КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ И МОДЕЛИРОВАНИЕ 2024 Т. 16 № 1 С. 105–122
DOI: 10.20537/2076-7633-2024-16-1-105-122

Ки&М

СПЕЦИАЛЬНЫЙ ВЫПУСК

УДК: 519.85

Субградиентные методы с шагом типа Б. Т. Поляка для задач минимизации квазивыпуклых функций с ограничениями-неравенствами и аналогами острого минимума

С. М. Пучинин[1,a], Е. Р. Корольков[1,b], Ф. С. Стонякин[1,2,c], М. С. Алкуса[1,d], А. А. Выгузов[1,e]

[1]Московский физико-технический институт (национальный исследовательский университет), Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., д. 9
[2]Крымский федеральный университет им. В. И. Вернадского, Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, д. 4

E-mail: [a] puchinin.sm@phystech.edu, [b] korolkov.er@phystech.edu, [c] fedyor@mail.ru, [d] mohammad.alkousa@phystech.edu, [e] vyguzov.aa@phystech.edu

Получено 06.12.2023, после доработки — 23.12.2023.
Принято к публикации 23.12.2023.

В работе рассмотрено два варианта понятия острого минимума для задач математического программирования с квазивыпуклой целевой функцией и ограничениями-неравенствами. Исследована задача описания варианта простого субградиентного метода с переключениями по продуктивным и непродуктивным шагам, для которого бы на классе задач с липшицевыми функциями можно было гарантировать сходимость со скоростью геометрической прогрессии ко множеству точных решений или его окрестности. При этом важно, чтобы для реализации метода не было необходимости знать параметр острого минимума, который обычно сложно оценить на практике. В качестве решения проблемы авторы предлагают использовать процедуру регулировки шага, аналогичную предложенной ранее Б. Т. Поляком. Однако при этом более остро по сравнению с классом задач без ограничений встает проблема знания точного значения минимума целевой функции. В работе описываются условия на погрешность этой информации, которые позволяют сохранить сходимость со скоростью геометрической прогрессии в окрестность множества точек минимума задачи. Рассмотрено два аналога понятия острого минимума для задач с ограничениями-неравенствами. В первом случае возникает проблема приближения к точному решению лишь до заранее выбранного уровня точности, при этом рассматривается случай, когда минимальное значение целевой функции неизвестно, вместо этого дано некоторое его приближение. Описаны условия на неточность минимума целевой функции, при которой все еще сохраняется сходимость к окрестности искомого множества точек со скоростью геометрической прогрессии. Второй рассматриваемый вариант острого минимума не зависит от желаемой точности задачи. Для него предложен несколько иной способ проверки продуктивности шага, позволяющий в случае точной информации гарантировать сходимость метода к точному решению со скоростью геометрической прогрессии. Доказаны оценки сходимости в условиях слабой выпуклости ограничений и некоторых ограничениях на выбор начальной точки, а также сформулирован результат-следствие для выпуклого случая, когда необходимость дополнительного предположения о выборе начальной точки пропадает. Для обоих подходов доказано убывание расстояния от текущей точки до множества решений с ростом количества итераций. Это, в частности, позволяет ограничить требования используемых свойств функций (липшицевость, острый минимум) лишь для ограниченного множества. Выполнены вычислительные эксперименты, в том числе для задачи проектирования механических конструкций.

Ключевые слова: субградиентный метод, липшицева функция, острый минимум, шаг Б. Т. Поляка, квазивыпуклая функция, слабовыпуклая функция

Исследования в §§ 2, 4, 5, а также по разработке алгоритма 1 и замечаний 2 и 4 выполнены при поддержке гранта Российского научного фонда и города Москвы № 22-21-20065 (https://rscf.ru/project/22-21-20065/).

COMPUTER RESEARCH AND MODELING
2024 VOL. 16 NO. 1 P. 105–122
DOI: 10.20537/2076-7633-2024-16-1-105-122

Ки&М

SPECIAL ISSUE

UDC: 519.85

Subgradient methods with B. T. Polyak-type step for quasiconvex minimization problems with inequality constraints and analogs of the sharp minimum

S. M. Puchinin[1,a], E. R. Korolkov[1,b], F. S. Stonyakin[1,2,c], M. S. Alkousa[1,d], A. A. Vyguzov[1,e]

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

E-mail: [a] puchinin.sm@phystech.edu, [b] korolkov.er@phystech.edu, [c] fedyor@mail.ru,
[d] mohammad.alkousa@phystech.edu, [e] vyguzov.aa@phystech.edu

Received 06.12.2023, after completion — 23.12.2023.
Accepted for publication 23.12.2023.

In this paper, we consider two variants of the concept of sharp minimum for mathematical programming problems with quasiconvex objective function and inequality constraints. It investigated the problem of describing a variant of a simple subgradient method with switching along productive and non-productive steps, for which, on a class of problems with Lipschitz functions, it would be possible to guarantee convergence with the rate of geometric progression to the set of exact solutions or its vicinity. It is important that to implement the proposed method there is no need to know the sharp minimum parameter, which is usually difficult to estimate in practice. To overcome this problem, the authors propose to use a step adjustment procedure similar to that previously proposed by B. T. Polyak. However, in this case, in comparison with the class of problems without constraints, it arises the problem of knowing the exact minimal value of the objective function. The paper describes the conditions for the inexactness of this information, which make it possible to preserve convergence with the rate of geometric progression in the vicinity of the set of minimum points of the problem. Two analogs of the concept of a sharp minimum for problems with inequality constraints are considered. In the first one, the problem of approximation to the exact solution arises only to a pre-selected level of accuracy, for this, it is considered the case when the minimal value of the objective function is unknown; instead, it is given some approximation of this value. We describe conditions on the inexact minimal value of the objective function, under which convergence to the vicinity of the desired set of points with a rate of geometric progression is still preserved. The second considered variant of the sharp minimum does not depend on the desired accuracy of the problem. For this, we propose a slightly different way of checking whether the step is productive, which allows us to guarantee the convergence of the method to the exact solution with the rate of geometric progression in the case of exact information. Convergence estimates are proved under conditions of weak convexity of the constraints and some restrictions on the choice of the initial point, and a corollary is formulated for the convex case when the need for an additional assumption on the choice of the initial point disappears. For both approaches, it has been proven that the distance from the current point to the set of solutions decreases with increasing number of iterations. This, in particular, makes it possible to limit the requirements for the properties of the used functions (Lipschitz-continuous, sharp minimum) only for a bounded set. Some computational experiments are performed, including for the truss topology design problem.

Keywords: subgradient method, Lipschitz-continuous function, sharp minimum, Polyak step-size, quasiconvex function, weakly convex function

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 1, pp. 105–122 (Russian).

This work in sections 2, 4, 5, Algorithm 1, Remarks 2 and 4 was supported by Russian Science Foundation and Moscow city, project 22-21-20065 (https://rscf.ru/project/22-21-20065/).

1. Введение

Задачи математического программирования возникают в самых разных приложениях. Одним из хорошо известных подходов к решению задач минимизации функций с ограничениями-неравенствами являются так называемые схемы с переключением между продуктивными и непродуктивными шагами, впервые предложенные Поляком в его работе [Поляк, 1967]. Общая идея подхода заключается в следующем: если в текущей точке значение функции-ограничения достаточно хорошее, то спуск выполняется по целевой функции, а в противном случае — по функции ограничения. Такого типа подходам посвящаются всё новые работы как для выпуклых задач большой и сверхбольшой размерности [Huang, Lin, 2023; Bayandina et al., 2018; Lagae, 2017; Nesterov, 2014], так и для некоторых классов невыпуклых задач [Huang, Lin, 2023].

В данной статье рассмотрены аналоги субградиентных методов такого типа с вариантом шага Б. Т. Поляка с использованием информации о минимальном значении целевой функции f^*, хорошо известного для задач без ограничений [Поляк, 1969]. Использование такого типа шага для липшицевых задач с острым минимумом позволяет доказать сходимость субградиентного метода со скоростью геометрической прогрессии. При этом важно, что реализация метода не потребует знания параметра острого минимума. Для других вариантов регулировки шага субградиентного метода такую скорость сходимости оказывается возможным доказать лишь за счет процедуры рестартов по параметру острого минимума, оценка которого проблематична для многих реально возникающих типов прикладных задач. Стоит отметить, что использование в методах градиентного типа шага Б. Т. Поляка довольно популярно (см., например, [Hazan, Kakade, 2019; Loizou et al., 2021]). Так, совсем недавно предложены и изучены различные современные варианты шага Поляка [Wang et al., 2023; Devanathan, Boyd, 2023; Abdukhakimov et al., 2023].

