rho^2}\right\}$. Тогда для алгоритма (9) с шагом (10) справедливо неравенство*

$$\begin{aligned} & f(x_{k+1}) - f^* \leqslant \\ & \leqslant \max\left\{1 - \frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}}; 1 - \frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{k+1} (f(x_0) - f^*) + \\ & + \Delta \max\left\{\frac{1}{L_1}; \frac{\Delta}{L_0}\right\} \max\left\{\frac{L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}}{\rho}; \frac{L_0}{\mu}\right\}. \end{aligned} \tag{18}$$

Д о к а з а т е л ь с т в о. Рассмотрим сначала случай $g_k > \frac{L_0}{L_1}$. Из квазар-выпуклости функции $f(x)$ при $x = x_k$ вытекает неравенство

$$\begin{aligned} f(x_k) - f^* &\leqslant \frac{1}{\rho}\langle \nabla f(x_k), x_k - x_* \rangle \leqslant \frac{1}{\rho}\|\nabla f(x_k)\|\|x_k - x_*\| = \\ &= \frac{1}{\rho}(g_k + \Delta)\|x_k - x_*\|. \end{aligned} \tag{19}$$

Применяя к (19) лемму 3, получим неравенство

$$f(x_k) - f^* \leqslant \frac{1}{\rho}\sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}(g_k + \Delta),$$

откуда

$$g_k \geqslant \frac{\rho(f(x_k) - f^*)}{\sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}} - \Delta. \tag{20}$$

Из неравенств (16) и (20) вытекает оценка

$$\begin{aligned} f(x_{k+1}) - f^* &\leqslant f(x_k) - f^* - \frac{\alpha(2-\alpha)}{6L_1}\left(\frac{\rho(f(x_k) - f^*)}{\sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}} - \Delta\right) = \\ &= \left(1 - \frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}}\right)\left(f(x_k) - f^*\right) + \frac{\alpha(2-\alpha)\Delta}{6L_1}. \end{aligned} \tag{21}$$

При $g_k \leqslant \frac{L_0}{L_1}$ из (17) следует, что

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leqslant -\frac{\alpha(2-\alpha)}{6L_0}g_k^2 = -\frac{\alpha(2-\alpha)}{6L_0}(\|\nabla f(x_k)\| - \Delta)^2 \leqslant \\ &\leqslant -\frac{\alpha(2-\alpha)}{6L_0}\left(\frac{1}{2}\|\nabla f(x_k)\|^2 - \Delta^2\right) = \\ &= -\frac{\alpha(2-\alpha)}{12L_0}\|\nabla f(x_k)\|^2 + \frac{\alpha(2-\alpha)\Delta^2}{6L_0} = \\ &= -\frac{\alpha(2-\alpha)\mu}{6L_0}\left(f(x_k) - f^*\right) + \frac{\alpha(2-\alpha)\Delta^2}{6L_0}. \end{aligned}$$

Таким образом,

$$f(x_{k+1}) - f^* \leqslant \left(1 - \frac{\alpha(2-\alpha)\mu}{6L_0}\right)\left(f(x_k) - f^*\right) + \frac{\alpha(2-\alpha)\Delta^2}{6L_0}. \tag{22}$$

Из неравенств (21) и (22) следует справедливость неравенства

$$\begin{aligned} f(x_{k+1}) - f^* \leqslant \max\left\{1 - \frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0) - f^*\right)}}; 1 - \frac{\alpha(2-\alpha)\mu}{6L_0}\right\}\left(f(x_k) - f^*\right) + \\ + \max\left\{\frac{\alpha(2-\alpha)\Delta}{6L_1}; \frac{\alpha(2-\alpha)\Delta^2}{6L_0}\right\}, \end{aligned}$$

или

$$f(x_{k+1})-f^*\leqslant\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}(f(x_k)-f^*)+$$

$$+\frac{\alpha(2-\alpha)\Delta}{6}\max\left\{\frac{1}{L_1};\frac{\Delta}{L_0}\right\}\leqslant$$

$$\leqslant\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{k+1}(f(x_0)-f^*)+$$

$$+\frac{\alpha(2-\alpha)\Delta}{6}\max\left\{\frac{1}{L_1};\frac{\Delta}{L_0}\right\}\sum_{i=0}^{k}\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};\right.$$

$$\left.1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{i}\leqslant$$

$$\leqslant\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{k+1}(f(x_0)-f^*)+$$

$$+\frac{\alpha(2-\alpha)\Delta}{6}\max\left\{\frac{1}{L_1};\frac{\Delta}{L_0}\right\}\frac{1}{1-\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}(f(x_0)-f^*)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}}.$$

При положительных $a,b$ верна цепочка равенств

$$\frac{1}{1-\max\{1-a;1-b\}}=\frac{1}{1-(1-\min\{a;b\})}=\frac{1}{\min\{a;b\}}=\max\left\{\frac{1}{a};\frac{1}{b}\right\},$$

следовательно,

$$f(x_{k+1})-f^*\leqslant$$

$$\leqslant\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{k+1}(f(x_0)-f^*)+$$

$$+\frac{\alpha(2-\alpha)\Delta}{6}\max\left\{\frac{1}{L_1};\frac{\Delta}{L_0}\right\}\frac{6}{\alpha(2-\alpha)}\max\left\{\frac{L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}}{\rho};\frac{L_0}{\mu}\right\}\leqslant$$

$$\leqslant\max\left\{1-\frac{\alpha(2-\alpha)\rho}{6L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}};1-\frac{\alpha(2-\alpha)\mu}{6L_0}\right\}^{k+1}(f(x_0)-f^*)=$$

$$+\Delta\max\left\{\frac{1}{L_1};\frac{\Delta}{L_0}\right\}\max\left\{\frac{L_1\sqrt{\frac{2}{\mu}\left(f(x_0)-f^*\right)}}{\rho};\frac{L_0}{\mu}\right\}.$$

