



Math-Net.Ru

All Russian mathematical portal

S. S. Ablav, F. S. Stonyakin, M. S. Alkousa, A. V. Gasnikov, Adaptive Subgradient Methods for Mathematical Programming Problems with Quasiconvex Functions, *Trudy Inst. Mat. i Mekh. UrO RAN*, 2023, Volume 29, Number 3, 7–25

DOI: 10.21538/0134-4889-2023-29-3-7-25

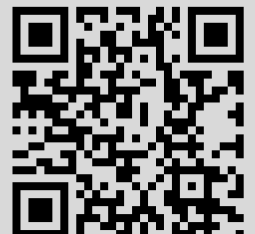
Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

<http://www.mathnet.ru/eng/agreement>

Download details:

IP: 176.57.72.44

February 3, 2025, 13:46:09



УДК 519.85

АДАПТИВНЫЕ СУБГРАДИЕНТНЫЕ МЕТОДЫ ДЛЯ ЗАДАЧ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ С КВАЗИВЫПУКЛЫМИ ФУНКЦИЯМИ¹

С. С. Аблаев, Ф. С. Стонякин, М. С. Алкуса, А. В. Гасников

Работа посвящена субградиентным методам с переключениями по продуктивным и непродуктивным шагам для задач минимизации квазивыпуклых функций при наличии функциональных ограничений-неравенств. Для задачи минимизации выпуклой функции с квазивыпуклыми ограничениями-неравенствами получен результат о сходимости субградиентного метода с адаптивным правилом останова. Далее, на базе аналога острого минимума для нелинейных задач с ограничениями-неравенствами получен результат о сходимости со скоростью геометрической прогрессии рестартованных версий субградиентных методов. Отдельно рассмотрены такие результаты в случае выпуклой целевой функции и квазивыпуклых ограничений, а также в случае квазивыпуклой целевой функции и выпуклых ограничений-неравенств. Выпуклость может позволять дополнительно предлагать адаптивные правила останова вспомогательных методов, гарантирующие достижение приемлемого качества решения. Приведены результаты вычислительных экспериментов, показывающие преимущества использования таких правил останова.

Ключевые слова: субградиентный метод, квазивыпуклая функция, острый минимум, рестарты, адаптивный метод.

S. S. Ablayev, F. S. Stonyakin, M. S. Alkousa, A. V. Gasnikov. Adaptive subgradient methods for mathematical programming problems with quasi-convex functions.

The paper is devoted to subgradient methods with switching over productive and non-productive steps for problems of minimization of quasi-convex functions with functional inequality constraints. For the problem of minimizing a convex function with quasi-convex inequality constraints, result is obtained on the convergence of the subgradient method with an adaptive stopping rule. Further, on the basis of an analogue of a sharp minimum for nonlinear problems with inequality constraints, results are obtained on the convergence with the rate of a geometric progression of restarted versions of subgradient methods. Such results are considered separately in the case of a convex objective function and quasi-convex inequality constraints, as well as in the case of a quasi-convex objective function and convex inequality constraints. The convexity may allow to additionally suggest adaptive stopping rules for auxiliary methods, which guarantee that an acceptable solution quality is achieved. The results of computational experiments are presented, showing the advantages of using such stopping rules.

Keywords: subgradient method, quasi-convex function, sharp minimum, restarts, adaptive method.

MSC: 90C25, 90C06, 49J52

DOI: 10.21538/0134-4889-2023-29-3-7-25

1. Введение. Аналог условия острого минимума для задач минимизации с ограничениями-неравенствами

Задачи математического программирования возникают в самых разнообразных приложениях. Одним из достаточно известных подходов к задачам минимизации функции с ограничениями-неравенствами могут служить так называемые схемы с переключениями по продуктивным и непродуктивным шагам, впервые предложенные Б. Т. Поляком [1]. Общая идея подхода заключается в следующем: если в текущей точке значение ограничения достаточно хорошее, то спуск выполняем по целевой функции, а в противном случае — по функции ограничения.

¹Исследования Ф. С. Стонякаина в разд. 2 и 5 выполнены при поддержке программы стратегического академического лидерства “Приоритет-2030” (соглашение № 075-02-2021-1316 от 30.09.2021). Исследования А. В. Гасникова в разд. 4 в рамках выполнения государственного задания Министерства науки и высшего образования Российской Федерации (проект № 0714-2020-0005).

Такого типа подходам, которые интересны ввиду малых затрат памяти на итерациях, посвящаются все новые работы как для выпуклых задач большой и сверхбольшой размерности [2–5], так и для некоторых классов невыпуклых задач [2].

В последние годы некоторыми из авторов статьи были исследованы адаптивные субградиентные методы с переключениями для липшицевых задач выпуклого программирования, в том числе и для различных предположений о росте целевой функции [3; 6; 7]. В настоящей статье развиваются эти исследования и предлагаются адаптивные методы как на классе задач минимизации выпуклой функции с квазивыпуклыми ограничениями-неравенствами, так и на классе задач минимизации квазивыпуклой функции с выпуклыми ограничениями-неравенствами. Основной упор в теоретическом плане сделан на обосновании сходимости со скоростью геометрической прогрессии для рестартованных версий субградиентных методов такого типа по параметру острого минимума в общности [8]. Однако, как оказалось, адаптивные правила останковки могут на практике приводить к сопоставимой скорости работы при отсутствии требований предполагать острый минимум и знать этот параметр.

Здесь и далее $\alpha > 0$, X — выпуклое замкнутое множество, норма $\|\cdot\|$ евклидова.

Хорошо известно, что предположение острого минимума вида

$$f(x) - f^* \geq \alpha \min_{x_* \in X^*} \|x - x_*\| \quad \forall x \in X, \quad (1.1)$$

где $f^* = f(x_*) = \min_{x \in X} f(x)$, а X^* — множество точек минимума функции f , позволяет доказывать эффективные скоростные гарантии численных методов для негладких задач. Например, для выпуклых (в том числе негладких) липшицевых задач можно доказать сходимость субградиентного метода со скоростью геометрической прогрессии. В частности [9], острым минимумом будет обладать задача вида

$$\begin{cases} \min \langle c, x \rangle, & x \in \mathbb{R}^n, \\ Ax \leq b, & b \in \mathbb{R}^m, \end{cases} \quad (1.2)$$

где c — n -мерный вектор, A — матрица порядка $m \times n$.

Работа в основном посвящена обобщению результатов о сходимости субградиентных методов со скоростью геометрической прогрессии для субградиентных схем с переключениями по продуктивным и непродуктивным шагам в случае как выпуклых, так и квазивыпуклых задач с ограничениями-неравенствами. Если для линейных задач (1.2) возможно обойтись использованием обычного условия острого минимума (1.1), то в общем случае для нелинейных задач это уже не так. Будем рассматривать задачу с функциональными ограничениями вида

$$\begin{cases} \min f(x), \\ x \in X, g(x) \leq 0, \end{cases} \quad (1.3)$$

где $f(x)$ и $g(x)$ — липшицевы функции. Всюду далее будем считать, что задача (1.3) разрешима.

Напомним, что функцию $g(x) : X \rightarrow \mathbb{R}$ называют липшицевой на X с константой $M_g > 0$, если $|g(x) - g(y)| \leq M_g \|x - y\| \quad \forall x, y \in X$, и функцию g называют квазивыпуклой на множестве X , если верно неравенство $g((1 - \alpha)x + \alpha y) \leq \max\{g(x), g(y)\} \quad \forall \alpha \in [0, 1] \quad \forall x, y \in X$.

Для такой постановки (1.3) будем использовать следующую вариацию понятия острого минимума, известную на классе выпуклых задач [8].

О п р е д е л е н и е 1. Будем говорить, что для задачи вида (1.3) выполняется условие острого минимума (“условный” острый минимум), если при некотором $\alpha > 0$ для всех x справедливо неравенство

$$\max\{f(x) - f^*, g(x)\} \geq \alpha \min_{x_* \in X^*} \|x - x_*\|. \quad (1.4)$$

Смысл такого подхода следующий. Поскольку задача (1.3) — с ограничениями-неравенствами, то в тех точках x , где эти неравенства нарушены, возможно $f(x) \leq f^*$, и тогда выполнение неравенства (1.4) возможно за счет положительного g (очевидный пример — когда в качестве ограничения выбирается функция расстояния от точки до допустимого множества). Если же $g(x) \leq 0$, то потенциально возможна ситуация, когда неравенство $f(x) - f^* \geq \alpha \min_{x_* \in X^*} \|x - x_*\|$ справедливо, а на всем допустимом множестве — было бы неверным. Рассмотрим некоторые примеры таких задач с “условным” острым минимумом.