В статье [Аблаев и др., 2023] для липшицевых задач математического программирования предложены схемы рестартов по параметру «условного» острого минимума, для которых гарантирована сходимость в окрестность точного решения за линейное время как для выпуклых, так и для квазивыпуклых задач. Однако представляется существенным моментом требование знать значение параметра острого минимума для реализации метода. Цель предлагаемой статьи — обойти эту проблему посредством исследования субградиентных схем с переключениями по продуктивным и непродуктивным шагам с вариантом шага Б. Т. Поляка. Однако наличие ограничений-неравенств приводит к нетривиальности вопроса оценки искомого точного минимального значения целевой функции. Более того, это значение может быть большим по сравнению с глобальным минимумом целевой функции на всем пространстве. Поэтому в настоящей статье исследуются аналоги шага Б. Т. Поляка с неточной информацией о минимуме целевой функции и возможность гарантировать сходимость метода со скоростью геометрической прогрессии.

Статья состоит из введения, 4 основных разделов и заключения.

Во втором разделе приводятся основные определения и постановка рассматриваемой задачи, а также некоторые известные прикладные результаты, которые потребуются для анализа предлагаемых в последующих разделах алгоритмов.

В третьем разделе рассматривается аналог острого минимума, а именно ε-острый минимум из работы [Стонякин и др., 2023]. Для данного варианта острого минимума предложен алгоритм, гарантирующий линейную скорость сходимости к ε-точному решению задачи по функции и по ограничению.

В четвертом разделе рассматривается альтернативный вариант задания острого минимума, содержащий максимум из невязки по функции и значения ограничения [Lin et al., 2020; Аблаев и др., 2023]. Данный подход к острому минимуму не предполагает задания параметра точности ε. Это позволяет предложить алгоритм с гарантией линейной скорости сходимости в окрестность множества точных решений, размер которой определяется степенью точности информации о минимальном значении функции. Сходимость к точному решению со скоростью геометрической прогрессии можно гарантировать только при условии доступности информации о точном значении минимума целевой функции, чего не удается доказать в случае использования ε-острого минимума.

В пятом разделе приведены результаты численных экспериментов и продемонстрирована эффективность предлагаемых алгоритмов для некоторых конкретных задач. Среди них можно выделить задачу геометрического программирования [Boyd et al., 2007] и задачу проектирования механических конструкций [Nesterov, Shpirko, 2014].

2. Постановка задачи и основные теоретические сведения

В работе рассматривается задача минимизации вида

$$f(x) \to \min_{g(x)\leqslant 0}, \quad x \in Q, \tag{1}$$

где $Q \subset \mathbb{R}^n$ — замкнутое выпуклое множество; $f\colon \mathbb{R}^n \to \mathbb{R}$ — квазивыпуклая функция, удовлетворяющая условию Липшица для некоторой константы $M_f > 0$ (то есть $|f(x)-f(y)| \leqslant M_f\|x-y\|_2$ для любых $x, y \in \mathbb{R}^n$); $g\colon Q \to \mathbb{R}$ — μ-слабовыпуклая функция, удовлетворяющая условию Липшица для некоторой константы $M_g > 0$.

Напомним, что функция $g\colon \mathbb{R}^n \to \mathbb{R}$ (аналогично и для функций $g\colon Q \to \mathbb{R}$) называется μ-слабовыпуклой ($\mu \geqslant 0$), если функция $x \mapsto g(x) + \frac{\mu}{2}\|x\|_2^2$ выпукла. Для недифференцируемых μ-слабовыпуклых функций g под субдифференциалом $\partial g(x)$ в точке x можно понимать (см. [Davis et al., 2018] и цитируемую там литературу) множество всех векторов $\upsilon \in \mathbb{R}^n$, удовлетворяющих неравенству

$$g(y) \geqslant g(x) + \langle \upsilon, y-x\rangle + o(\|y-x\|_2) \quad \text{при } y \to x. \tag{2}$$

Известно [Davis et al., 2018], что все векторы-субградиенты $\upsilon \in \mathbb{R}^n$ из (2) автоматически удовлетворяют неравенству

$$g(y) \geqslant g(x) + \langle \upsilon, y-x\rangle - \frac{\mu}{2}\|y-x\|_2^2 \quad \forall x, y \in \mathbb{R}^n, \ \upsilon \in \partial g(x).$$

Можно проверить, что слабовыпуклые функции локально липшицевы, и поэтому в качестве субградиентов можно использовать произвольный вектор из субдифференциала Кларка (см., например, [Дудов, Осипцев, 2021]). Если целевая функция или функция ограничений недифференцируемы в некоторой точке x, то ввиду сделанного выше предположения об их липшицевости будем считать их субдифференцируемыми по Кларку и под $\nabla f(x)$ ($\nabla g(x)$) понимать произвольный ненулевой элемент субдифференциала Кларка в данной точке. Если же f (g) дифференцируемы в обычном смысле, то под $\nabla f(x)$ ($\nabla g(x)$) будем понимать обычный градиент f (g) в точке x.

Определение 1 (см. [Нестеров, 1989, п. 1.5]). Функция f, определенная на выпуклом множестве $Q \subset \mathbb{R}^n$, называется квазивыпуклой на этом множестве, если

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

Будем обозначать через f^* оптимальное значение целевой функции задачи (1), а через X_* — множество таких $x \in Q$, на которых это значение достигается. Также для произвольных $x, y \in Q$ введем вспомогательную величину

$$v_f(x, y) := \begin{cases} \left\langle \dfrac{\nabla f(x)}{\|\nabla f(x)\|_2}, x-y \right\rangle, & \nabla f(x) \neq 0, \\ 0, & \nabla f(x) = 0, \end{cases}$$

и докажем для нее следующую техническую лемму.

Лемма 1. *Пусть* $f\colon \mathbb{R}^n \to \mathbb{R}$ *квазивыпукла и* M_f*-липшицева, тогда для любого* $x \in Q$ *такого, что* $\nabla f(x) \neq 0$*, выполнено*

$$f(x) - f^* \leqslant M_f v_f(x, x_*) \quad \forall x_* \in X_*. \tag{3}$$

Доказательство. В силу [Нестеров, 1989, сл. 1.5.1], а также условия M_f-липшицевости функции f для произвольного $x_* \in X_*$

$$f(x) - f^* \leqslant \max\{f(y) - f(z) \mid \|y - z\|_2 \leqslant v_f(x, x_*)\} \leqslant M_f v_f(x, x_*),$$

что и требовалось. □

Будем в дальнейшем предполагать, что у функции f нет стационарных точек в Q, за исключением, может быть, точек из X_* (замечание 4, [Стонякин и др., 2023]). Легко видеть, что неравенство (3) верно также и для точек x из X_*, несмотря на возможное нулевое значение градиента. Тем самым дополнительное ограничение на x ($\nabla f(x) \neq 0$) в условии леммы 1 может быть опущено.

Зачастую на практике точное значение f^* неизвестно. Поэтому будем считать, что алгоритмам, предложенным в данной работе, доступно некоторое значение $\overline{f}$, связанное с f^* следующим образом:

$$f(x) - \overline{f} = c(x)\big(f(x) - f^*\big), \tag{4}$$

где $c(x) \in [C, 2 - C]$, $C \in (0; 1]$ — некоторая константа. Очевидно, что равенство (4) не может быть верно ни при каком значении $f(x)$, достаточно близком к $\overline{f}$ или f^*, где допустимая степень близости задается параметром C — чем он меньше, тем для более близких значений равенство (4) выполняется. Все последующие исследования предлагаемых далее алгоритмов подразумевают, что значения функции в рассматриваемых последовательностях точек x^k не приближаются к $\overline{f}$ или f^* слишком близко, то есть для них равенство (4) подразумевается верным. Однако при $C = 1$ равенство (4), очевидно, всегда верно, и никаких дополнительных ограничений не накладывается.

Также для анализа приведенных ниже алгоритмов нам потребуется следующее утверждение (см. [Поляк, 1969]).

Лемма 2. *Пусть* $\varphi\colon Q \to \mathbb{R}$ *— некоторая функция* ($Q \subset \mathbb{R}^n$ *— замкнутое выпуклое множество*)*, тогда для любых* $x \in Q$, $x_* \in X_* \subset Q$ *и любого* $h \geqslant 0$ *верно следующее неравенство*:

$$\|\overline{x} - x_*\|_2^2 \leqslant \|x - x_*\|_2^2 - 2h\langle \nabla\varphi(x), x - x_*\rangle + h^2\|\nabla\varphi(x)\|_2^2,$$

где $\overline{x} = \mathrm{Pr}_Q\{x - h\nabla\varphi(x)\}$.