*Замечание 2.* Количество итераций алгоритма (9), для которых справедливо неравенство $\|\widetilde{\nabla} f(x_k)\| \geqslant 5\Delta$, можно оценить как

$$N \leqslant \left\lceil \frac{6L_1^2(f(x_0) - f^*)}{\alpha(2-\alpha)L_0} \right\rceil + \left\lceil \frac{2L_0(f(x_0) - f^*)}{3\alpha(2-\alpha)\Delta^2} \right\rceil .$$

*Доказательство.* Оценим число итераций, когда $g_k > \frac{L_0}{L_1}$. Из (16) следует неравенство

$$f(x_k) - f(x_{k+1}) > \frac{\alpha(2-\alpha)L_0}{6L_1^2}.$$

Пусть $i_1, \ldots, i_m$, $p = \overline{1,m}$ – индексы таких итераций, для которых $g_{i_p} > \frac{L_0}{L_1}$. Тогда после суммирования верно неравенство

$$f(x_0) - f^* \geqslant f(x_{i_1}) - f(x_{i_m}) \geqslant \frac{\alpha(2-\alpha)L_0}{6L_1^2} m,$$

откуда

$$m \leqslant \left\lceil \frac{6L_1^2(f(x_0) - f^*)}{\alpha(2-\alpha)L_0} \right\rceil, \quad L_0 \neq 0.$$

Оценим теперь количество итераций, для которых $g_k \leqslant \frac{L_0}{L_1}$. Принимая во внимание, что $g_k^2 \geqslant 9\Delta^2$, получаем из неравенства (17) оценку $f(x_k) - f(x_{k+1}) > \frac{9\alpha\,(2-\alpha)\,\Delta^2}{6L_0}$. Пусть $j_1, \ldots, j_n$, $l = \overline{1,n}$ — индексы итераций, для которых $g_{j_l} \leqslant \frac{L_0}{L_1}$. После суммирования имеем неравенство

$$f(x_0) - f^* \geqslant f(x_{j_1}) - f(x_{j_n}) \geqslant \frac{9\alpha(2-\alpha)\Delta^2}{6L_0} n,$$

откуда следует ограничение на число таких итераций

$$n \leqslant \left\lceil \frac{2L_0(f(x_0) - f^*)}{3\alpha(2-\alpha)\Delta^2} \right\rceil .$$

*Замечание 3.* Неравенство (21) указывает на сходимость метода (9) с шагом (10) со скоростью, близкой к скорости геометрической прогрессии, при достаточно больших значениях норм градиентов, т.е. $g_k > \frac{L_0}{L_1}$.

*Замечание 4.* Будем считать, что если $\|\widetilde{\nabla} f(x_k)\| < 5\Delta$, то приемлемое качество решения уже достигнуто. Действительно, если $\|\widetilde{\nabla} f(x_k)\| < 5\Delta$, то $\|\nabla f(x_k)\| < 6\Delta$. Тогда из условия Поляка–Лоясиевича (1) следует

$$f(x_k) - f^* \leqslant \frac{1}{2\mu}\|\nabla f(x)\|^2 < \frac{18\Delta^2}{\mu}.$$

*Замечание 5.* Поскольку количество итераций, для которых возможно выполнение условия $g_k > \frac{L_0}{L_1}$, ограничено, начиная с некоторого номера $N$ норма градиента будет меньше, чем $\frac{L_0}{L_1}$. Тогда можно утверждать о сходимости метода (9) с шагом (10) на классе задач с условием Поляка–Лоясиевича (без требования $\rho$-квазар-выпуклости) начиная с этого номера. Таким образом, начиная с некоторого номера справедлива альтернатива:

1) либо из неравенства (22) по рекурсии следует

$$f(x_{k+1}) - f^* \leqslant \left(1 - \frac{\alpha(2-\alpha)\mu}{6L_0}\right)^k (f(x_N) - f^*) + \frac{\Delta^2}{\mu},$$

2) либо приемлемая точность решения уже достигнута.

## 4. Адаптивный градиентный метод для квазар-выпуклых $(0, M)$-гладких функций

Рассмотрим случай $L_0 = 0$, соответствующий, например, задаче логистической регрессии. Для удобства введем обозначение $M = L_1$. Тогда условие (4) принимает вид

$$\|\nabla f(y) - \nabla f(x)\| \leqslant M\|\nabla f(x)\|\|y - x\| \ \forall x, y \in \mathbb{R}^n \text{ таких, что } \|y - x\| \leqslant \frac{1}{M}. \tag{23}$$

Далее исследуется адаптивный вариант градиентного метода с использованием неточной информации о градиенте на классе $\rho$-квазар выпуклых функций.

Пусть $f$ – $(0, M)$-гладкая $\rho$-квазар-выпуклая функция (в отличие от предыдущего раздела уже не обязательно относительно всякого точного решения), причем $\rho \in \left(\frac{1}{2}; 1\right]$ . Рассмотрим следующий адаптивный вариант градиентного метода с использованием неточных градиентов на итерациях.

*Алгоритм 1* (адаптивный градиентный метод для $(0, M)$-гладких задач).

1. Вход $x_0 \in \mathbb{R}^n$, $\Delta > 0$, $0 < M_0 < 2M$, $\alpha > 0$.
2. $k = 0$.
3. Повторять
4. $\quad M_{k+1} = \max\{M_0/2, M_k/2\}$,
5. $\quad \eta_k = \frac{\alpha}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}$,
6. $\quad x_{k+1} = x_k - \eta_k \widetilde{\nabla} f(x_k)$,
7. $\quad$ Если $f(x_k) - f(x_{k+1}) \geqslant \frac{\alpha(2-\alpha)}{4M_{k+1}}\left(\|\widetilde{\nabla} f(x_k)\| - 2\Delta\right)$ и
$$f(x_k) - f(x_{k+1}) \geqslant \alpha\left(1 - \frac{\alpha}{2}\right)\frac{\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} - \frac{\Delta\alpha}{M_{k+1}},$$
8. $\qquad k := k + 1$.
9. $\quad$ Иначе
10. $\qquad M_{k+1} = 2M_{k+1}$ и перейти к 5.
11. Выход $\widehat{x} = x_{k+1}$.