Пример 1. Специальный вариант линейной задачи [8]

$$\min\{-x_1\}, \tag{1.5}$$

$$\rho \cos\left(\frac{j\pi}{10}\right)x_1 + \rho \sin\left(\frac{j\pi}{10}\right)x_2 \leq \rho, \quad j = 0, 1, \dots, 19. \tag{1.6}$$

Как известно (см. [8]), в этой задаче точка минимума $x_* = (1, 0)$, а оптимальное значение $f^* = -1$. Отдельно целевая функция $-x_1$, вообще говоря, не удовлетворяет условию острого минимума (1.1), поскольку зависит только от одной переменной из двух. Но при наличии ограничений (1.6) для этой задачи выполнен “условный” вариант острого минимума (1.4) при $\alpha = \frac{\rho}{2}$ [8]. Выбор масштабирующего коэффициента ρ может влиять на значение параметра такого острого минимума.

Пример 2. Рассмотрим задачу

$$\begin{cases} \min \{f(x_1, x_2) = |x_1| + |x_2|\}, \\ g(x_1, x_2) = \max\{\varepsilon - x_1, \varepsilon - x_2\} \leq 0, \end{cases}$$

где $\varepsilon > 0$. В этом примере целевая функция удовлетворяет условию острого минимума (1.1). Действительно, $f(x) = |x_1| + |x_2|$, $f^* = 0$, $x_* = (0, 0)$ и неравенство $|x_1| + |x_2| \geq \alpha\|x\|$ выполняется, например, при $\alpha = 1$. Однако функциональные ограничения смещают точку минимума $x_* = (\varepsilon, \varepsilon)$, и условие острого минимума нарушается. Но условному острому минимуму эта задача удовлетворяет, учитывая, что $x_* = (\varepsilon, \varepsilon)$, а минимальное значение функции $f^* = 2\varepsilon$.

Пример 3. Задача классификации с ограничениями [8]. Рассмотрим множество, состоящее из n пар: $\mathcal{D} = \{(a_i, b_i)\}_{i=1}^n$, где $a \in \mathbb{R}^p$ — вектор признаков, а $b_i \in \{1, -1\}$ — множество меток при $i = 1, 2, \dots, n$. Пусть $\mathcal{D}_M \in \mathbb{R}^p$, $\mathcal{D}_F \in \mathbb{R}^p$ — два типа признаков. Мы хотим найти линейный классификатор $x \in \mathbb{R}^p$, который не только минимизирует функцию потерь, но и правильным образом обрабатывает каждый элемент из множеств \mathcal{D}_F и \mathcal{D}_M . Указанная задача сводится к следующей постановке:

$$\begin{aligned} \min_x \frac{1}{n} \sum_{i=1}^n \max_x \{0, 1 - b_i a_i^T x\}, \\ \frac{1}{n_F} \sum_{a \in \mathcal{D}_F} \sigma(a^T x) \geq \frac{\kappa}{n_M} \sum_{a \in \mathcal{D}_M} \sigma(a^T x), \\ \frac{1}{n_M} \sum_{a \in \mathcal{D}_M} \sigma(a^T x) \geq \frac{\kappa}{n_F} \sum_{a \in \mathcal{D}_F} \sigma(a^T x), \end{aligned}$$

где $\kappa \in (0, 1]$ — константа, n_M и n_F — количество экземпляров из \mathcal{D}_M и \mathcal{D}_F соответственно и $\sigma(a^T x) := \max\{0, \min\{1, \{0.5 + a^T x\}\}\} \in [0, 1]$ — вероятность присваивания метки +1 предсказанию a .

Напомним далее некоторые вспомогательные понятия и результаты.

Для всякого субградиента $\nabla f(x)$ выпуклой целевой функции $f(x)$ для точки $y \in X$ введем в рассмотрение вспомогательную величину аналогично [10]:

$$v_f(x, y) = \begin{cases} \left\langle \frac{\nabla f(x)}{\|\nabla f(x)\|}, x - y \right\rangle, & \nabla f(x) \neq 0, \\ 0, & \nabla f(x) = 0, \end{cases} \quad x \in X. \quad (1.7)$$

Как показано в [10], при $y = x_*$ величина $v_f(x, x_*)$ может быть равна длине проекции точки x_* на гиперплоскость, ортогональную вектору $\nabla f(x)$. Если численный метод может выдавать точку x_k , для которой $v_f(x_k, x_*) \leq \varepsilon$, то для такой точки в случае предположения о непрерывности функции можно гарантировать достижение приемлемого качества решения. При этом можно рассматривать не только выпуклые, но и квазивыпуклые задачи.

Лемма 1 ([10], теорема 1.5.5). Пусть функция $f(x)$ квазивыпукла и $f(x) \geq f(x_*) \quad \forall x \in X$. Определим следующую функцию: $\omega(\tau) = \max_{x \in X} \{f(x) - f(x_*) : \|x - x_*\| \leq \tau\}$, где τ — положительное число. Тогда для любого $y \in X$

$$f(y) - f(x_*) \leq \omega(v_f(y, x_*)).$$

В качестве субградиента локально липшицевой квазивыпуклой функции f можно рассматривать, например, элемент субдифференциала Кларка [11] в соответствующей точке. Для квазивыпуклых задач возможны ситуации, когда (суб)градиент $\nabla f(x)$ равен 0 или не существует, но минимума в точке x нет. В таком случае в методах градиентного типа вместо субградиента можно рассматривать согласно [10] всякий ненулевой элемент из набора векторов нормалей ко множеству уровня функции f в точке x : $\hat{D}f(x) = \{p \mid \langle p, x - y \rangle \geq 0 \quad \forall y \in X : f(y) \leq f(x)\}$. Это множество есть непустой замкнутый выпуклый конус. Далее будем обозначать через $Df(x)$ произвольный ненулевой вектор из $\hat{D}f(x)$.

Работа состоит из введения, заключения и четырех основных разделов. В разд. 2 рассматривается новый адаптивный метод (алгоритм 1) для задач минимизации выпуклой целевой функции в случае квазивыпуклых ограничений-неравенств и его рестартованная версия (алгоритм 2) для случая вариации острого минимума. Выпуклость целевой функции и адаптивность метода позволяют применять данные методы на практике без знания константы Липшица целевой функции M_f . Раздел 3 посвящен ситуации, когда, наоборот, целевая функция квазивыпукла, а ограничения — выпуклы. В нем рассмотрены алгоритмы 3 и 4, которые можно применять без знания константы Липшица ограничения M_g . Наконец, в разд. 4 исследуется самая общая ситуация, когда квазивыпуклыми могут быть как f , так и g (алгоритмы 5 и 6). Но для реализации алгоритма 6 необходимо знать обе константы Липшица M_f и M_g . Отметим, что полученные в статье результаты для рестартованных по параметру “условного” острого минимума (1.4) алгоритмов новы и для задач выпуклого программирования, поскольку ранее схемы с переключениями не рассматривались для задач с острым минимумом.

2. Случай выпуклой целевой функции и квазивыпуклых ограничений

В этом разделе исследуем ситуацию, когда в задаче (1.3) $f(x)$ — выпуклая функция, а $g(x)$ — квазивыпуклая функция.

Пусть $\text{Pr}_X(y) := \arg \min_{x \in X} \{\|y - x\|\}$ — оператор проектирования на множество X . Обозначим через I и J множество индексов продуктивных и непродуктивных шагов соответственно. $|I|$ и $|J|$ обозначают количество продуктивных и непродуктивных шагов соответственно. Рассмотрим следующий алгоритм. В нем и всюду далее x_* — некоторое решение задачи (например, ближайшее к x_0).

Algorithm 1. Адаптивный субградиентный метод для выпуклой целевой функции

Require: $\delta > 0, M_g > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2}\|x_* - x_0\|^2$, множество X .