3. Субградиентный метод с шагом типа Б. Т. Поляка для липшицевых задач с ограничениями-неравенствами с ε-острым минимумом

В данном разделе будет рассмотрен вариант острого минимума и субградиентного метода, зависящий от желаемой точности решения $\varepsilon > 0$ [Стонякин и др., 2023]. Такой подход достаточно общий, но позволяет даже при доступности точной информации об f^* гарантировать лишь попадание в окрестность некоторого точного решения задачи.

Определение 2. Будем говорить, что задача (1) обладает ε-острым минимумом ($\varepsilon > 0$), если для фиксированного $\alpha > 0$ верна альтернатива

- при $g(x) > \varepsilon$
$$g(x) \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \forall x \in Q,$$
- при $g(x) \leqslant \varepsilon$
$$f(x) - f^* \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \text{или} \quad f(x) - f^* \leqslant \varepsilon \quad \forall x \in Q.$$

Пример 1. Рассмотрим задачу минимизации сильно выпуклой функции
$$\min_{g(x) \leqslant 0,\ x \in Q} f(x), \tag{5}$$
где Q — выпуклое ограниченное замкнутое множество, а $g(x) = \operatorname{dist}(x, Q) = \min_{y \in Q} \|x - y\|_2$. Действительно, задача (5) обладает ε-острым минимумом: так как Q — ограниченное множество, то $\operatorname{diam}(Q) \leqslant C$ для некоторого $C > 0$, поэтому в случае $g(x) = \operatorname{dist}(x, Q) > \varepsilon$ воспользуемся неравенством треугольника
$$\min_{x_* \in X_*} \|x - x_*\|_2 \leqslant \operatorname{dist}(x, Q) + \operatorname{diam}(Q) \leqslant g(x) + C,$$
поэтому
$$g(x) \geqslant \min_{x_* \in X_*} \|x - x_*\|_2 - C = \alpha \min_{x_* \in X_*} \|x - x_*\|_2.$$
При $g(x) \leqslant \varepsilon$ все так же потребуем от целевой функции условие острого минимума:
$$f(x) - f^* \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \text{или} \quad f(x) - f^* \leqslant \varepsilon$$
для некоторого $\alpha > 0$.

Покажем, что с использованием некоторого варианта шага Б. Т. Поляка можно достичь сходимости следующего метода со скоростью геометрической прогрессии к ε-точному решению поставленной задачи.

Теорема 1. *Пусть функция f квазивыпукла и M_f-липшицева ($M_f > 0$), а g μ-слабовыпукла и задача имеет ε-острый минимум ($\alpha > 0$, $\mu > 0$). Пусть $\overline{f}$ такое, что $f(x) - \overline{f} = c(x)\,(f(x) - f^*)$, где $c(x) \in [C, 2 - C]$ для некоторого $C \in (0, 1)$. Начальная точка x_0 такова, что $\min_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha\gamma}{\mu}$ для некоторого фиксированного $\gamma \in (0; 1)$. Тогда после $k + 1$ итераций алгоритма 1 справедлива следующая альтернатива: либо достигнуто ε-точное решение задачи, т. е.*
$$f(x_{k+1}) - f^* \leqslant \varepsilon, \quad g(x_{k+1}) \leqslant \varepsilon,$$
или верно следующее неравенство:
$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2}{M_f^2}\left(2C - C^2\right)\right)^{|I|} \prod_{i \in J} \left(1 - \frac{(1 - \gamma_i)\alpha^2}{\|\nabla g(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2,$$
где последовательность $\{\gamma_i\}_{i=0}^k$ такова, что $\gamma_0 := \gamma \in (0; 1)$, а далее, при $i = 0, 1, 2, \ldots, k$, определим
$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla g(x_i)\|_2^2}}$$
для всякого непродуктивного шага ($i \in J$) и
$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{\alpha^2}{M_f^2}\,(2C - C^2)}$$
для всякого продуктивного шага ($i \in I$).

Алгоритм 1. Схема с переключениями по продуктивным и непродуктивным шагам для липшицевых задач с ограничениями-неравенствами с ε-острым минимумом

Require: $x_0 : \min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha\gamma}{\mu}$, $\gamma \in (0;\, 1)$, $\varepsilon > 0$, $M_f > 0$.

```
I ← ∅, J ← ∅, k ← 0;
repeat
    if g(x_k) ⩽ ε then
        h_k^f = (f(x_k) − f̄) / (M_f ‖∇f(x_k)‖_2);
        x_{k+1} = Pr_Q {x_k − h_k^f ∇f(x_k)};        «продуктивные шаги»
        k → I;
    else
        h_k^g = g(x_k) / ‖∇g(x_k)‖_2^2;
        x_{k+1} = Pr_Q {x_k − h_k^g ∇g(x_k)};        «непродуктивные шаги»
        k → J;
    end if
    k ← k + 1;
until k = N.
```

Доказательство. 1. Для продуктивных шагов по лемме 2 справедливо следующее неравенство:

$$2h_k^f \langle \nabla f(x_k),\, x_k - x_* \rangle \leqslant \left(h_k^f\right)^2 \|\nabla f(x_k)\|_2^2 + \|x_k - x_*\|_2^2 - \|x_{k+1} - x_*\|_2^2 \quad \forall k \in I.$$

Поэтому имеем

$$\begin{aligned}
\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 &\leqslant \min_{x_* \in X_*} \|x_k - x_*\|_2^2 + \left(h_k^f\right)^2 \|\nabla f(x_k)\|_2^2 - 2h_k^f \langle \nabla f(x_k),\, x_k - x_* \rangle = \\
&= \min_{x_* \in X_*} \|x_k - x_*\|_2^2 + \frac{(f(x_k) - \overline{f})^2}{M_f^2} - \frac{2(f(x_k) - \overline{f})}{M_f} \left\langle \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2},\, x_k - x_* \right\rangle = \\
&= \min_{x_* \in X_*} \|x_k - x_*\|_2^2 + \left(\frac{f(x_k) - \overline{f}}{M_f}\right)^2 - \frac{2(f(x_k) - \overline{f})}{M_f} v_f(x_k,\, x_*) \leqslant \\
&\leqslant \min_{x_* \in X_*} \|x_k - x_*\|_2^2 + c^2(x_k) \frac{\left(f(x_k) - f^*\right)^2}{M_f^2} - 2 \cdot \frac{f(x_k) - f^*}{M_f^2} \cdot c(x_k) \left(f(x_k) - f^*\right) = \\
&= \min_{x_* \in X_*} \|x_k - x_*\|_2^2 - \left(\frac{f(x_k) - f^*}{M_f}\right)^2 \left(2c(x_k) - c^2(x_k)\right) \leqslant \\
&\leqslant \min_{x_* \in X_*} \|x_k - x_*\|_2^2 \left(1 - \frac{\alpha^2}{M_f^2} \left(2c(x_k) - c^2(x_k)\right)\right) \leqslant \\
&\leqslant \min_{x_* \in X_*} \|x_k - x_*\|_2^2 \left(1 - \frac{\alpha^2}{M_f^2} \left(2C - C^2\right)\right),
\end{aligned}$$

так как $v_f(x_k,\, x_*) \geqslant \frac{f(x_k) - f^*}{M_f}$.

2. Для непродуктивных шагов при $g(x_k) > \varepsilon$ верно $g(x_k) \geqslant \alpha \min\limits_{x_* \in X_*} \|x_k - x_*\|_2$. Тогда по лемме 2 для ближайшего к x_k точного решения $x_* \in X_*$ ввиду $g(x_*) \leqslant 0$ имеем

$$
\begin{aligned}
\|x_{k+1} - x_*\|_2^2 &\leqslant \|x_k - x_*\|_2^2 + \frac{2g(x_k)\langle \nabla g(x_k),\, x_* - x_k\rangle}{\|\nabla g(x_k)\|_2^2} + \left(\frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}\right)^2 \|\nabla g(x_k)\|_2^2 \leqslant \\
&\leqslant \|x_k - x_*\|_2^2 + \frac{2g(x_k)}{\|\nabla g(x_k)\|_2^2}\left(g(x_*) - g(x_k) + \frac{\mu}{2}\|x_k - x_*\|_2^2\right) + \frac{(g(x_k))^2}{\|\nabla g(x_k)\|_2^2} \leqslant \\
&\leqslant \|x_k - x_*\|_2^2 + \frac{g(x_k)\mu}{\|\nabla g(x_k)\|_2^2}\|x_k - x_*\|_2^2 - \frac{2(g(x_k))^2}{\|\nabla g(x_k)\|_2^2} + \frac{(g(x_k))^2}{\|\nabla g(x_k)\|_2^2} = \\
&= \|x_k - x_*\|_2^2 + \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}\left(\mu\|x_k - x_*\|_2^2 - g(x_k)\right) \leqslant \\
&\leqslant \|x_k - x_*\|_2^2 + \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}\left(\mu\|x_k - x_*\|_2^2 - \alpha\|x_k - x_*\|_2\right) \leqslant \\
&\leqslant \|x_k - x_*\|_2^2 + \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}(\alpha\gamma_k - \alpha)\|x_k - x_*\|_2 \leqslant \left(1 - \frac{(1-\gamma_k)\alpha^2}{\|\nabla g(x_k)\|_2^2}\right)\|x_k - x_*\|_2^2.
\end{aligned}
$$