Докажем несколько вспомогательных утверждений для $(0, M)$-гладких функций. В частности, они обоснуют корректность алгоритма 1, т.е. выполнение для $(0, M)$-гладких функций при достаточно больших $M_{k+1}$ критерия выхода из итерации согласно п. 7 алгоритма 1.

*Лемма 4. Пусть $f(x)$ – $(0, M)$-гладкая функция. Предположим, что $\|\widetilde{\nabla} f(x_k)\| > 5\Delta$ и $\{M_i\}_{i=0}^{k}$ – последовательность, причем $M_{i+1} \geqslant M$. Тогда на итерациях алгоритма*

$$x_{k+1} = x_k - \frac{\alpha}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} \widetilde{\nabla} f(x_k), \quad \alpha \in (0; 1]$$

*значения функции монотонно не возрастают $f(x_{k+1}) \leqslant f(x_k)$, причем справедливо неравенство*

$$f(x_k) - f(x_{k+1}) \geqslant \frac{\alpha(2-\alpha)}{2} \frac{(\|\widetilde{\nabla} f(x_k)\| - 2\Delta)^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)}. \tag{24}$$

*Доказательство.* Из неравенства (11) при $L_0 = 0$, $L_1 = M$ следует неравенство

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{M\|\nabla f(x)\|}{2}\|y - x\|^2.$$

Положим в этом неравенстве $y = x_{k+1}$, $x = x_k$. Тогда

$$\begin{aligned}
&f(x_{k+1}) - f(x_k) \leqslant \\
&\leqslant \frac{-\alpha\langle \nabla f(x_k), \widetilde{\nabla} f(x_k)\rangle}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} + \frac{M\|\nabla f(x_k)\|}{2M_{k+1}^2\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)^2}\alpha^2\|\widetilde{\nabla} f(x_k)\|^2 \leqslant \\
&\leqslant \frac{-\alpha\langle \nabla f(x_k), \widetilde{\nabla} f(x_k)\rangle}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} + \frac{M\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}{2M\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} \frac{\alpha^2\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} = \\
&= \frac{\alpha^2\|\widetilde{\nabla} f(x_k)\|^2 - 2\alpha\langle \nabla f(x_k), \widetilde{\nabla} f(x_k)\rangle}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} = \\
&= \frac{\alpha^2\|\widetilde{\nabla} f(x_k)\|^2 - 2\alpha\langle \nabla f(x_k), \widetilde{\nabla} f(x_k)\rangle + \|\nabla f(x_k)\|^2 - \|\nabla f(x_k)\|^2}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} = \\
&= \frac{\|\alpha\widetilde{\nabla} f(x_k) - \nabla f(x_k)\|^2 - \|\nabla f(x_k)\|^2}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}.
\end{aligned}$$

Следовательно,

$$f(x_k) - f(x_{k+1}) \geqslant \frac{\|\nabla f(x_k)\|^2 - \|\alpha\widetilde{\nabla} f(x_k) - \nabla f(x_k)\|^2}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}.$$

Воспользуемся следующим соотношением для $\alpha \in (0;1]$, $a, b \in \mathbb{R}^n$:

$$\|\alpha a - b\| = \|\alpha(a-b) + (\alpha - 1)b\| \leqslant \alpha\|a-b\| + (1-\alpha)\|b\|.$$

$$\begin{aligned}
&f(x_k) - f(x_{k+1}) \geqslant \\
&\geqslant \frac{\|\nabla f(x_k)\|^2 - \left(\alpha\|\widetilde{\nabla} f(x_k) - \nabla f(x_k)\| + (1-\alpha)\|\nabla f(x_k)\|\right)^2}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} \geqslant \\
&\geqslant \frac{\|\nabla f(x_k)\|^2 - (\alpha\Delta + (1-\alpha)\|\nabla f(x_k)\|)^2}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} = \\
&= \frac{\alpha\left(\|\nabla f(x_k)\| - \Delta\right)\left((2-\alpha)(\|\nabla f(x_k)\| - \Delta) + 2\Delta\right)}{2M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} > \\
&> \frac{\alpha(2-\alpha)}{2}\frac{(\|\nabla f(x_k)\| - \Delta)^2}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)} \geqslant \\
&\geqslant \frac{\alpha(2-\alpha)}{2}\frac{\left(\|\widetilde{\nabla} f(x_k)\| - 2\Delta\right)^2}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}.
\end{aligned}$$

Пусть $\{M_i\}_{i=0}^k$ – последовательность положительных чисел, удовлетворяющих предположениям предыдущей теоремы. Тогда из (24) при $\|\widetilde{\nabla} f(x_k)\| \geqslant 5\Delta$ вытекает

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\geqslant \frac{\alpha(2-\alpha)(\|\widetilde{\nabla} f(x_k)\| - 2\Delta)^2}{2\left(M_{k+1}(\|\widetilde{\nabla} f(x_k)\| - 2\Delta) + 3M_{k+1}\Delta\right)} \geqslant \\
&\geqslant \frac{\alpha(2-\alpha)}{4M_{k+1}}\left(\|\widetilde{\nabla} f(x_k)\| - 2\Delta\right).
\end{aligned} \tag{25}$$

*Лемма 5. Пусть $f(x)$ – $(0, M)$-гладкая функция, $\{M_i\}_{i=0}^k$ – некоторая последовательность, причем $M_{i+1} \geqslant M$. Тогда для алгоритма*

$$x_{k+1} = x_k - \frac{\alpha}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)}\widetilde{\nabla} f(x_k), \quad \alpha \in (0;1]$$