1: $I =: \emptyset$
 2: $N \leftarrow 0$
 3: **repeat**
 4: **if** $g(x_N) \leq \delta M_g$ **then**
 5: $h_N^f = \frac{\delta}{\|\nabla f(x_N)\|^2}$,
 6: $x_{N+1} = \text{Pr}_X(x_N - h_N^f \nabla f(x_N))$, // продуктивные шаги
 7: $N \rightarrow I$,
 8: **else**
 9: $h_N^g = \frac{\delta}{\|Dg(x_N)\|}$,
 10: $x_{N+1} = \text{Pr}_X(x_N - h_N^g Dg(x_N))$, // непродуктивные шаги
 11: **end if**
 12: $N \leftarrow N + 1$,
 13: **until** $\frac{2\theta_0^2}{\delta^2} \leq \sum_{k \in I} \frac{1}{\|\nabla f(x_k)\|^2} + N - |I|$.

Ensure: $\hat{x} := \arg \min_{x_k, k \in I} f(x_k)$.

Справедлива следующая теорема.

Теорема 1. После остановки алгоритма 1 для всякого M_g -липшицева квазивыпуклого ограничения g верно $f(\hat{x}) - f(x_*) \leq \delta$ и $g(\hat{x}) \leq \delta M_g$. При этом в случае M_f -липшицевой функции f достаточное количество итераций для выполнения критерия остановки алгоритма 1 оценивается следующим образом:

$$N \geq \frac{2\theta_0^2 \max\{1, M_f^2\}}{\delta^2}.$$

Доказательство. Если $k \in I$, то для всякого $x_* \in X_*$ имеем согласно ([12], соотношение (3) из доказательства теоремы 1)

$$\begin{aligned} h_k(f(x_k) - f(x_*)) &\leq h_k \langle \nabla f(x_k), x_k - x_* \rangle \leq \frac{h_k^2}{2} \|\nabla f(x_k)\|^2 + \frac{1}{2} \|x_k - x_*\|^2 - \frac{1}{2} \|x_{k+1} - x_*\|^2 \\ &= \frac{\delta^2}{2} \frac{1}{\|\nabla f(x_k)\|^2} + \frac{1}{2} \|x_k - x_*\|^2 - \frac{1}{2} \|x_{k+1} - x_*\|^2. \end{aligned} \quad (2.1)$$

Если $k \notin I$, то $g(x_k) > \delta M_g$ и $\delta < \frac{g(x_k)}{M_g} \leq \frac{g(x_k) - g(x_*)}{M_g} \leq v_g(x_*, x_k)$; $v_g(x_*, x_k)$ определяется по аналогии с (1.7).

Поэтому верны неравенства

$$\delta^2 < h_k v_g(x_*, x_k) \leq \frac{h_k^2}{2} \|Dg(x_k)\|^2 + \frac{1}{2} \|x_k - x_*\|^2 - \frac{1}{2} \|x_{k+1} - x_*\|^2,$$

или

$$\frac{\delta^2}{2} < \frac{1}{2} \|x_k - x_*\|^2 - \frac{1}{2} \|x_{k+1} - x_*\|^2. \quad (2.2)$$

После суммирования неравенств (2.1) и (2.2) имеем

$$\begin{aligned} \sum_{k \in I} h_k(f(x_k) - f(x_*)) &\leq \sum_{k \in I} \frac{\delta^2}{2\|\nabla f(x_k)\|^2} - \frac{\delta^2 |J|}{2} + \frac{1}{2} \|x_0 - x_*\|^2 - \frac{1}{2} \|x_{k+1} - x_*\|^2 \\ &\leq \frac{\delta}{2} \sum_{k \in I} h_k - \frac{\delta^2 |J|}{2} + \theta_0^2 = \delta \sum_{k \in I} h_k - \frac{\delta^2}{2} \left(\sum_{k \in I} \frac{1}{\|\nabla f(x_k)\|^2} + |J| \right) + \theta_0^2. \end{aligned}$$

После выполнения критерия остановки алгоритма 1 будет верно

$$\sum_{k \in I} h_k (f(x_k) - f(x_*)) \leq \delta \sum_{k \in I} h_k,$$

откуда справедливо $f(\hat{x}) - f(x_*) \leq \delta$.

При этом $g(x_k) \leq \delta M_g \forall k \in I$, поэтому $g(\hat{x}) \leq \delta M_g$.

Покажем, что множество продуктивных шагов I непусто. Если $I = \emptyset$, то $|J| = N$ означает, что $N \geq \frac{2\theta_0^2}{\delta^2}$. С другой стороны, $\frac{\delta^2 N}{2} < \frac{1}{2} \|x_0 - x_*\|^2 \leq \theta_0^2$. Получили противоречие, следовательно, $I \neq \emptyset$.

Оценим достаточное число итераций для выполнения критерия остановки алгоритма 1. Если функция $f(x)$ M_f -липшицева, то $\|\nabla f(x_k)\| \leq M_f \forall k \in I$. Следовательно,

$$\sum_{k \in I} \frac{1}{\|\nabla f(x_k)\|^2} + |J| \geq \frac{|I|}{M_f^2} + |J| \geq (|I| + |J|) \min \left\{ 1, \frac{1}{M_f^2} \right\} = \frac{N}{\max\{1, M_f^2\}}.$$

Теорема доказана.

З а м е ч а н и е 1. На практике для реализации алгоритма 1 знание константы Липшица M_f функции $f(x)$ необязательно, достаточно остановить алгоритм после выполнения критерия остановки.

З а м е ч а н и е 2. В случае выпуклых $f(x)$ и $g(x)$ можно рассматривать следующую точку выхода алгоритма 1: $\hat{x} = \frac{1}{\sum_{k \in I} h_k} \sum_{k \in I} h_k x_k$.

Построим схему рестартов алгоритма 1 — алгоритм 2 — в предположении, что верно условие (1.4).

Algorithm 2. Рестарты алгоритма 1

Require: $\varepsilon > 0, \alpha > 0, M_g > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2} \|x_* - x_0\|^2$, множество X .

1: Set $p = 1$.

2: **repeat**

3: \hat{x}^p — результат работы алгоритма 1 с параметрами δ_p, θ_p, x_0 , где

4: $x_0 = \hat{x}^p$,

5: $\theta_p = \frac{1}{\sqrt{2^p}} \theta_0$,

6: $\delta_p = \frac{\alpha \theta_p}{\sqrt{2} \max\{1, M_g\}}$.

7: Set $p = p + 1$.

8: **until** $p > \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil$.

Ensure: x^p .

Для алгоритмов 1 и 2 верна следующая

Теорема 2. Пусть $f(x)$ и $g(x)$ — липшицевы функции с константами M_f и M_g соответственно, удовлетворяющие условию (1.4), и известна константа $\theta_0 > 0$ такая, что $2\theta_0^2 \geq \|x_* - x_0\|^2$. Тогда для алгоритма 1 можно подобрать параметр $\delta > 0$ так, чтобы после

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil$$

итераций было выполнено неравенство $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$. После $p - 1$ запусков алгоритма 2 (рестартов алгоритма 1) имеем

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}}\theta_0.$$

Тогда для достижения ε -точного решения вида $\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \varepsilon$ достаточно количество обращений к субградиенту f или g можно оценить как

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Доказательство. Выполнение критерия останова алгоритма 1 влечет неравенство $\max\{f(\hat{x}) - f^*, g(\hat{x})\} \leq \delta \max\{1, M_g\}$. Отсюда, с учетом условия острого минимума (1.4), вытекает соотношение

$$\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{\delta \max\{1, M_g\}}{\alpha}.$$

Выбрав

$$\delta = \frac{\alpha\theta_0}{\sqrt{2} \max\{1, M_g\}}, \quad (2.3)$$

получим, что $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$.

Подставим выбранное значение δ в критерий останова алгоритма 1. Имеем

$$\frac{2\theta_0^2}{\delta^2} = \frac{2\theta_0^2}{\left(\frac{\alpha\theta_0}{\sqrt{2} \max\{1, M_g\}}\right)^2} = \frac{4 \max\{1, M_g^2\}}{\alpha^2}.$$

Критерий останова принимает вид

$$\frac{4 \max\{1, M_g^2\}}{\alpha^2} \leq \sum_{k \in I} \frac{1}{\|\nabla f(x_k)\|^2} + |J|.$$

С ростом p происходит изменение параметров δ_p , θ_p и \hat{x}^p , но оценка достаточного числа итераций сохраняется для всех p , поскольку величина $\frac{\theta_p}{\delta_p}$ не зависит от p :

$$\frac{\theta_p}{\delta_p} = \frac{\theta_p \sqrt{2} \max\{1, M_g\}}{\alpha \theta_p} = \frac{\sqrt{2} \max\{1, M_g\}}{\alpha}.$$

Следовательно, после

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \quad (2.4)$$

итераций критерий останова алгоритма 1 заведомо выполнен.