Тогда ясно, что

$$
\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2}{M_f^2}\left(2C - C^2\right)\right)^{|I|} \prod_{i \in J}\left(1 - \frac{(1-\gamma_i)\alpha^2}{\|\nabla g(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2,
$$

что и требовалось. При этом на «непродуктивном шаге» не может быть ситуации $\nabla g(x_k) = 0$ ввиду предложенного варианта выбора начальной точки x_0 ([Стонякин и др., 2023], замечание 4).

□

Если рассматривать задачи с выпуклыми ограничениями ($\mu = 0$), то пропадает необходимость ограничивать выбор начальной точки x_0 в алгоритме 1. Сформулируем этот результат.

Следствие 1. *Пусть f квазивыпукла и M_f-липшицева ($M_f > 0$), а g выпукла и задача имеет ε-острый минимум ($\alpha > 0$). Тогда после $k + 1$ итераций алгоритма* 1 *справедлива следующая альтернатива: либо достигнуто ε-точное решение задачи, т. е.*

$$
f(x_{k+1}) - f^* \leqslant \varepsilon, \quad g(x_{k+1}) \leqslant \varepsilon,
$$

или верно следующее неравенство:

$$
\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2}{M_f^2}\left(2C - C^2\right)\right)^{|I|} \prod_{i \in J}\left(1 - \frac{\alpha^2}{\|\nabla g(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2.
$$

Доказательство. Аналогично теореме 1 с $\mu = 0$. □

Замечание 1. Отметим, что равномерная ограниченность норм $\|\nabla g(x_k)\|_2 \leqslant M_g$, $\forall k \in \overline{0, N}$, где $M_g > 0$, влечет в выпуклом случае оценку

$$
\begin{aligned}
\min_{x_* \in X_*} \|x_N - x_*\|_2^2 &\leqslant \left(1 - \frac{\alpha^2}{M_f^2}\left(2C - C^2\right)\right)^{|I|}\left(1 - \frac{\alpha^2}{M_g^2}\right)^{|J|} \min_{x_* \in X_*} \|x_0 - x_*\|_2^2 \leqslant \\
&\leqslant \left(1 - \frac{\alpha^2}{M^2}\left(2C - C^2\right)\right)^{N} \min_{x_* \in X_*} \|x_0 - x_*\|_2^2,
\end{aligned}
$$

где $M = \max\{M_f, M_f\}$. Это, в свою очередь, указывает на сходимость со скоростью геометрической прогрессии, причем при $C = 1$ получаем $\overline{f} = f^*$ и наибольшую скорость сходимости.

Замечание 2. Также шаг Б. Т. Поляка позволяет доказать, что невязка по аргументу не возрастает по мере увеличения количества итераций. А значит, можно считать, что для любого $k = 0, 1, 2, \ldots$ выполняется $x_k \in Q \cap G$, где

$$G = \left\{ x \in \mathbb{R}^n \,\middle|\, \|x - x_*\|_2 \leqslant \min_{x_* \in X_*} \|x_0 - x_*\|_2 \right\}.$$

Поэтому достаточно требовать липшицевость и квазивыпуклость функции f, а также ограниченность норм градиента из предыдущего замечания на множестве G.

Замечание 3. В некоторых задачах возникает условие *слабого острого минимума*

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

для некоторого фиксированного $p \in [1, +\infty)$ и некоторой постоянной $C > 0$. Тогда при достаточно большом расстоянии от текущей точки до множества решений X_* выполняется

$$f(x) - f^* \geqslant C \min_{x_* \in X_*} \|x - x_*\|_2^p \geqslant C_1 \min_{x_* \in X_*} \|x - x_*\|_2,$$

где C_1 — некоторая постоянная. Таким образом, описанная методика позволит достигнуть определенного уровня близости к точному решению по аргументу.

Замечание 4. Пусть теперь задача ставится следующим образом:

$$\min_{g(x) \leqslant 0,\ x \in Q} f(x), \tag{6}$$

где Q — выпуклое замкнутое подмножество $\mathbb{R}^n$, а $g(x) = \max_{i \in \overline{1, m}} g_i(x)$. Тогда следует переформулировать определение ε-острого минимума задачи (6).

Определение 3. Будем говорить, что задача (6) обладает ε-острым минимумом ($\varepsilon > 0$), если существует набор $\alpha_{g_i} > 0$ для $i \in \overline{1, m}$ и $\alpha_f > 0$, для которых верна альтернатива

- при $g_i(x) > \varepsilon$
$$g_i(x) \geqslant \alpha_{g_i} \min_{x_* \in X_*} \|x - x_*\|_2 \quad \forall x \in Q,$$
- при $g_i(x) \leqslant \varepsilon$
$$f(x) - f^* \geqslant \alpha_f \min_{x_* \in X_*} \|x - x_*\|_2 \quad \text{или} \quad f(x) - f^* \leqslant \varepsilon \quad \forall x \in Q.$$

Тогда множество продуктивных шагов алгоритма обозначим как $I = \{k \in \mathbb{N} \mid \forall i \in \overline{1, m} \hookrightarrow g(x_k) \leqslant \varepsilon\}$, непродуктивных — как $J = \{k \in \mathbb{N} \mid \exists i\colon g_i(x_k) > \varepsilon\}$. То есть пропадает необходимость считать все субградиенты $\nabla g_i(x)$, вместо этого работая с первым нарушенным ограничением. При этом на непродуктивных шагах в качестве параметров ограничений (α, M, μ и т. д.) можно выбрать наиболее удачные с точки зрения сходимости.

4. Субградиентный метод с шагом типа Б. Т. Поляка для липшицевых задач с ограничениями-неравенствами с вариантом острого минимума, не зависящим от ожидаемой точности решения задач

В этом разделе мы рассмотрим другой способ задания острого минимума, предложенный в [Lin et al., 2020].

Определение 4. Будем говорить, что задача (1) обладает «условным» острым минимумом, если существует такое $\alpha > 0$, для которого выполнено

$$\max\{f(x) - f^*, g(x)\} \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2. \tag{7}$$

В недавней работе [Аблаев и др., 2023] для липшицевых задач с «условным» острым минимумом доказана сходимость со скоростью геометрической прогрессии рестартованных по параметру острого минимума субградиентных методов. Недостаток этого подхода — необходимость знать оценку такого параметра, что довольно затруднительно для многих реально возникающих задач на практике. Мы поставим задачу получить результат о линейной скорости сходимости (хотя бы в окрестность множества решений) для варианта субградиентного метода, не предполагающего для реализации знания величины этого параметра. В отличие от рассмотренного в предыдущем разделе ε-острого минимума данный способ задания острого минимума не зависит от параметра ε, который, как видно из теоремы 1, ограничивает точность, которую может гарантировать алгоритм 1. Предлагаемый же алгоритм 2 для задачи с острым минимумом (7), как показано далее в этом разделе, в теореме 2, имеет линейную скорость сходимости и при достаточно большом числе итераций может позволить достичь любого заданного качества решения по аргументу в зависимости от точности информации о минимальном значении f^*.

Алгоритм 2. Схема с переключениями по продуктивным и непродуктивным шагам для липшицевых задач с ограничениями-неравенствами с вариантом острого минимума, не зависящим от ожидаемой точности решения задач

Require: $x_0\colon \min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{C\alpha\gamma_0}{\mu}$, $\gamma_0 \in (0;\,1)$, $M_f > 0$, $N > 0$.