*справедливо неравенство*

$$\frac{\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} \leqslant \frac{2}{\alpha(2-\alpha)}(f(x_k) - f(x_{k+1})) + \frac{2\Delta}{M_{k+1}(2-\alpha)}. \tag{26}$$

*Доказательство.* Из неравенства (12) при $L_0 = 0$, $L_1 = M$ следует

$$f(y) \leqslant f(x) + \langle \widetilde{\nabla} f(x), y - x \rangle + \frac{M(\|\widetilde{\nabla} f(x)\| + \Delta)}{2}\|y - x\|^2 + \Delta\|y - x\|.$$

Положим в этом неравенстве $y = x_{k+1}$ и $x = x_k$:

$$\begin{aligned}
f(x_{k+1}) \leqslant f(x_k) &+ \langle \widetilde{\nabla} f(x_k), x_{k+1} - x_k \rangle + \\
&+ \frac{M(\|\widetilde{\nabla} f(x_k)\| + \Delta)}{2}\|x_{k+1} - x_k\|^2 + \Delta\|x_{k+1} - x_k\| \leqslant \\
&\leqslant f(x_k) - \frac{\alpha\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} + \frac{\alpha^2\|\widetilde{\nabla} f(x_k)\|^2}{2M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} + \\
&+ \frac{\Delta\alpha\|\widetilde{\nabla} f(x_k)\|}{M_{k+1}(\|\widetilde{\nabla} f(x)\| + \Delta)} \\
&= f(x_k) - \alpha\left(1 - \frac{\alpha}{2}\right)\frac{\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} + \frac{\Delta\alpha\|\widetilde{\nabla} f(x_k)\|}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)}.
\end{aligned}$$

Оценим слагаемое, содержащее $\Delta$:

$$\frac{\Delta\alpha\|\widetilde{\nabla} f(x_k)\|}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} \leqslant \frac{\frac{\Delta\alpha}{M_{k+1}}M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} = \frac{\Delta\alpha}{M_{k+1}}.$$

Таким образом, получаем неравенство

$$f(x_k) - f(x_{k+1}) \geqslant \alpha\left(1 - \frac{\alpha}{2}\right)\frac{\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} - \frac{\Delta\alpha}{M_{k+1}},$$

или в другом виде

$$\frac{\|\widetilde{\nabla} f(x_k)\|^2}{M_{k+1}(\|\widetilde{\nabla} f(x_k)\| + \Delta)} \leqslant \frac{2}{\alpha(2 - \alpha)}(f(x_k) - f(x_{k+1})) + \frac{2\Delta}{M_{k+1}(2 - \alpha)}.$$

Доказанные выше неравенства (25) и (26) означают, что при $M_{k+1} \geqslant M$ автоматически выполнен критерий выхода из итерации алгоритма 1.

*Лемма 6. Пусть $f(x)$ – $\rho$-квазар-выпуклая функция при $\rho \in \left(\frac{1}{2}; 1\right]$, $\{M_i\}_{i=0}^k$ – некоторая последовательность положительных чисел из алгоритма 1. Тогда при $\alpha \leqslant \frac{2\rho - 1}{\rho}$ справедливо неравенство*

$$\|x_{k+1} - x_*\| \leqslant \|x_k - x_*\| + \Delta\eta_k + \frac{1}{M_{k+1}}. \tag{27}$$

*Доказательство.* Поскольку $x_{k+1} = x_k - \eta_k\widetilde{\nabla} f(x_k)$, то верны следующие равенства:

$$\begin{aligned}
\|x_{k+1} - x_*\|^2 &= \|x_k - x_* - \eta_k\widetilde{\nabla} f(x_k)\|^2 = \\
&= \|x_k - x_*\|^2 - 2\eta_k\langle \widetilde{\nabla} f(x_k), x_k - x_* \rangle + \eta_k^2\|\widetilde{\nabla} f(x_k)\|^2.
\end{aligned}$$

Из квазар-выпуклости функции $f$ следует, что

$$\|x_{k+1}-x_*\|^2 \leqslant$$

$$\leqslant \|x_k-x_*\|^2 - 2\eta_k(\rho(f(x_k)-f^*)-\Delta\|x_k-x_*\|)+\eta_k^2\|\widetilde{\nabla} f(x_k)\|^2 \overset{(26)}{\leqslant}$$

$$\overset{(26)}{\leqslant} \|x_k-x_*\|^2 - 2\eta_k(\rho(f(x_k)-f^*)-\Delta\|x_k-x_*\|)+$$

$$+\alpha\eta_k\left(\frac{2}{\alpha(2-\alpha)}(f(x_k)-f^*)+\frac{2\Delta}{M_{k+1}(2-\alpha)}\right)=$$

$$=\|x_k-x_*\|^2+\eta_k\left(\frac{2}{2-\alpha}-2\rho\right)(f(x_k)-f^*)+$$

$$+2\Delta\eta_k\left(\|x_k-x_*\|+\frac{\alpha}{M_{k+1}(2-\alpha)}\right).$$

Выберем параметр $\alpha$ так, чтобы $\frac{2}{2-\alpha}-2\rho\leqslant 0$. Поскольку $\alpha\in(0;1]$, получаем, что это возможно тогда и только тогда, когда $\rho\in\left(\frac{1}{2};1\right]$. В этом случае имеем:

$$\|x_{k+1}-x_*\|^2 \leqslant \|x_k-x_*\|^2+2\Delta\eta_k\|x_k-x_*\|+$$
$$+\Delta^2\eta_k^2+\frac{2\alpha\Delta\eta_k}{M_{k+1}(2-\alpha)}-\Delta^2\eta_k^2=$$
$$=(\|x_k-x_*\|+\Delta\eta_k)^2+\Delta\eta_k\left(\frac{2\alpha}{M_{k+1}(2-\alpha)}-\Delta\eta_k\right).$$