Итак, после первого запуска алгоритма с начальной точкой x_0 при δ , как в (2.3), выполняется неравенство $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$. Рестартуем метод с начальной точкой $\hat{x}^0 = \hat{x}$. Тогда после (2.4) шагов

$$\min_{x_* \in X_*} \|\hat{x}^1 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_1 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\theta_0 = \frac{1}{2}\theta_0.$$

После второго перезапуска с начальной точкой \hat{x}^1 будет

$$\min_{x_* \in X_*} \|\hat{x}^2 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_2 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{2}\theta_0 = \frac{1}{\sqrt{2^3}}\theta_0.$$

После третьего — $\min_{x_* \in X_*} \|\hat{x}^3 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_3 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2^3}}\theta_0 = \frac{1}{2^2}\theta_0$. Методом математической индукции покажем, что после $p - 1$ рестартов алгоритма 1 будет справедливо неравенство

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}}\theta_0.$$

Базу индукции мы проверили для $p = 1$. Пусть после $p - 2$ рестартов алгоритма 1 верно следующее неравенство: $\min_{x_* \in X_*} \|\hat{x}^{p-2} - x_*\| \leq \frac{1}{\sqrt{2^{p-1}}}\theta_0$. Тогда после $(p - 1)$ -го рестарта алгоритма 1 имеем

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_{p-2} \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2^{p-1}}}\theta_0 = \frac{1}{\sqrt{2^p}}\theta_0.$$

Отметим, что $\frac{1}{\sqrt{2^p}}\theta_0 \leq \varepsilon$ при $p = \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil$. Общее, достаточное для достижения требуемой точности, количество обращений к субградиенту f или g можно оценить как произведение количества итераций алгоритма 1 на количество этих рестартов, т. е.

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Теорема доказана.

3. Случай квазивыпуклой целевой функции и выпуклых ограничений

В данном разделе мы рассмотрим адаптивный алгоритм для решения квазивыпуклых задач минимизации с выпуклыми ограничениями при условии, что целевая функция удовлетворяет предложенному условию острого минимума (1.4). Для указанного класса задач (1.3) рассмотрим следующий алгоритм ([7], Algorithm 1).

Algorithm 3. Субградиентный метод для квазивыпуклых задач с ограничениями.

Require: $\delta > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2}\|x_* - x_0\|^2$, множество X .

- 1: $I =: \emptyset$
- 2: $N \leftarrow 0$
- 3: **repeat**
- 4: **if** $g(x_N) \leq \delta$ **then**
- 5: $h_N^f = \frac{\delta}{\|Df(x_N)\|}$,
- 6: $x_{N+1} = \text{Pr}_X(x_N - h_N^f Df(x_N))$, // продуктивные шаги
- 7: $N \rightarrow I$
- 8: **else**
- 9: $h_N^g = \frac{\delta}{\|\nabla g(x_N)\|^2}$,
- 10: $x_{N+1} = \text{Pr}_X(x_N - h_N^g \nabla g(x_N))$, // непродуктивные шаги
- 11: **end if**
- 12: $N \leftarrow N + 1$,
- 13: **until** $\frac{2\theta_0^2}{\delta^2} \leq |I| + \sum_{k \notin I} \frac{1}{\|\nabla g(x_k)\|^2}$.

Ensure: $\hat{x} := \arg \min_{x_k, k \in I} f(x_k)$.

Построим схему рестартов алгоритма 3 — алгоритм 4 — по параметру “условного” острого минимума по аналогии с алгоритмом 2 (отличие в выборе параметра δ_p). Отметим, что с каждым перезапуском происходит обновление параметров δ_p, θ_p, x_0 .

Algorithm 4. Рестарты алгоритма 3

Require: $\varepsilon > 0, \alpha > 0, M_f > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2} \|x_* - x_0\|^2$, множество X .

1: \hat{x}^p — результат работы алгоритма 3.

2: $\delta_p = \frac{\alpha \theta_p}{\sqrt{2} \max\{1, M_f\}}$.

Ensure: x^p .

Из ([7], теорема 1) известно, что если $f(x)$ и $g(x)$ — липшицевы функции, то после $N = \left\lceil \frac{2 \max\{1, M_g^2\} \theta_0^2}{\delta^2} \right\rceil$ итераций алгоритма 3 справедливо неравенство $\min_{k \in I} v_f(x_k, x_*) < \delta$, откуда

$$\min_{0 \leq k \leq N} f(x_k) - f(x_*) \leq \delta M_f,$$

где M_f — константа Липшица функции $f(x)$. Отсюда следует, что после $N = \left\lceil \frac{2 \max\{1, M_g^2\} \theta_0^2}{\delta^2} \right\rceil$ шагов достигается δ -точное решение задачи (1.3):

$$\begin{cases} f(\hat{x}) - f^* \leq \delta M_f, \\ g(\hat{x}) \leq \delta. \end{cases}$$

Для алгоритмов 3 и 4 имеет место следующий результат.

Теорема 3. Пусть $f(x)$ и $g(x)$ — липшицевы функции с константами M_f и M_g соответственно, удовлетворяющие условию (1.4), и известна константа $\theta_0 > 0$ такая, что $2\theta_0^2 \geq \|x_* - x_0\|^2$ для некоторого решения $x_* \in X_*$. Тогда для алгоритма 3 можно подобрать параметр $\delta > 0$ так, чтобы после

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil$$

итераций было выполнено неравенство $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}} \theta_0$. После $p - 1$ запусков алгоритма 4 (рестартов алгоритма 3) имеем

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}} \theta_0.$$

Тогда для достижения ε -точного решения вида $\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \varepsilon$ достаточно количество обращений к субградиенту f или g можно оценить как

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Доказательство. Выполнение критерия остановки алгоритма 3 влечет неравенство $\max\{f(\hat{x}) - f^*, g(\hat{x})\} \leq \delta \max\{1, M_f\}$. Отсюда, с учетом условия острого минимума (1.4), вытекает соотношение

$$\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{\delta \max\{1, M_f\}}{\alpha}.$$

Выбрав

$$\delta = \frac{\alpha \theta_0}{\sqrt{2} \max\{1, M_f\}}, \tag{3.1}$$

получим, что $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}} \theta_0$.

Подставим выбранное значение δ в критерий остановки алгоритма 3. Имеем

$$\frac{2\theta_0^2}{\delta^2} = \frac{2\theta_0^2}{\left(\frac{\alpha\theta_0}{\sqrt{2}\max\{1, M_f\}}\right)^2} = \frac{4\max\{1, M_f^2\}}{\alpha^2}.$$

Критерий остановки принимает вид

$$\frac{4\max\{1, M_f^2\}}{\alpha^2} \leq |I| + \sum_{k \notin I} \frac{1}{\|\nabla g(x_k)\|^2}. \quad (3.2)$$

С ростом p происходит изменение параметров δ_p , θ_p и \hat{x}^p , но оценка достаточного числа итераций сохраняется для всех p , поскольку величина $\frac{\theta_p}{\delta_p}$ не зависит от p :

$$\frac{\theta_p}{\delta_p} = \frac{\theta_p \sqrt{2} \max\{1, M_f\}}{\alpha \theta_p} = \frac{\sqrt{2} \max\{1, M_f\}}{\alpha}.$$

Следовательно, достаточное количество итераций для выполнения критерия остановки алгоритма 3 оценивается сверху следующим образом: $N = \left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil$. После

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \quad (3.3)$$

итераций критерий остановки алгоритма 3 заведомо выполнен.