$I \leftarrow \varnothing$, $J \leftarrow \varnothing$, $k \leftarrow 0$;
repeat
  if $f(x_k) - \overline{f} \geqslant g(x_k)$ **then**
    $h_k^f = \frac{f(x_k) - \overline{f}}{M_f \|\nabla f(x_k)\|_2}$;
    $x_{k+1} = \mathrm{Pr}_Q\left\{x_k - h_k^f \nabla f(x_k)\right\}$;   «продуктивные шаги»
    $k \to I$;
  else
    $h_k^g = \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}$;
    $x_{k+1} = \mathrm{Pr}_Q\left\{x_k - h_k^g \nabla g(x_k)\right\}$;   «непродуктивные шаги»
    $k \to J$;
  end if
  $k \leftarrow k + 1$;
until $k = N$.

Проведем анализ алгоритма 2 для задачи (1) с учетом всех предположений (отсутствие стационарных точек в $Q \setminus X_*$, наличие острого минимума (7) и доступ к $\overline{f}$).

Начнем с анализа «продуктивных шагов». Возможны два случая:

- $f(x_k) - f^* \geqslant g(x_k)$: в силу условия острого минимума (7) и условия (4) на $\overline{f}$ выполнено

$$f(x_k) - \overline{f} = c(x_k)\left(f(x_k) - f^*\right) \geqslant c(x_k)\alpha \min_{x_* \in X_*} \|x_k - x_*\|_2;$$

- $f(x_k) - f^* \leqslant g(x_k)$: тогда, так как шаг «продуктивный» и в силу условия острого минимума (7), выполнено

$$f(x_k) - \overline{f} \geqslant g(x_k) \geqslant \alpha \min_{x_* \in X_*} \|x_k - x_*\|_2.$$

То есть

$$f(x_k) - \overline{f} \geqslant \min\{1,\ c(x_k)\}\alpha \min_{x_* \in X_*} \|x_k - x_*\|_2. \tag{8}$$

Рассмотрим ближайшее $x_* \in X_*$ к x^k и, применив лемму 2 к x_{k+1} на «продуктивном шаге», получим

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - 2h_k^f \langle \nabla f(x_k),\ x_k - x_* \rangle + \left(h_k^f\right)^2 \|\nabla f(x_k)\|_2^2 =$$
$$= \|x_k - x_*\|_2^2 - 2\frac{f(x_k) - \overline{f}}{M_f} v_f(x_k,\ x_*) + \left(\frac{f(x_k) - \overline{f}}{M_f}\right)^2.$$

Воспользовавшись леммой 1 и условием (4) на $\overline{f}$, получаем

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - \frac{2 - c(x_k)}{c(x_k)} \left(\frac{f(x_k) - \overline{f}}{M_f}\right)^2.$$

Наконец, применив (8),

$$\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{2 - c(x_k)}{c(x_k)} (\min\{1,\ c(x_k)\})^2 \frac{\alpha^2}{M_f^2}\right) \|x_k - x_*\|_2^2 \leqslant \left(1 - \frac{C}{2 - C} \frac{\alpha^2}{M_f^2}\right) \|x_k - x_*\|_2^2.$$

Следовательно,

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{C}{2 - C} \frac{\alpha^2}{M_f^2}\right) \min_{x_* \in X_*} \|x_k - x_*\|_2^2. \tag{9}$$

Теперь перейдем к анализу «непродуктивного шага». Действовать будем по аналогичной схеме. Возможны два случая:

- $f(x_k) - f^* \geqslant g(x_k)$: в силу условия острого минимума (7), условия (4) на $\overline{f}$ и того, что шаг «непродуктивный», выполнено

$$g(x_k) \geqslant f(x_k) - \overline{f} = c(x_k)\left(f(x_k) - f^*\right) \geqslant c(x_k)\alpha \min_{x_* \in X_*} \|x_k - x_*\|_2 \geqslant C\alpha \min_{x_* \in X_*} \|x_k - x_*\|_2;$$

- $f(x_k) - f^* \leqslant g(x_k)$: тогда, в силу условия острого минимума (7), выполнено

$$g(x_k) \geqslant \alpha \min_{x_* \in X_*} \|x_k - x_*|_2.$$

То есть, так как $C \leqslant 1$,

$$g(x_k) \geqslant C\alpha \min_{x_* \in X_*} \|x_k - x_*\|_2. \tag{10}$$

Снова взяв ближайшее $x_* \in X_*$ к x^k и воспользовавшись леммой 2 для x_{k+1} с «непродуктивного шага», получим

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - 2h_k^g \left\langle \nabla g(x_k),\ x_k - x_* \right\rangle + \left(h_k^g\right)^2 \|\nabla g(x_k)\|_2^2 =$$
$$= \|x_k - x_*\|_2^2 - 2\frac{g(x_k)}{|\nabla g(x_k)\|_2^2} \left\langle \nabla g(x_k),\ x_k - x_* \right\rangle + \frac{(g(x_k))^2}{\|\nabla g(x_k)\|_2^2}.$$

В силу μ-слабой выпуклости g имеем

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 + 2\frac{g(x_k)}{\|\nabla g(x_k)\|_2^2} \left(g(x_*) - g(x_k) + \frac{\mu}{2}\|x_k - x_*\|_2^2\right) + \frac{(g(x_k))^2}{\|\nabla g(x_k)\|_2^2} \leqslant$$
$$\leqslant \|x_k - x_*\|_2^2 - \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2} \left(g(x_k) - \mu\|x_k - x_*\|_2^2\right),$$

где в последнем неравенстве учтено, что $g(x_*) \leqslant 0$. Далее, дважды воспользовавшись неравенством (10), получим

$$\begin{aligned} \|x_{k+1} - x_*\|_2^2 &\leqslant \|x_k - x_*\|_2^2 - \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}\left(C\alpha\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2\right) \leqslant \\ &\leqslant \|x_k - x_*\|_2^2 - \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2}(1-\gamma_k)C\alpha\|x_k - x_*\|_2 \leqslant \\ &\leqslant \left(1 - \frac{(1-\gamma_k)C^2\alpha^2}{\|\nabla g(x_k)\|_2^2}\right)\|x_k - x_*\|_2^2, \end{aligned}$$

где последовательность $\{\gamma_{i+1}\}_{i=0}^{k-1}$ задается следующим образом:

$$\gamma_{i+1} = \begin{cases} \gamma_i\sqrt{1 - \dfrac{C}{2-C}\dfrac{\alpha^2}{M_f^2}}, & i \in I; \\ \gamma_i\sqrt{1 - \dfrac{(1-\gamma_i)C^2\alpha^2}{\|\nabla g(x_i)\|_2^2}}, & i \in J, \end{cases} \tag{11}$$

гарантирующим, что $\min\limits_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \frac{C\alpha\gamma_k}{\mu}$. Следовательно,

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - C^2\frac{(1-\gamma_k)\alpha^2}{\|\nabla g(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2. \tag{12}$$

Осталось показать, что на «непродуктивном шаге» не может быть ситуации $\nabla g(x_k) = 0$. Предположим противное. Тогда, в силу μ-слабой выпуклости функции g, для любого $x_* \in X_*$

$$g(x_*) \geqslant g(x_k) + \left\langle \nabla g(x_k),\, x_* - x_k\right\rangle - \frac{\mu}{2}\|x_* - x_k\|_2^2 = g(x_k) - \frac{\mu}{2}\|x_* - x_k\|_2^2. \tag{13}$$

Так как любое $x_* \in X_*$ является решением, то $g(x_*) \leqslant 0$. Следовательно, заметив, что неравенство (13) верно для любого $x_* \in X_*$, и воспользовавшись неравенством (10), имеем

$$C\alpha\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant g(x_k) \leqslant \frac{\mu}{2}\min_{x_*\in X_*}\|x_* - x_k\|_2^2.$$

То есть $\min\limits_{x_*\in X_*}\|x_* - x_k\|_2 \geqslant \frac{2C\alpha}{\mu}$, что неверно, так как $\min\limits_{x_*\in X_*}\|x_0 - x_*\|_2 \leqslant \frac{C\alpha\gamma_0}{\mu}$, где $\gamma_0 \in (0;\,1)$, и, в силу оценок (9) и (12), расстояния от каждой следующей получаемой алгоритмом точки до множества решений меньше, чем от каждой предыдущей.

Таким образом, верна следующая

Теорема 2. *Пусть $Q \subset \mathbb{R}^n$ — замкнутое выпуклое множество, $f\colon \mathbb{R}^n \to \mathbb{R}$ — M_f-липшицевая квазивыпуклая функция, $g\colon Q \to \mathbb{R}$ — μ-слабовыпуклая функция, а начальная точка x_0 такова, что $\min\limits_{x_*\in X_*}\|x_0 - x_*\|_2 \leqslant \frac{C\alpha\gamma_0}{\mu}$, где $\gamma_0 \in (0;\,1)$. Тогда верно следующее неравенство для выходной точки x_N алгоритма 2 после N итераций*:

$$\min_{x_*\in X_*}\|x_N - x_*\|_2^2 \leqslant \left(1 - \frac{C}{2-C}\frac{\alpha^2}{M_f^2}\right)^{|I|}\prod_{k\in J}\left(1 - C^2\frac{(1-\gamma_k)\alpha^2}{\|\nabla g(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_0 - x_*\|_2^2,$$

где последовательность $\{\gamma_{i+1}\}_{i=0}^{N-1}$ задается согласно (11).