Рассмотрим функцию $y(t)=-t^2+\frac{2\alpha}{M_{k+1}(2-\alpha)}t$. В точке $t=\frac{\alpha}{M_{k+1}(2-\alpha)}$ она достигает максимального значения $\frac{\alpha^2}{M_{k+1}^2(2-\alpha)^2}$, причем при $\alpha\leqslant 1$ справедливо $\frac{\alpha^2}{M_{k+1}^2(2-\alpha)^2}\leqslant\frac{1}{M_{k+1}^2}$. Таким образом,

$$\|x_{k+1}-x_*\|^2\leqslant(\|x_k-x_*\|+\Delta\eta_k)^2+\frac{1}{M_{k+1}^2}.$$

Заметим, что при $a,b,c>0$ из неравенства $a^2\leqslant b^2+c^2$ вытекает $a^2\leqslant(b+c)^2$, откуда $a\leqslant b+c$. Следовательно,

$$\|x_{k+1}-x_*\|\leqslant\|x_k-x_*\|+\Delta\eta_k+\frac{1}{M_{k+1}}.$$

Отметим, что неравенства (25) и (26) заведомо выполняются, если $M_{k+1}\geqslant\geqslant M$. Если $\|\widetilde{\nabla} f(x_k)\|\geqslant 5\Delta$ и для некоторого $M_0>0$ верно $M_i\geqslant\frac{M_0}{2}\ \forall i==\overline{1,k+1}$, то

$$\eta_k=\frac{\alpha}{M_{k+1}\left(\|\widetilde{\nabla} f(x_k)\|+\Delta\right)}\leqslant\frac{\alpha}{6M_{k+1}\Delta}\leqslant\frac{\alpha}{3M_0\Delta}.$$

*Теорема 2. Пусть $f(x)$ – $\rho$-квазар-выпуклая $(0, M)$-гладкая функция, где $\rho \in \left(\frac{1}{2}; 1\right]$, а $\alpha \in (0; 1]$ такое, что $\alpha \leqslant \frac{2\rho-1}{\rho}$. Тогда после $\left\lceil \frac{8M(f(x_0)-f^*)}{3\alpha(2-\alpha)\Delta} \right\rceil$ итераций алгоритма 1 найдется $k$, для которого либо $\|\widetilde{\nabla} f(x_k)\| < 5\Delta$, либо будет достигнуто точное решение.*

*Доказательство.* Из неравенства (25) с учетом того, что $\|\widetilde{\nabla} f(x_k)\| \geqslant \geqslant 5\Delta$, после суммирования по $k$ от 0 до $N-1$ вытекают соотношения

$$f(x_0) - f^* \geqslant f(x_0) - f(x_N) \geqslant \frac{3\alpha(2-\alpha)\Delta}{4} \sum_{k=0}^{N-1} \frac{1}{M_{k+1}},$$

откуда

$$\sum_{k=0}^{N-1} \frac{1}{M_{k+1}} \leqslant \frac{4(f(x_0) - f^*)}{3\alpha(2-\alpha)\Delta}.$$

Учитывая, что $M_i \leqslant 2M$, оценим количество итераций, для которых $\|\widetilde{\nabla} f(x_k)\| \geqslant 5\Delta$. Имеем $\frac{N}{2M} \leqslant \frac{4(f(x_0)-f^*)}{3\alpha(2-\alpha)\Delta}$, или

$$N \leqslant \left\lceil \frac{8M(f(x_0) - f^*)}{3\alpha(2-\alpha)\Delta} \right\rceil.$$

*Замечание 6.* Будем считать, что если $\|\widetilde{\nabla} f(x_k)\| < 5\Delta$, то приемлемое качество решения уже достигнуто. Действительно, из условия $\rho$-квазар-выпуклости функции $f$ и неравенства (27) следует

$$\begin{aligned} f(x_k) - f^* &\leqslant \frac{1}{\rho}\left(\|\widetilde{\nabla} f(x_k)\| + \Delta\right)\|x_k - x_*\| \leqslant \\ &\leqslant \frac{6\Delta}{\rho}\left(\|x_0 - x_*\| + \frac{(\alpha+6)N}{3M_0}\right). \end{aligned}$$

Отсюда следует, что при небольшом количестве итераций можно гарантировать, что качество решения находится в пределах $O(\Delta)$. В случае больших значений $N$ этого утверждать нельзя. Отметим, что если целевая функция удовлетворяет условию Поляка–Лоясиевича, то согласно замечанию 4 малая норма неточного градиента гарантирует достижение приемлемого качества решения.

В частном случае, при $\Delta = 0$, верно следующее утверждение.

*Теорема 3. Пусть $f(x)$ – $\rho$-квазар-выпуклая $(0, M)$-гладкая функция, где $\rho \in \left(\frac{1}{2}; 1\right]$, а $\alpha \in (0; 1]$ такое, что $\alpha \leqslant \frac{2\rho-1}{\rho}$. Тогда алгоритм 1 в случае $\Delta = 0$ сходится со скоростью геометрической прогрессии.*

*Доказательство.* Рассмотрим случай $\Delta = 0$, т.е. $\widetilde{\nabla} f(x) = \nabla f(x)$ для всякого $x$. Тогда из п. 7 листинга алгоритма 1 следует, что

$$f(x_k) - f(x_{k+1}) \geqslant \frac{\alpha(2-\alpha)}{2M_{k+1}} \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_k)\|} = \frac{(2-\alpha)\eta_k}{2}\|\nabla f(x_k)\|^2. \tag{28}$$