Итак, после первого запуска алгоритма 3 с начальной точкой x_0 при δ , как в (3.1), выполняется неравенство $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$. Рестартуем метод с начальной точкой $\hat{x}^0 = \hat{x}$. Тогда после (3.3) шагов $\min_{x_* \in X_*} \|\hat{x}^1 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_1 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\theta_0 = \frac{1}{2}\theta_0$. После второго перезапуска с начальной точкой \hat{x}^1 будет

$$\min_{x_* \in X_*} \|\hat{x}^2 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_2 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{2}\theta_0 = \frac{1}{\sqrt{2^3}}\theta_0.$$

После третьего — $\min_{x_* \in X_*} \|\hat{x}^3 - x_*\| \leq \frac{1}{\sqrt{2}}\theta_3 \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2^3}}\theta_0 = \frac{1}{2^2}\theta_0$. Методом математической индукции покажем, что после $p - 1$ рестартов алгоритма 3 будет справедливо неравенство

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}}\theta_0.$$

Базу индукции мы проверили для $p = 1$. Пусть после $p - 2$ рестартов алгоритма 3 верно следующее неравенство: $\min_{x_* \in X_*} \|\hat{x}^{p-2} - x_*\| \leq \frac{1}{\sqrt{2^{p-1}}}\theta_0$. Тогда после $p - 1$ рестартов алгоритма 3 имеем

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_{p-2} \leq \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2^{p-1}}}\theta_0 = \frac{1}{\sqrt{2^p}}\theta_0.$$

Отметим, что $\frac{1}{\sqrt{2}}\theta_0 \leq \varepsilon$ при $p = \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil$. Достаточное количество обращений к субградиенту f или g можно оценить как произведение количества итераций алгоритма 3 на количество этих рестартов, т. е.

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Теорема доказана.

Теорема 3 показывает оценку количества итераций алгоритма 3 с учетом введенного условия острого минимума, а также оценку алгоритма 4 по аргументу и общее число итераций.

З а м е ч а н и е 3. На практике для реализации алгоритмов 3 и 4 знание константы Липшица M_g функции $g(x)$ необязательно. Достаточно рестартовать алгоритм 3 согласно листингу алгоритма 4 с обновлением значений параметров после выполнения критерия остановки (3.2).

4. Случай квазивыпуклой целевой функции и квазивыпуклых ограничений

В этом разделе исследуем задачу (1.3) в случае, когда и $f(x)$, и $g(x)$ — квазивыпуклые функции, удовлетворяющие условию Липшица с константами M_f и M_g соответственно.

Рассмотрим следующий алгоритм ([7], Algorithm 3).

Algorithm 5. Случай квазивыпуклых ограничений

Require: $\delta > 0, M_g > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2}\|x_* - x_0\|^2$, множество X .

```

1:  $I =: \emptyset$ 
2:  $N \leftarrow 0$ 
3: repeat
4:   if  $g(x_N) \leq \delta M_g$  then
5:      $h_N^f = \frac{\delta}{\|Df(x_N)\|}$ ,
6:      $x_{N+1} = \text{Pr}_X(x_N - h_N^f Df(x_N))$ , // продуктивные шаги
7:      $N \rightarrow I$ 
8:   else
9:      $h_N^g = \frac{\delta}{\|Dg(x_N)\|}$ ,
10:     $x_{N+1} = \text{Pr}_X(x_N - h_N^g Dg(x_N))$ , // непродуктивные шаги
11:   end if
12:    $N \leftarrow N + 1$ ,
13: until  $\frac{2\theta_0^2}{\delta^2} \leq N$ .

```

Ensure: $\hat{x} := \arg \min_{x_k, k \in I} f(x_k)$.

Построим схему рестартов алгоритма 5 по параметру “условного” острого минимума по аналогии с алгоритмом 2 (отличие в выборе параметра δ_p).

Algorithm 6. Схема рестартов алгоритма 5

Require: $\varepsilon > 0, \alpha > 0, M_f > 0, M_g > 0, x_0, \theta_0 : \theta_0^2 \geq \frac{1}{2}\|x_* - x_0\|^2$, множество X .

1: \hat{x}^p — результат работы алгоритма 5

2: $\delta_p = \frac{\alpha \theta_p}{\sqrt{2} \max\{M_f, M_g\}}$,

Ensure: x^p .

В ([7], теорема 3) было установлено, что если $f(x)$ — квазивыпуклая функция, $g(x)$ — квазивыпуклая липшицева функция с константой Липшица M_g , то для алгоритма 5 после $N = \left\lceil \frac{2\theta_0^2}{\delta^2} \right\rceil$ шагов выполняются неравенства $\min_{k \in I} v_f(x_k, x_*) \leq \delta$, $\max_{k \in I} g(x_k) \leq \delta M_g$. Следовательно, после остановки алгоритма 5 будет верна оценка

$$f(\hat{x}) - f^* \leq \delta M_f, \quad g(\hat{x}) \leq \delta M_g. \quad (4.1)$$

Использование схемы рестартов алгоритма 5 по параметру “условного” острого минимума позволяет доказать следующий результат для алгоритмов 5, 6.

Теорема 4. Пусть $f(x)$ и $g(x)$ — квазивыпуклые липшицевы функции с константами $M_f > 0$ и $M_g > 0$ соответственно, удовлетворяющие условию (1.4), и известна константа $\theta_0 > 0$ такая, что $2\theta_0^2 \geq \|x_* - x_0\|^2$. Тогда для алгоритма 5 можно подобрать параметр $\delta > 0$ так, чтобы после

$$\left\lceil \frac{4 \max\{M_f^2, M_g^2\}}{\alpha^2} \right\rceil$$

итераций было гарантированно верно неравенство $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$. После $p-1$ запусков алгоритма 6 верна оценка

$$\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}}\theta_0.$$

Тогда для достижения ε -точного решения вида $\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \varepsilon$ достаточное количество обращений к субградиенту f или g можно оценить как

$$\left\lceil \frac{4 \max\{M_f^2, M_g^2\}}{\alpha^2} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Доказательство. Неравенства (4.1) приводят к оценке $\max\{f(\hat{x}) - f^*, g(\hat{x})\} \leq \delta \max\{M_f, M_g\}$. Согласно условию острого минимума (1.4) имеем

$$\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{\delta \max\{M_f, M_g\}}{\alpha}.$$

Выбрав $\delta = \frac{\alpha\theta_0}{\sqrt{2} \max\{M_f, M_g\}}$, получим, что $\min_{x_* \in X_*} \|\hat{x} - x_*\| \leq \frac{1}{\sqrt{2}}\theta_0$.

В данном случае критерий останова имеет вид

$$\frac{4 \max\{M_f^2, M_g^2\}}{\alpha^2} \leq N.$$

Достаточное количество итераций для выполнения критерия останова алгоритма 5 можно оценить сверху следующим образом: $N = \left\lceil \frac{4 \max\{M_f^2, M_g^2\}}{\alpha^2} \right\rceil$.

После $p-1$ рестартов алгоритма 5 получим $\min_{x_* \in X_*} \|\hat{x}^{p-1} - x_*\| \leq \frac{1}{\sqrt{2^p}}\theta_0$. Отметим, что $\frac{1}{\sqrt{2^p}}\theta_0 \leq \varepsilon$ при $p = \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil$.

Общее количество обращений к субградиенту f или g можно оценить как произведение количества итераций алгоритма 5 на число рестартов, т. е.

$$\left\lceil \frac{4 \max\{M_f^2, M_g^2\}}{\alpha^2} \right\rceil \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil.$$

Теорема доказана.

5. Численные эксперименты

В данном разделе для иллюстрации работоспособности предложенных выше алгоритмов приведем некоторые результаты вычислительных экспериментов при различных вариантах выпуклости целевой функции и функциональных ограничений. В подразд. 5.1 мы продемонстрируем работу алгоритмов 1–6 в случае, когда целевая функция и функциональные ограничения выпуклы, но будет проверено выполнение “условного” острого минимума (доказанные в статье теоретические результаты новы и для такого класса задач). Далее, в подразд. 5.2 рассмотрим работу алгоритмов 3 и 5 для примера квазивыпуклой целевой функции с выпуклыми

ограничениями-неравенствами. Наконец, в подразд. 5.3 мы покажем работу алгоритмов 1 и 5, когда целевая функция выпукла, а функциональные ограничения квазивыпуклы. При этом подбор ограничений приводит к неочевидности положения минимума и проверке условия типа острого минимума, что естественно подтверждает справедливость идеи использовать предложенные правила останковки, которые гарантируют достижение приемлемого качества решения.

Все эксперименты проводились на Python 3.4 на компьютере с Intel(R) Core(TM) i7-8550U CPU (1.80GHz, 4 ядра, 8 потоков). Оперативная память компьютера составляла 8 ГБ.