Следствие 2. *Если в условиях теоремы* 2 *дополнительно предположить выпуклость функции* g, *или, иными словами, что* $\mu = 0$, *то будет верно следующее неравенство для выходной точки* x_N *алгоритма* 2 *после* N *итераций*:

$$\min_{x_*\in X_*}\|x_N - x_*\|_2^2 \leqslant \left(1 - \frac{C}{2-C}\frac{\alpha^2}{M_f^2}\right)^{|I|}\prod_{k\in J}\left(1 - C^2\frac{\alpha^2}{\|\nabla g(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_0 - x_*\|_2^2.$$

При этом выбор начальной точки ничем не ограничен.

Доказательство. Аналогично приведенным выше рассуждениям при $\mu = 0$. □

Следствие 3. *Если, помимо выпуклости функции* g, *в условиях теоремы* 2 *предположить ограниченность норм* $\|\nabla g(x_k)\|_2 \leqslant M_g$ *для всех* $k = \overline{0, N-1}$ (*например, в случае* M_g*-липшицевости функции* g), *то будет верно следующее неравенство для выходной точки* x_N *алгоритма* 2 *после* N *итераций*:

$$\min_{x_*\in X_*}\|x_N - x_*\|_2^2 \leqslant \left(1 - C^2\frac{\alpha^2}{M^2}\right)^N \min_{x_*\in X_*}\|x_0 - x_*\|_2^2,$$

где $M = \max\{M_f, M_g\}$.

Доказательство. Аналогично приведенным выше рассуждениям с учетом того, что $\frac{C}{2-C} \geqslant C^2$, так как $C \in (0;\ 1]$. □

Замечание 5. Аналогично замечанию 2 для алгоритма 1 липшицевость функции f, а также ограниченность норм в следствии 3 можно требовать не на всем множестве Q, а лишь на некотором ограниченном подмножестве G.

5. Вычислительные эксперименты

В настоящем разделе для иллюстрации работоспособности предложенных выше алгоритмов 1 и 2 приведем некоторые результаты вычислительных экспериментов для четырех примеров в сравнении с алгоритмом 3 из [Аблаев и др., 2023], который является одним из недавно предложенных алгоритмов для класса квазивыпуклых задач оптимизации с функциональными ограничениями-неравенствами.

Пример 2 (задача геометрического программирования [Boyd et al., 2007]). Рассмотрим целевую функцию вида

$$f(x) = \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p}, \quad p \geqslant 1, \tag{14}$$

и следующую функцию ограничений:

$$g(x) = \max_{i\in\overline{1,m}}\left\{g_i(x) = a_i x_1^{\alpha_{i1}} x_2^{\alpha_{i2}}\cdots x_n^{\alpha_{in}} - b_i,\ x_j > 0\ \forall j \in \overline{1, n}\right\}, \tag{15}$$

где $a_i > 0$, $b_i \in \mathbb{R}$, $(\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in}) \in \mathbb{R}^n$ и каждая функция $g_i(x)$ ($i \in \overline{1, m}$) является позиномиальной функцией.

Пример 3. Предположим, что $a, b \in \mathbb{R}^n$, и определим следующую целевую функцию:

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \tag{16}$$

то есть отношение евклидова расстояния до a к расстоянию до b. Функция f квазивыпукла на полупространстве $\{x \in \mathbb{R}^n : \|x - a\|_2 \leqslant \|x - b\|_2\}$. Функция ограничений имеет следующий вид:

$$g(x) = \|x\|_2 + \max\{\langle -a, x\rangle, \|x\|_2\} - b, \tag{17}$$

где $a \in \mathbb{R}^n$ — фиксированный вектор и $b \in \mathbb{R}$, или имеет следующий вид:

$$g(x) = \max_{i \in \overline{1, m}} \{g_i(x) = \langle \alpha_i, x\rangle - \beta_i\}, \tag{18}$$

где $\alpha_i \in \mathbb{R}^n$, $\beta_i \in \mathbb{R}$ $\forall i \in \overline{1, m}$.

Для этого примера мы берем $a = \mathbf{0} \in \mathbb{R}^n$, точка $b \in \mathbb{R}^n$ выбрана так, что $\|b - a\|_2 = 2$, и в качестве множества Q выберем шар из $\mathbb{R}^n$ с центром в $a = \mathbf{0} \in \mathbb{R}^n$ радиусом 1.

Пример 4 (задача проектирования механических конструкций [Nesterov, Shpirko, 2014]). Интересным примером приложений может быть использование схем с переключениями к задаче проектирования механических конструкций, которые могут быть сведены [Nesterov, Shpirko, 2014] к оптимизационной задаче вида

$$\max_{x \in Q} \langle \alpha, x\rangle, \quad \text{удовл.} \quad g(x) := \max_{i \in \overline{1, m}} \{\pm\langle a_i, x\rangle - 1\} = \max_{i \in \overline{1, 2m}} g_i(x), \tag{19}$$

где $g_i(x) = \pm\langle a_i, x\rangle - 1$ (у нас есть $2m$ функциональных ограничений $g_i(\cdot)$).

Выберем в качестве множества Q для примеров 3 и 4 евклидов шар из $\mathbb{R}^n$ с центром в $\mathbf{0} \in \mathbb{R}^n$ радиусом r, а для примера 2 выбираем положительную часть этого шара.

Для примера 2 запускался алгоритм 1 (на практике, по крайней мере для рассмотренных примеров, заметной разницы в работе алгоритмов 1 и 2 нет, поэтому мы рассматриваем здесь только работу алгоритма 1) со стартовой точкой $\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ при $n = 1000$, $m = 100$, $r = 1$, $\varepsilon = 10^{-3}$, $p = 5$. Коэффициенты в (15) генерируются случайным образом с равномерным распределением на интервале $[0, 1)$ и константы b_i генерируются случайным образом с нормальным (гауссовским) распределением с математическим ожиданием в 0 и среднеквадратичным отклонением, равным 1. В этом случае оптимальное значение $f^* = 0$.

Для примера 3 с ограничением вида (17) запускались алгоритмы со стартовой точкой $x_0 = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ и $n = 10^5$, $r = 1$. Вектор a и константа b в (17) сгенерированы случайным образом с равномерным распределением на интервале $[0, 1)$. В случае ограничений вида (18) алгоритмы запускались с начальной точкой $\left(-\frac{1}{\sqrt{n}}, \ldots, -\frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ при $n = 1000$, $m = 100$, $r = 1$. Коэффициенты α_i и β_i в (18) генерируются случайным образом с равномерным распределением на интервале $[0, 1)$.

Для примера 4 мы запускали алгоритмы со стартовой точкой $x_0 = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ и $n = 1000$, $m = 100$, $r = 1$, $\varepsilon = 10^{-4}$. Вектор α в (19) сгенерирован случайным образом с равномерным распределением на интервале $[0, 1)$. Коэффициенты a_i в (19) генерируются случайным образом с нормальным (гауссовским) распределением с математическим ожиданием в 0 и среднеквадратичным отклонением, равным 0,1 и 1.

Результаты сравниваемых алгоритмов (алгоритм 1 и алгоритм 3 из [Аблаев и др., 2023] «субградиентный метод с переключениями») представлены на рис. 1, 2, 3 и 4.