Также в этом случае справедливы соотношения

$$\begin{aligned}\|x_{k+1}-x_*\|^2 &= \|x_k-x_*\|^2 - 2\eta_k\langle\nabla f(x_k), x_k-x_*\rangle + \eta_k^2\|\nabla f(x_k)\|^2 \leqslant\\ &\leqslant \|x_k-x_*\|^2 - 2\eta_k\rho(f(x_k)-f^*) + \eta_k^2\|\nabla f(x_k)\|^2 \leqslant\\ &\leqslant \|x_k-x_*\|^2 - 2\eta_k\rho(f(x_k)-f^*) + \frac{2\eta_k}{(2-\alpha)}(f(x_k)-f^*) =\\ &= \|x_k-x_*\|^2 + \eta_k(f(x_k)-f^*)\left(\frac{2}{2-\alpha}-2\rho\right) \leqslant\\ &\leqslant \|x_k-x_*\|^2 \leqslant \|x_0-x_*\|^2 = R^2,\end{aligned}$$

где $\alpha \in (0;1]$, $\rho \in \left(\frac{1}{2};1\right]$. Далее, из квазар-выпуклости функции $f$ имеем

$$f(x_k)-f^* \leqslant \frac{1}{\rho}\|\nabla f(x_k)\|\|x_k-x_*\| \leqslant \frac{1}{\rho}\|\nabla f(x_k)\|\|x_0-x_*\| = \frac{R}{\rho}\|\nabla f(x_k)\|.$$

Вспомним, что ввиду лемм 4 и 5 при $M_{k+1} \geqslant M$ заведомо выполнен критерий выхода из итерации алгоритма 1 (п. 7 листинга). Тогда для всякого $k \geqslant 0$ верно $M_{k+1} \leqslant 2M$, и из (28) вытекает

$$\begin{aligned}f(x_{k+1})-f^* &\leqslant f(x_k)-f^* - \frac{\alpha(2-\alpha)}{2M_{k+1}}\|\nabla f(x_k)\| \leqslant\\ &\leqslant f(x_k)-f^* - \frac{\alpha(2-\alpha)}{2M_{k+1}}\frac{\rho}{R}(f(x_k)-f^*) \leqslant\\ &\leqslant \left(1-\frac{\alpha(2-\alpha)\rho}{2M_{k+1}R}\right)(f(x_k)-f^*) \leqslant\\ &\leqslant \left(1-\frac{\alpha(2-\alpha)\rho}{4MR}\right)(f(x_k)-f^*) \leqslant\\ &\leqslant \left(1-\frac{\alpha(2-\alpha)\rho}{4MR}\right)^{k+1}(f(x_0)-f^*),\end{aligned}$$

что и завершает доказательство.

## 5. Вычислительные эксперименты

В данном разделе представлены результаты численных экспериментов, проведенных для оценки эффективности предложенных алгоритмов при минимизации логистической функции вида (29) и квазар-выпуклой функции вида (30).

### *5.1. Результаты тестирования адаптивного алгоритма 1*

Сравним предложенный алгоритм 1 с адаптивным универсальным градиентным методом из [30] на задаче минимизации логистической функции из примера 3:

$$f(x) = \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-a_i^\top x)), \quad a_i \in \mathbb{R}^n. \tag{29}$$

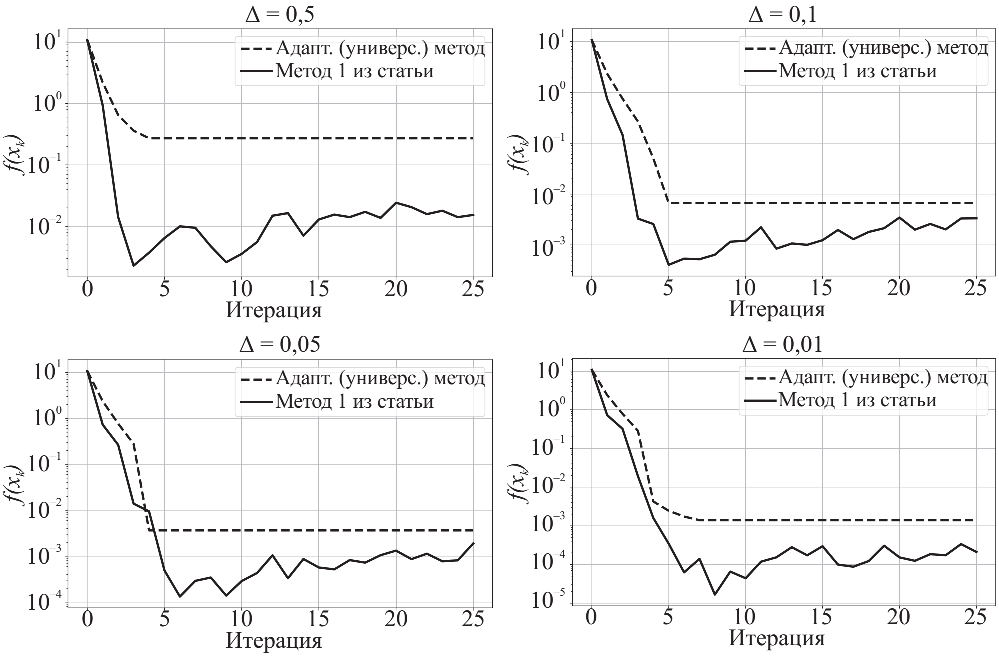


Рис. 1. Результаты алгоритма 1 и адаптивного универсального градиентного метода [30] для задачи логистической регрессии с (29) при $n = 10^3, m = 100$.

Известно, что логистическая функция (29) является выпуклой $L$-гладкой с константой $L = \frac{1}{m}\sum\limits_{i=1}^{m} \|a_i\|^2$, а также $(L_0, L_1)$-гладкой с параметрами $L_0 = 0$ и $L_1 = \max\limits_{1 \leqslant i \leqslant m} \|a_i\|$.