5.1. Выпуклые задачи с “условным” острым минимумом

Начнем с описания результатов экспериментов для работы алгоритмов 1–6 в случае, когда удастся доказать выполнение “условного” острого минимума (1.4). В первом из таких примеров целевая функция имеет вид (1.5), (1.6) (см. пример 1) [8] для различных значений параметра $\rho > 0$ со следующей функцией-ограничением:

$$g(x) := \max_{j=0,1,\dots,19} \left\{ g_j(x) = \rho \cos\left(\frac{j\pi}{10}\right)x_1 + \rho \sin\left(\frac{j\pi}{10}\right)x_2 - \rho \right\} \leq 0.$$

Рассмотрим еще такой пример уже в пространстве большой размерности.

Пример 4. Рассмотрим функции вида

$$f(x) = \|x\| + \max\{-\langle a, x \rangle, \|x\|\}, \tag{5.1}$$

где $a \in \mathbb{R}^n$ — фиксированный вектор, а ограничения имеют следующий вид:

$$g(x) = \langle a, x \rangle \leq 0. \tag{5.2}$$

Для этого примера мы имеем $\alpha = 1$, $M_f = 1 + \max\{\|a\|, 1\}$, $M_g = \|a\|$, $x^* = \mathbf{0} \in \mathbb{R}^n$ и $f^* = f(x^*) = 0$.

Выберем в качестве множества X евклидов шар из \mathbb{R}^n с центром в $\mathbf{0} \in \mathbb{R}^n$ радиуса 10. Для примера 1 мы запускали алгоритмы 1–6 со стартовой точкой $x_0 = (0, 0)$ и различными значениями ρ . Для примера 4 мы запускали эти алгоритмы со стартовой точкой $x_0 = \left(\frac{10}{\sqrt{n}}, \dots, \frac{10}{\sqrt{n}}\right) \in \mathbb{R}^n$ и $n = 1000$. Вектор a в (5.1) и (5.2) сгенерирован случайным образом с равномерным распределением на интервале $[0, 1)$.

Результаты работы этих алгоритмов с различными значениями ε (которые являются значениями δ в алгоритмах 1, 3 и 5) представлены на рис. 1 и 2. Эти результаты демонстрируют количество выполненных итераций до выполнения критерия останковки, теоретически гарантирующего соответствующую точность ε (справа) и наблюдаемые на практике значения целевой функции $f(\hat{x}) = \hat{f}$ в точке выхода \hat{x} каждого алгоритма (слева).

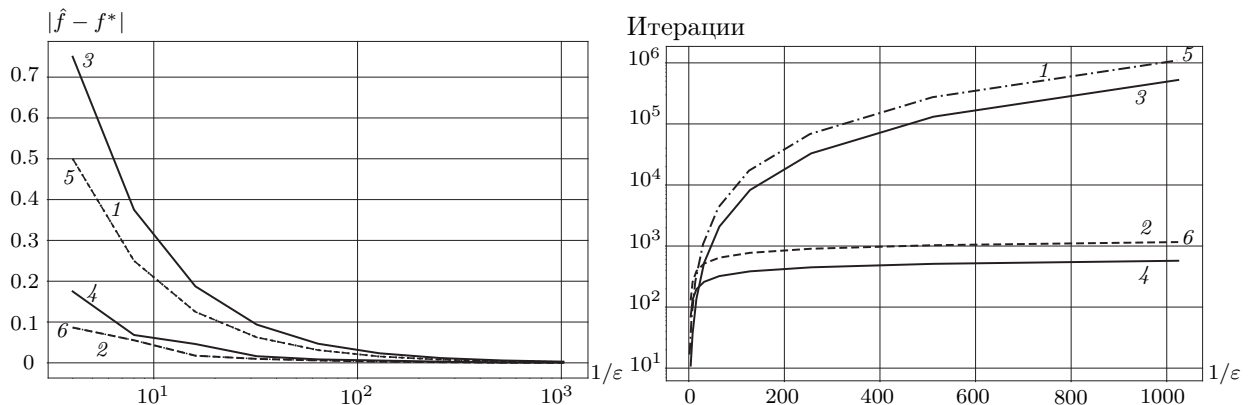
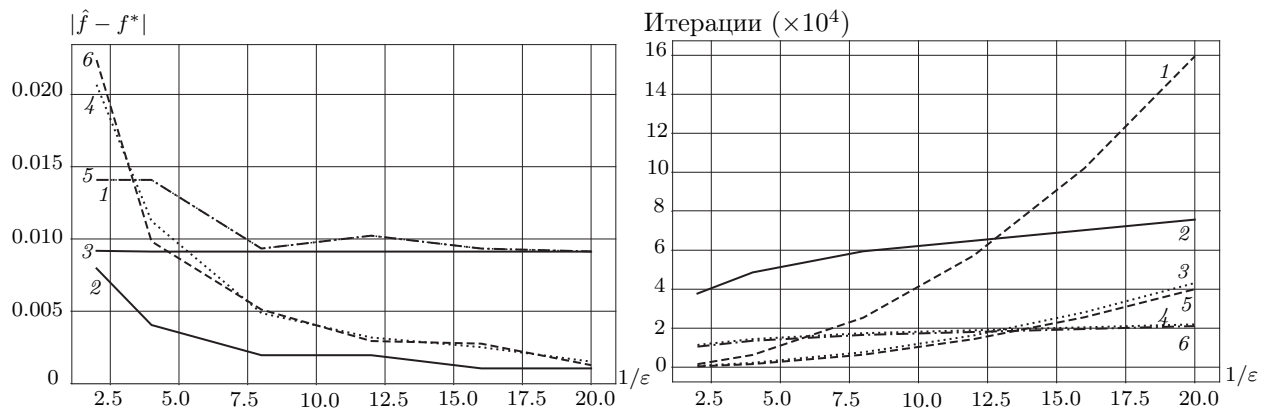


Рис. 1. Результаты алгоритмов 1–6 для примера 1 с $\rho = 0.5$.

Рис. 2. Результаты алгоритмов 1–6 для примера 4 с $n = 1000$.

На рисунках номера линий являются номерами соответствующих им алгоритмов.

На рис. 1 и 2 мы можем увидеть результаты работы предложенных алгоритмов 2, 4 и 6, которые основаны на технике рестартов по параметру острого минимума. Отметим, что увеличение числа итераций алгоритмов с перезапусками практически не происходит при уменьшении значения ϵ . Это указывает на преимущества рестартованных версий алгоритмов 1, 3 и 5. Однако применение адаптивных правил остановки алгоритмов 1, 3 и 5 без рестартов тоже приводит к неплохим результатам, но при этом не предполагает столь существенных требований, как знание параметра острого минимума.

Для примера 1 по рис. 1 можно увидеть, что алгоритмы 2 и 6 (которые эквивалентны в проведенных экспериментах) работают лучше, чем алгоритм 4. Но алгоритм 4 работает быстрее, чем алгоритмы 2 и 6, потому что ему требуется меньше итераций для достижения критерия остановки. Также отметим, что значение параметра ρ (т.е. значение параметра острого минимума α) экспериментально влияет на работу алгоритмов 1, 3 и 5, но не влияет на работу алгоритмов с рестартами. С увеличением значения ρ значения целевой функции в выходных точках алгоритма 3 точнее, чем у алгоритмов 1 и 5, но он работает медленнее.

Кроме того, для примера 4 на рис. 2 видно, что алгоритмы 4 и 6 работают быстрее, чем алгоритм 2, но алгоритм 2 выдает наилучшие значения целевой функции на выходе. Также мы видим, что алгоритм 3 работает лучше, чем алгоритмы 1 и 5.

5.2. Минимизация квазивыпуклой целевой функции с выпуклыми ограничениями-неравенствами

Приведем теперь результаты экспериментов для алгоритмов 3 и 5 для следующего примера [13] с квазивыпуклой целевой функцией.

Пример 5. Предположим, что $a, b \in \mathbb{R}^n$, и определим следующую целевую функцию:

$$f(x) = \frac{\|x - a\|}{\|x - b\|}, \quad (5.3)$$

т.е. отношение евклидова расстояния до a к расстоянию до b . Функция f квазивыпукла на полупространстве $\{x \in \mathbb{R}^n : \|x - a\| \leq \|x - b\|\}$.