Эти результаты отражают значения целевой функции в примерах 2 и 3, а также значения функции ограничений в точках на каждой итерации. А результаты на рис. 4 (для примера 4) отражают значении $f(x_k) - f^*$, когда коэффициенты a_i в (19) генерируются случайным образом

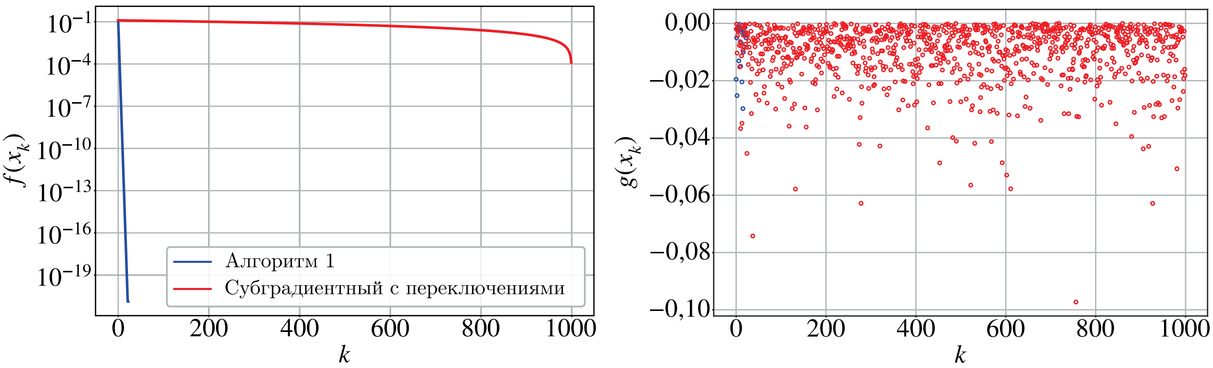

Рис. 1. Результаты алгоритма 1 и субградиентного метода (алгоритм 3 из [Аблаев и др., 2023]) с переключениями для примера 2. На этих рисунках показана динамика значений $f(x_k)$ (14) и значения $g(x_k)$ (15)

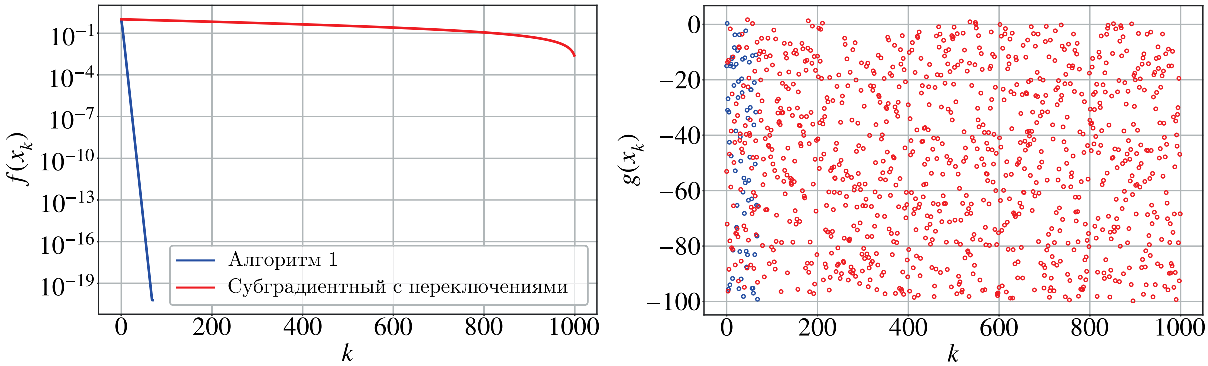

Рис. 2. Результаты алгоритма 1 и субградиентного метода с переключениями (алгоритм 3 из [Аблаев и др., 2023]) для примера 3 с ограничениями (17). На этих рисунках показана динамика значений $f(x_k)$ (16) и значения $g(x_k)$ (17)

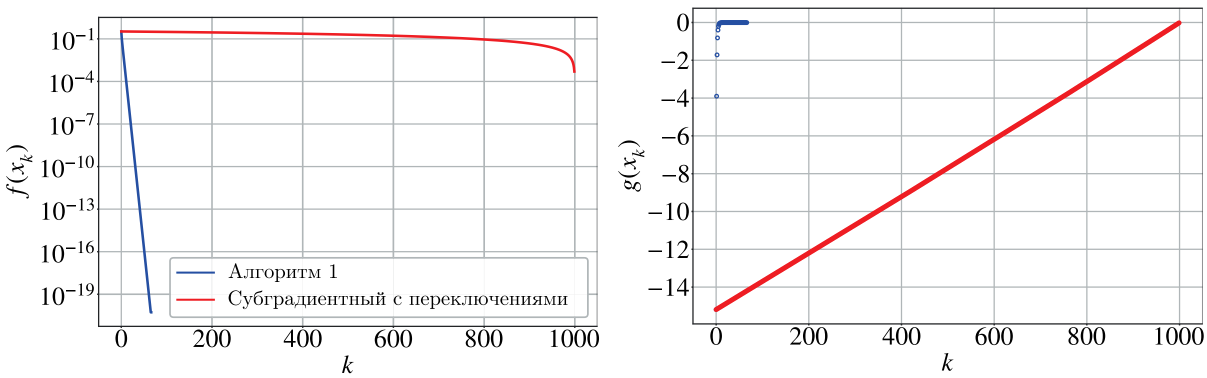

Рис. 3. Результаты алгоритма 1 и субградиентного метода с переключениями (алгоритм 3 из [Аблаев и др., 2023]) для примера 3 с ограничениями (18). На этих рисунках показана динамика значений $f(x_k)$ (16) и значения $g(x_k)$ (18)

с нормальным распределением с математическим ожиданием 0 и стандартным средним квадратичным отклонением, равным 0,1 (слева), и с математическим ожиданием 0 и стандартным средним квадратичным отклонением 1 (справа).

Из рис. 1, 2, 3 и 4 мы видим эффективность предложенного алгоритма 1, с помощью этого алгоритма мы можем достичь решения задачи с очень высокой скоростью по сравнению с другими схемами, такими как алгоритм 3 из [Аблаев и др., 2023], которые медленно сходятся

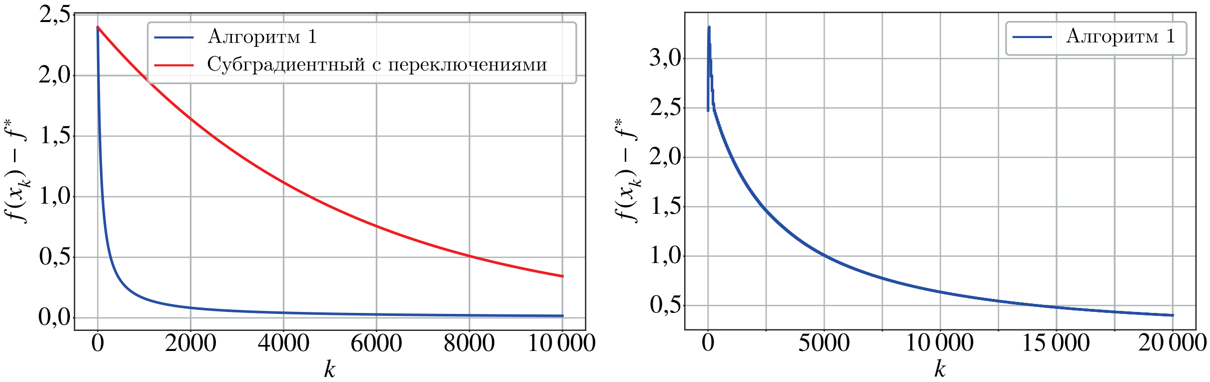

Рис. 4. Результаты алгоритма 1 и субградиентного метода с переключениями (алгоритм 3 из [Аблаев и др., 2023]) для примера 4 (задача (19)). На этих рисунках показана динамика значений $f(x_k) - f^*$, когда коэффициенты a_i в (19) генерируются случайным образом с нормальным распределением с центром, равным 0, и стандартным отклонением, равным 0,1 (слева), и с центром, равным 0, и стандартным отклонением, равным 1 (справа)

к решению. Также на рис. 4 (справа) мы видим результаты только для алгоритма 1, потому что для случая, когда коэффициенты a_i в (19) генерируются случайным образом с нормальным распределением с математическим ожиданием 0 и средним квадратичным отклонением 1, по субградиентному алгоритму с переключением до 20 000 итераций нет ни одного продуктивного шага.

Сравним теперь скорость сходимости метода для шага вида $h_k^f = \frac{\varepsilon}{\|\nabla f(x_k)\|_2^2}$ и шага Поляка из алгоритма 1: $h_k^f = \frac{f(x_k) - \overline{f}}{M_f \|\nabla f(x_k)\|_2}$. Шаг для ограничений же у обоих методов поставим одинаковый: $h_k^g = \frac{1}{\|\nabla g(x_k)\|_2}$. Далее, рассмотрим следующий пример.

Пример 5. Предположим, что $x \in \mathbb{R}^n_{++}$, и определим следующую целевую функцию:

$$f(x) = -\sqrt{x^\top a},$$

где $a \in \mathbb{R}^n_{++}$ — фиксированный вектор, а функция ограничений имеет следующий вид:

$$g(x) = \sum_{i=0}^{n} x \log\left(\frac{x_i}{a_i}\right) - x_i - a_i,$$

т. е. $g(x)$ отражает расстояние Кульбака – Лейблера между a и x. Положим $g(x) \leqslant 1000$ и размерность задачи $n = 10^5$.