Алгоритмы запускались с параметрами $n = 1000$, $m = 100$ и начальной точкой $x_0 = (1, \ldots, 1) \in \mathbb{R}^n$. Параметр неточности градиента $\Delta$ принимал значения из множества $\{0{,}5; 0{,}1; 0{,}05; 0{,}01\}$. Векторы $\{a_i\}_{i=1}^m$ генерировались случайно из стандартного нормального распределения. На рис. 1 представлена динамика значений целевой функции в зависимости от номера итерации для сравниваемых алгоритмов. Результаты показывают, что предложенный алгоритм 1, использующий $(L_0, L_1)$-гладкость, демонстрирует более высокую эффективность по сравнению с адаптивным универсальным градиентным методом [30].

### *5.2. Пример невыпуклой задачи из класса квазар-выпуклых функций*

Представлено сравнение предложенного метода (9) с шагом (10) и градиентного метода с постоянным шагом размера 0,001 на примере квазар-выпуклой функции из [27]:

$$f(x) = h(\|x\|_2)\, g\left(\frac{x}{\|x\|_2}\right), \tag{30}$$

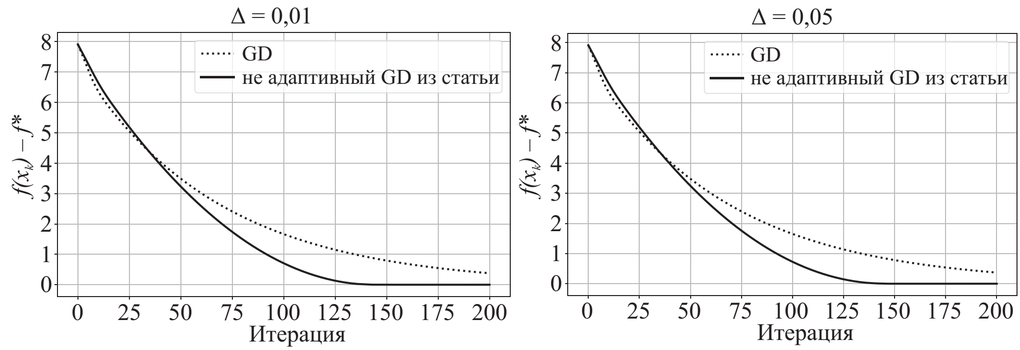


Рис. 2. Результаты применения метода (9) с шагом (10) и градиентного метода с постоянным шагом для минимизации функции (30).

где $h(t) = t^2$ и

$$g(x_1, x_2) = 1 + \frac{1}{4N} \sum_{i=1}^{N} a_i \sin^2(b_i x_1) + c_i \cos^2(d_i x_2).$$

В вычислительных экспериментах использовались последовательности $\{a_i\}_i$ и $\{c_i\}_i$, элементы которых независимо и равномерно распределены на отрезке $[0, 20)$. Аналогично, последовательности $\{b_i\}_i$ и $\{d_i\}_i$ независимо и равномерно распределены на отрезке $[-25, 25)$. Все методы инициализировались из начальной точки $(1, 1)$ с параметрами $N = 10$ и $\alpha = 0{,}01$. На рис. 2 представлена динамика изменения значений целевой функции в зависимости от номера итерации. Полученные результаты демонстрируют, что предложенный алгоритм (9) с выбором шага (10), основанным на $(L_0, L_1)$-гладкости целевой функции, обладает большей эффективностью по сравнению с градиентным методом с постоянным шагом.

## 6. Заключение

В работе предложен и проанализирован ряд методов градиентного типа, предназначенных для решения задач минимизации $(L_0, L_1)$-гладких функций при условии аддитивной неточности в значениях градиента на итерациях. Внимание было сосредоточено на исследовании сходимости методов для некоторых классов невыпуклых функций, таких как $\rho$-квазар-выпуклые функции и функции, удовлетворяющие условию Поляка–Лоясиевича.

В работе представлен обзор современного состояния исследований в этой области. Все ключевые определения и вспомогательные результаты адаптированы на случай $\Delta$-неточного градиента. Сформулированы и доказаны леммы, характеризующие влияние аддитивной неточности градиента на траекторию градиентного метода. Получена теоретическая оценка скорости сходимости градиентного метода на классе квазар-выпуклых функций, удовлетворяющих условию Поляка–Лоясиевича. В частности, показано, что обеспечивается сходимость метода с близкой к линейной скоростью (с точностью до параметров, связанных с погрешностями). Доказано, что начиная с некоторого номера итерации метод сходится в окрестность точного решения на классе функций с условием Поляка–Лоясиевича. Отдельное внимание уделено детальному исследованию случая $L_0 = 0$ (класс $(0, M)$-гладких функций), который представляет значительный интерес для задач машинного обучения, таких как обучение моделей логистической регрессии. Разработан адаптивный вариант градиентного метода, использующий неточные градиенты, для которого получена теоретическая оценка скорости сходимости. Показано, что в случае точных градиентов (при $\Delta = 0$) метод сходится со скоростью геометрической прогрессии. Эффективность предложенных методов подтверждена результатами численных экспериментов, проведенных как на задаче логистической регрессии, так и на некоторой невыпуклой квазар-выпуклой задаче.

В качестве направлений дальнейших исследований представляется интересным изучение ускоренных методов для $(L_0, L_1)$-гладких задач с неточным градиентом, а также разработка методов с адаптивной настройкой на параметры $L_0$ и $L_1$.