Функция ограничений имеет вид

$$g(x) = \max_{1 \leq i \leq m} \{g_i(x) = \langle \alpha_i, x \rangle + \beta_i, \alpha_i \in \mathbb{R}^n, \beta_i \in \mathbb{R}\}. \quad (5.4)$$

Для этого примера мы принимаем $a = \mathbf{0} \in \mathbb{R}^n$, точка $b \in \mathbb{R}^n$ выбрана так, что $\|b - a\| = 10$ и в качестве множества X возьмем шар из \mathbb{R}^n с центром в $a = \mathbf{0} \in \mathbb{R}^n$ радиуса 5.

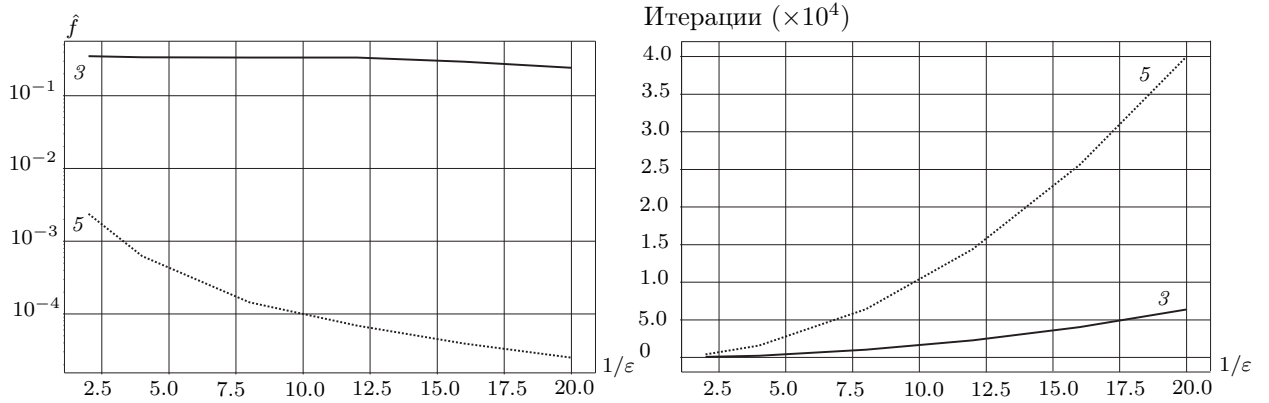


Рис. 3. Результаты алгоритмов 3 и 5 для примера 5 с $n = 1000, m = 10$.

Алгоритмы 3 и 5 запускались с начальной точкой $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ при $n = 1000, m = 10$. Коэффициенты α_i генерируются случайным образом с нормальным (гауссовским) распределением с центром 0 и стандартным отклонением 0.01, константы β_i в (5.4) генерируются случайно с равномерным распределением на интервале $[-10, 10)$. В этом случае оптимальное значение f^* неизвестно.

Результаты работы алгоритмов 3 и 5 представлены на рис. 3. Приводятся значения целевой функции (5.3) в точке выхода \hat{x} каждого алгоритма, где $\hat{f} = f(\hat{x})$ (слева) и количество выполненных итераций до выполнения критерия останова, теоретически гарантирующего соответствующую точность δ (справа).

Как показано на рис. 3, алгоритм 5 приводит к меньшим значениям целевой функции по сравнению с алгоритмом 3. Однако гарантирующий приемлемое качество решения критерий останова алгоритма 3 выполняется быстрее по сравнению с алгоритмом 5. Поэтому результаты для каждого из этих методов по-своему интересны.

5.3. Случай квазивыпуклых ограничений

В этом подразделе мы покажем работу алгоритмов 1 и 5 для задачи (1.3) с целевой функцией $f(x) = \|x\|$ и функциональными ограничениями вида

$$g_k(x) = \sqrt{\alpha_k \|x - a_k\|} - \beta_k, \quad k = 1, \dots, m, \tag{5.5}$$

где $\alpha_k > 0, \beta_k > 0$ и $a_k \in \mathbb{R}^n, \forall k = 1, \dots, m$. А также

$$g_k(x) = \begin{cases} \|x - a_k\| + (\rho - 1)r_k - \gamma_k, & \text{если } \|x - a_k\| \geq r_k, \\ \rho \|x - a_k\| - \gamma_k, & \text{в противном случае,} \end{cases} \tag{5.6}$$

где $a_k \in \mathbb{R}^n$ — некоторые фиксированные точки, $\rho > 1$ и $r_k > 0$ для всех $k = 1, \dots, m$. Функционалы (5.5) и (5.6) квазивыпуклы и липшицевы. Отметим, что $M_{g_k} = \rho$ для всех $k = 1, \dots, m$.

Выберем в качестве множества X евклидов шар в \mathbb{R}^n с центром в $\left(\frac{2}{\sqrt{n}}, \dots, \frac{2}{\sqrt{n}}\right)$ радиуса 2.

Алгоритмы 1 и 5 запускались с начальной точкой $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$ при $n = 1000, m = 100$ и различных значениях δ . Коэффициенты α_k и точки $a_k \in \mathbb{R}^n$ в (5.5) случайно генерируются с равномерным распределением на интервале $[0, 1)$, β_k также случайно генерируются с равномерным распределением на интервале $[2, 100)$, γ_k в (5.6) случайно генерируются с равномерным распределением на интервале $[2, 10)$, точки a_k случайно генерируются с равномерным распределением на интервале $[0, 1)$, а затем нормируются таким образом, что $1 \leq \|a_k\|_2 \leq 2$,

Т а б л и ц а 1

Результаты работы алгоритмов 1 и 5 с ограничениями (5.5)

$1/\delta$	Алгоритм 1			Алгоритм 5		
	\hat{f}	итерации	время, с	\hat{f}	итерации	время, с
2	1.62499999	14	0.046904	1.0	17	0.035938
4	1.14772727	44	0.067855	1.0	65	0.115693
8	1.09294872	181	0.295239	0.99999999	257	0.467748
16	1.04425335	728	1.278552	0.99999999	1025	1.866044
32	1.02380952	2964	5.080439	1.0	4097	7.861005
64	1.00874814	11685	18.629206	0.99999999	16385	32.899023

Т а б л и ц а 2

Результаты работы алгоритмов 1 и 5 с ограничениями (5.6)

$1/\delta$	Алгоритм 1			Алгоритм 5		
	\hat{f}	итерации	время, с	\hat{f}	итерации	время, с
2	1.55555555	9	0.025958	0.99999999	17	0.034944
4	1.21990537	34	0.046874	0.99999999	65	0.118692
8	1.06559401	132	0.174533	0.99999999	257	0.538587
16	1.02753418	520	0.628345	0.99999999	1025	2.037566
32	1.01326641	2064	2.397629	0.99999999	4097	7.839096
64	1.00533693	8228	9.700079	0.99999999	16385	30.710967

$\rho = 2$ и $r_k = 1$ для всех $k = 1, \dots, m$. По сути, рассматривается задача проектирования точки на некоторое подмножество X , которое задается рассматриваемыми ограничениями-неравенствами.

Результаты работы алгоритмов 1 и 5 представлены в табл. 1, 2. Эти результаты демонстрируют количество итераций до выполнения критериев остановки, время работы в секундах и значения целевой функции $f(\hat{x}) = \hat{f}$ в точке выхода \hat{x} каждого алгоритма.

Согласно данным табл. 1 и 2 гарантирующий достижение приемлемой точности решения критерий остановки алгоритма 1 работает быстрее аналогичного критерия для алгоритма 5. Но при этом алгоритм 5 может выдавать на практике несколько лучшие значения целевой функции.

Заключение

В статье получен результат о сходимости субградиентного метода с адаптивным правилом остановки для задачи минимизации выпуклой целевой функции при квазивыпуклых липшицевых ограничениях. Исследована задача математического программирования с условием, аналогичным классическому острому минимуму (“условный” острый минимум (1.4)). В работе рассмотрены следующие классы задач.

- Целевая функция выпукла, ограничения квазивыпуклы. Этот случай интересен тем, что выпуклость позволяет предложить адаптивные критерии остановки алгоритма, причем метод можно использовать без знания константы Липшица целевой функции.
- Целевая функция квазивыпукла, ограничения выпуклы. Для этого случая построены рестарты субградиентного метода по параметру острого минимума, гарантирующие линейную скорость сходимости.
- Целевая функция квазивыпукла, ограничения квазивыпуклы. Наиболее общая постановка задачи, метод требует знания константы Липшица целевой функции и ограничений.