Для приближенного решения задачи минимизации $f(x)$ с ограничениями $g(x)$ на множестве $\mathbb{R}^n_{++}$ рассчитаем решение в библиотеке cvxpy. Воспользуемся этим примерным решением для расчета шага Поляка на продуктивных итерациях. Из рис. 5 видим, что метод с шагом Поляка сильно превосходит шаг вида $h_k^f = \frac{\varepsilon}{\|\nabla f(x_k)\|_2^2}$ по скорости сходимости, его убывание гораздо резче, но затем, после достаточного убывания, оба шага отбрасывает назад непродуктивными шагами.

6. Заключение

В статье исследованы некоторые субградиентные методы с переключениями по продуктивным и непродуктивным шагам для задач минимизации квазивыпуклых липшицевых функций с ограничениями-неравенствами двух предположений об остром минимуме [Стонякин и др.,

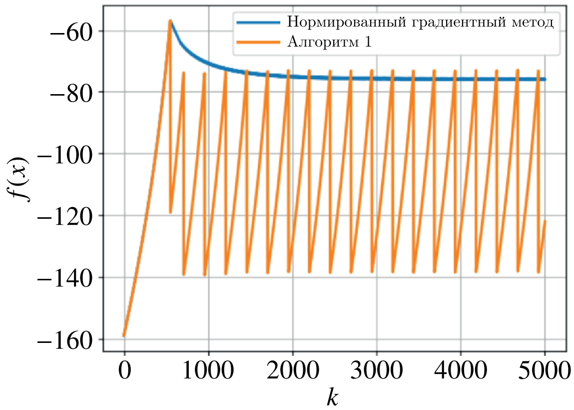

Рис. 5. Результаты сходимости алгоритма 1 с шагом Поляка и с нормированным градиентным шагом для примера (5)

2023; Lin et al., 2020]. Для такого класса задач развивается подход [Аблаев и др., 2023] на базе субградиентных методов с переключениями по продуктивным и непродуктивным шагам. Однако важно, что вместо рестартов по параметру острого минимума используется процедура регулировки шага в такого типа субградиентных методах, аналогичная подходу Б. Т. Поляка [Поляк, 1969]. Известно, что использование в методах градиентного типа шага Б. Т. Поляка довольно популярно (см., например, [Hazan, Kakade, 2019; Loizou et al., 2021]). Совсем недавно предложены и изучены различные современные варианты шага Поляка [Wang et al., 2023; Devanathan, Boyd, 2023; Abdukhakimov et al., 2023]. При этом, однако, естественно возникает проблема доступности информации о минимальном значении целевой функции, что еще более важно для задач с дополнительными ограничениями. В статье проработан этот вопрос и предложен подход к такого типа процедуре регулировки шага при условии с неточной информацией об оптимальном значении целевой функции задачи. Получены условия, при которых для такого типа шагов можно ожидать сходимость метода со скоростью геометрической прогрессии в окрестность множества точных решений. Описана методика повышения скорости работы данного алгоритма в случае задач со многими ограничениями-неравенствами, основанная на отказе от прохода по всем ограничениям. Доказаны оценки качества выдаваемого методом решения для предложенных алгоритмов в зависимости от точности информации о минимальном значении f^*. Выполненные вычислительные эксперименты, особенно для задач геометрического программирования и проектирования механических конструкций, показали хорошую эффективность предложенного метода в решении прикладных задач. Существенный момент в том, что использование предлагаемой в настоящей работе процедуры регулировки шага позволяет при реализации уйти от необходимости знания параметра острого минимума, который на практике часто довольно проблематично оценить. Вообще, такого типа шаги потенциально можно применять к любым оптимизационным задачам с ограничениями-неравенствами. Поэтому представляется, что исследованные в статье методы перспективны для использования на классах самых разных минимизационных задач с ограничениями.

Список литературы (References)

Аблаев С. С., Стонякин Ф. С., Алкуса М. С., Гасников А. В. Адаптивные субградиентные методы для задач математического программирования с квазивыпуклыми функциями // Труды Института математики и механики УрО РАН. — 2023. — Т. 29, № 3. — С. 7–25.

Ablaev S. S., Stonyakin F. S., Alkousa M. S., Gasnikov A. V. Adaptivnye subgradientnye metody dlja zadach matematicheskogo programmirovanija s kvazivypuklymi funkcijami [Adaptive subgradient methods for mathematical programming problems with quasiconvex functions] // Proceedings of the Institute of Mathematics and Mechanics: Ural Branch of the Russian Academy of Sciences. — 2023. — Vol. 29, No. 3. — P. 7–25 (in Russian).

Дудов С. И., Осипцев М. А. Характеризация решения задач сильно-слабовыпуклого программирования // Матем. сб. — 2021.— Т. 212, № 6. — С. 43–72.

Dudov S. I., Osipcev M. A. Harakterizacija reshenija zadach sil'no-slabovypuklogo programmirovanija [Characterization of the solutions to strongly-weakly convex programming problems] // Math. book. — 2021. — Vol. 212, No. 6. — P. 43–72 (in Russian).

Нестеров Ю. Е. Эффективные методы в нелинейном программировании. — М.: Радио и связь, 1989. — 301 с.

Nesterov Yu. E. Jeffektivnye metody v nelinejnom programmirovanii [Effective methods in nonlinear programming]. — Moscow: Radio and communications, 1989. — 301 p. (in Russian).

Поляк Б. Т. Минимизация негладких функционалов // Ж. вычисл. матем. и матем. физ. — 1969.— Т. 9, № 3. — С. 509–521.

Poljak B. T. Minimizacija negladkih funkcionalov [Minimization of nonsmooth functionals] // Journal of Computational Mathematics and Mathematical Physics. — 1969.— Vol. 9, No 3. — P. 509–521 (in Russian).

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

Poljak B. T. Odin obshhij metod reshenija jekstremal'nyh zadach [One general method for solving extremal problems] // Reports of the USSR Academy of Sciences. — 1967. — Vol. 174, No 1. — P. 33–36 (in Russian).

Стонякин Ф. С., Аблаев С. С., Баран И. В., Алкуса М. С. Субградиентные методы для слабо выпуклых и относительно слабо выпуклых задач с острым минимумом // Компьютерные исследования и моделирование. — 2023. — Т. 15, № 2. — С. 393–412.

Stonjakin F. S., Ablaev S. S., Baran I. V., Alkusa M. S. Subgradientnye metody dlja slabo vypuklyh i otnositel'no slabo vypuklyh zadach s ostrym minimumom [Subgradient methods for weakly convex and relatively weakly convex problems with a sharp minimum] // Computer research and modeling. — 2023. — Vol. 15, No. 2. — P. 393–412 (in Russian).

Abdukhakimov F., Xiang C., Kamzolov D., Takáč M. Stochastic gradient descent with preconditioned Polyak step-size // arXiv:2310.02093

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

Boyd S., Kim S. J., Vandenberghe L., Hassibi A. A tutorial on geometric programming // Optimization and engineering. — 2007. — Vol. 8. — P. 67–127.

Davis D., Drusvyatskiy D., Kellie M., Paquette C. Subgradient methods for sharp weakly convex functions // Journal of Optimization Theory and Applications. — 2018. — Vol. 179. — P. 962–982.

Devanathan N., Boyd S. Polyak minorant method for convex optimization // arXiv:2310.07922

Hazan E., Kakade S. Revisiting the Polyak step size // arXiv:1905.00313

Huang Y., Lin Q. Single-loop switching subgradient methods for non-smooth weakly convex optimization with non-smooth convex constraints // arXiv:2301.13314

Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. — 2017.

Lin Q., Ma R., Nadarajah S., Soheili N. A parameter-free and projection-free restarting level set method for adaptive constrained convex optimization under the error bound condition // arXiv:2010.15267

Loizou N., Vaswani S., Laradji I. H., Lacoste-Julien S. Stochastic Polyak step-size for SGD: An adaptive learning rate for fast convergence // International Conference on Artificial Intelligence and Statistics. — PMLR, 2021. — P. 1306–1314.

Nesterov Y. Subgradient methods for huge-scale optimization problems // Mathematical Programming. — 2014. — Vol. 146, No. 1–2. — P. 275–297.

Nesterov Y., Shpirko S. Primal-dual subgradient method for huge-scale linear conic problems // SIAM Journal on Optimization. — 2014. — Vol. 24, No. 3. — P. 1444–1457.

Wang X., Johansson M., Zhang T. Generalized Polyak step size for first order optimization with momentum // arXiv:2305.12939