## СПИСОК ЛИТЕРАТУРЫ


1. *Zhang J., He T., Sra S., Jadbabaie A.* Why gradient clipping accelerates training: A theoretical justification for adaptivity // International Conference on Learning Representations. 2020. P. 1–14.
2. *Schmidt M., Le Roux N., Bach F.* Minimizing finite sums with the stochastic average gradient // Mathematical Programming. 2017. V. 162. P. 83–112.
3. *Johnson R., Zhang T.* Accelerating stochastic gradient descent using predictive variance reduction // Advances in neural information processing systems. 2013. No. 26. P. 315–323.
4. *Defazio A., Bach F., Lacoste-Julien S.* SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives // Advances in neural information processing systems. 2014. No. 2. P. 1646–1654.
5. *Nguyen L., Liu J., Scheinberg K., Takač M.* Sarah: A novel method for machine learning problems using stochastic recursive gradient // 34th International Conference on Machine Learning (ICML). 2014. V. 6. P. 4009–4023.
6. *Nguyen L., Scheinberg K., Takač M.* Inexact sarah algorithm for stochastic optimization // Optimization Methods and Software. 2021. V. 36. No. 1. P. 237–258.
7. *Beznosikov A. Takač M.* Random-reshuffled SARAH does not need full gradient computations // Optim Lett. 2024. V. 18. P. 727–749.
8. *Shi Z., Sadiev A., Loizou N., et al.* Ai-sarah: Adaptive and implicit stochastic recursive gradient methods // Transactions on Machine Learning Research. 2023. P. 1–40.
9. *Defazio A., Bottou L.* On the ineffectiveness of variance reduced optimization for deep learning // Advances in Neural Information Processing Systems. 2019. V. 32. P. 1753–1763.

10. *Zhang B., Jin J., Fang C., Wang L.* Improved Analysis of Clipping Algorithms for Non-convex Optimization // Advances in Neural Information Processing Systems. 2020. V. 19. P. 15511–15522.

11. *Chen Z., Zhou Y., Liang Y., Lu Z.* Generalized-smooth nonconvex optimization is as efficient as smooth nonconvex optimization // International Conference on Machine Learning (PMLR). 2023. P. 5396–5427.

12. *Zhao S.-Y., Xie Y.-P., Li W.-J.* On the convergence and improvement of stochastic normalized gradient descent // Science China Information Sciences. 2021. V. 64. P. 1–13.

13. *Faw M., Rout L., Caramanis C., Shakkottai S.* Beyond uniform smoothness: A stopped analysis of adaptive sgd // The Thirty Sixth Annual Conference on Learning Theory (PMLR). 2023. P. 89–160.

14. *Wang B., Zhang H., Ma Z., Chen W.* Convergence of adagrad for non-convex objectives: Simple proofs and relaxed assumptions // The Thirty Sixth Annual Conference on Learning Theory (PMLR). 2023. P. 161–190.

15. *Li H., Rakhlin A., Jadbabaie A.* Convergence of adam under relaxed assumptions // Advances in Neural Information Processing Systems. 2024. P. 1792–1804.

16. *Hubler F., Yang J., Li X., He N.* Parameter-agnostic optimization under relaxed smoothness // International Conference on Artificial Intelligence and Statistics (PMLR). 2024. P. 4861–4869.

17. *Pascanu R., Mikolov T., Bengio Y.* On the difficulty of training recurrent neural networks // International Conference on Machine Learning. 2013. V. 28. P. 1310–1318.

18. *Polyak B.* Introduction to Optimization // Optimization Software. 1987. 438 p.

19. *Koloskova A., Hendrikx H., Stich S.* Revisiting gradient clipping: Stochastic bias and tight convergence guarantees // International Conference on Machine Learning. 2023. P. 17343–17363.

20. *Takezawa Y., Bao H., Sato R. et al.* Polyak meets parameter-free clipped gradient descent // 2024. arXiv:2405.15010

21. *Li H., Qian J., Tian Y., Rakhlin A., Jadbabaie A.* Convex and non-convex optimization under generalized smoothness // Advances in Neural Information Processing Systems. 2024. V. 36. P. 2675–2686.

22. *Gorbunov E., Tupitsa N., Choudhury S., et al.* Methods for Convex $(L0, L1)$-Smooth Optimization: Clipping, Acceleration, and Adaptivity // 2024. arXiv:2409.14989

23. *Vankov D., Rodomanov A., Nedich A., et al.* Optimizing (L0, L1)-Smooth Functions by Gradient Methods // 2024. arXiv:2410.10800

24. *Lobanov A., Gasnikov A., Gorbunov E., Takáč M.* Linear Convergence Rate in Convex Setup is Possible! Gradient Descent Method Variants under (L0, L1)-Smoothness // 2024. arXiv:2412.17050

25. *Stonyakin F., Kuruzov I., Polyak B.* Stopping Rules for Gradient Methods for Nonconvex Problems with Additive Noise in Gradient // Journal of Optimization Theory and Applications. 2022. V. 198. P. 531–551.

26. *Wang J., Wibisono A.* Continuized Acceleration for Quasar Convex Functions in Non-Convex Optimization // 2023. 10.48550/arXiv.2302.07851

27. *Hermant J., Aujol J.F., Dossal C., Rondepierre A.* Study of the behaviour of Nesterov accelerated gradient in a non convex setting: the strongly quasar convex case // 2024. arXiv:2405.19809

28. *Hinder O., Sidford A., Sohoni N.* Near-Optimal Methods for Minimizing Star-Convex Functions and Beyond // 2019. arXiv.1906.11985.

29. *Karimi H., Nutini J., Schmidt M.* Linear convergence of gradient and proximal-gradient methods under the Polyak–Lojasiewicz condition // Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2016. P. 795–811.

30. *Stonyakin F., Tyurin A., Gasnikov A., et al.* Inexact Relative Smoothness and Strong Convexity for Optimization and Variational Inequalities by Inexact Model // 2024. arXiv preprint arXiv:2402.06319.

*Статья представлена к публикации членом редколлегии П.С. Щербаковым.*

Поступила в редакцию 03.03.2025

После доработки 20.05.2025

Принята к публикации 27.06.2025