Для каждого из указанных выше классов задач, обладающих “условным” острым минимумом (1.4), построены рестартованные версии субградиентных методов, показана сходимость со скоростью геометрической прогрессии и получены оценки достаточного для достижения приемлемой точности количества итераций. С целью сравнительного анализа предложенных методов проведены вычислительные эксперименты для нескольких задач с выпуклыми и квазивыпуклыми ограничениями. Проведенные эксперименты показали, что в случае известного параметра “условного” острого минимума (1.4) рестартованные версии субградиентных методов (алгоритмы 2, 4 и 6) позволяют улучшить скоростные гарантии по сравнению с базовыми версиями, не использующими острый минимум (алгоритмы 1, 3 и 5 соответственно). Однако проблема (известная и ранее в выпуклом случае) в том, что точно этот параметр оценить проблематично, а он существенно используется для организации рестартов (не только в теоретических оценках). В таком случае результаты о линейной скорости сходимости методов носят скорее теоретический интерес. Для задач оптимизации с обычным острым минимумом (1.1) известен альтернативный подход Б. Т. Поляка [12] с использованием f^* к выбору шага субградиентного метода, который позволяет реализовать субградиентный метод для задач с острым минимумом без знания его параметра (этот параметр возникает лишь в оценках скорости сходимости). Интересной на будущее представляется задача построения методов с шагами такого типа на рассмотренных в настоящей статье классах нелинейных задач с ограничениями. При этом ограничения задачи приводят к сложности оценки f^* для реальных задач, а также, по-видимому, к необходимости пересмотреть подход к обобщению острого минимума (1.4) и к выбору шага.

СПИСОК ЛИТЕРАТУРЫ

1. Поляк Б.Т. Один общий метод решения экстремальных задач // Докл. АН СССР. 1967. Т. 174, № 1. С. 33–36.
2. Huang Y., Lin Q. Single-loop switching subgradient methods for non-smooth weakly convex optimization with non-smooth convex constraints [e-resource]. 2023. 49 p. URL: <https://arxiv.org/pdf/2301.13314.pdf>.
3. Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A. Mirror descent and convex optimization problems with non-smooth inequality constraints // Large-Scale and Distributed Optimization / eds. Pontus Giselsson and Anders Rantzer. 2018. P. 181–213. (Ser. Lecture Notes in Math. 2018; vol. 2227.) doi: 10.1007/978-3-319-97478-1_8
4. Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization // Ecole polytechnique de Louvain. Université catholique de Louvain, 2017. URL: <https://dial.uclouvain.be/memoire/ucl/object/thesis:12934>.
5. Nesterov Yu. Subgradient methods for huge-scale optimization problems // Math. Prog. 2015. Vol. 146, no. 1–2. P. 275–297. doi: 10.1007/s10107-013-0686-4
6. Стонякин Ф.С., Алкуса М.С., Степанов А.Н., Баринов М.А. Адаптивные алгоритмы зеркального спуска в задачах выпуклого программирования с липшицевыми ограничениями // Тр. Ин-та математики и механики УрО РАН. 2018. Т. 24, № 2. С. 266–279. doi: 10.21538/0134-4889-2018-24-2-266-279
7. Stonyakin F.S., Stepanov A.N., Gasnikov A.V., Titov A.A. Mirror descent for constrained optimization problems with large subgradient values of functional constraints // Computer Research and Modeling. 2020. Vol. 12, no. 2. P. 301–317. doi: 10.20537/2076-7633-2020-12-2-301-317
8. 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 [e-resource]. 2022. 29 p. URL: <https://arxiv.org/pdf/2010.15267.pdf>.
9. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.
10. Нестеров Ю.Е. Эффективные методы нелинейного программирования. М.: Радио и связь, 1989. 301 с.
11. Кларк Ф. Оптимизация и негладкий анализ: пер. с англ. / под ред. В. И. Благодатских. М.: Наука, 1988. 280 с. ISBN 5-02-013781-2.
12. Поляк Б.Т. Минимизация негладких функционалов // Журн. вычисл. математики и мат. физ. 1969. Т. 9, № 3. С. 509–521.

13. **Boyd S.P., Vandenberghe L.** Convex optimization. Cambridge: Cambridge Univ. Press., 2004. 716 p. ISBN: 9780521833783

Поступила 14.05.2023

После доработки 4.07.2023

Принята к публикации 10.07.2023

Аблаев Сейдамет Серверович
аспирант

Крымский федеральный университет им. В. И. Вернадского
г. Симферополь;

Московский физико-технический институт (национальный исследовательский университет)
г. Долгопрудный
e-mail: seydamet.ablaev@yandex.ru

Стонякин Федор Сергеевич

д-р физ.-мат. наук, ведущий науч. сотрудник

Московский физико-технический институт (национальный исследовательский университет)
г. Долгопрудный;
профессор
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: fedyor@mail.ru

Алкуса Мохаммад

канд. физ.-мат. наук

Московский физико-технический институт (национальный исследовательский университет)
г. Долгопрудный
e-mail: mohammad.alkousa@phystech.edu

Гасников Александр Владимирович

д-р физ.-мат. наук, профессор

Московский физико-технический институт (национальный исследовательский университет)
г. Долгопрудный;
Институт проблем передачи информации им. А. А. Харкевича
г. Москва;
Кавказский математический центр Адыгейского государственного университета
г. Майкоп
e-mail: gasnikov@yandex.ru

REFERENCES

1. Polyak B.T. A general method of solving extremum problems. *Sov. Math. Dokl.*, 1967, vol. 8, no. 3, pp. 593–597.
2. Huang Y., Lin Q. *Single-loop switching subgradient methods for non-smooth weakly convex optimization with non-smooth convex constraints*. 2023, 49 p. Available at <https://arxiv.org/pdf/2301.13314.pdf>.
3. Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A. Mirror descent and convex optimization problems with non-smooth inequality constraints. In: *Large-Scale and Distributed Optimization*, eds. Pontus Giselsson and Anders Rantzer, 2018, Ser. Lecture Notes in Math., vol. 2227, pp. 181–213. doi: 10.1007/978-3-319-97478-1_8
4. Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. In: *Ecole polytechnique de Louvain, Université catholique de Louvain*, 2017. Available at <https://dial.uclouvain.be/memoire/ucl/object/thesis:12934>.
5. Nesterov Yu. Subgradient methods for huge-scale optimization problems. *Math. Prog.*, 2015, vol. 146, no. 1–2, pp. 275–297. doi: /10.1007/s10107-013-0686-4

6. Stonyakin F.S., Alkousa M.S., Stepanov A.N., Barinov M.A. Adaptive mirror descent algorithms in convex programming problems with Lipschitz constraints. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2018, vol. 24, no. 2, pp. 266–279 (in Russian). doi: 10.21538/0134-4889-2018-24-2-266-279
7. Stonyakin F.S., Stepanov A.N., Gasnikov A.V., Titov A.A. Mirror descent for constrained optimization problems with large subgradient values of functional constraints. *Computer Research and Modeling*, 2020, vol. 12, no. 2, pp. 301–317. doi: 10.20537/2076-7633-2020-12-2-301-317
8. Lin Q., Ma R., Nadarajah S., Soheili N. *A Parameter-free and projection-free restarting level set method for adaptive constrained conver optimization under the error bound condition*, 2022, 29 p. Available at <https://arxiv.org/pdf/2010.15267.pdf>.
9. Polyak B.T. *Vvedenie v optimizatsiyu* [Introduction to Optimization]. Moscow: Nauka Publ., 1983, 384 p.
10. Nesterov Yu. *Effektivnye metody nelinejnogo programmirovaniya* [Effective methods of nonlinear programming]. Moscow: Radio and Communication Publ., 1989, 301 p.
11. Clarke F. *Optimization and nonsmooth analysis*. NY: Wiley-Interscience, 1983, 308 p. Translated to Russian under the title *Optimizatsiya i negladkii analiz*, Moscow: Nauka Publ, 1988. 280 p.
12. Polyak B.T. Minimization of nonsmooth functionals. *Comp. Math. and Math. Phys.*, 1969, vol. 9, no. 3, pp. 509–521 (in Russian).
13. Boyd S., Vandenberghe L. *Convex optimization*. Cambridge: Cambridge Univ. Press, 2004, 716 p. ISBN: 9780521833783

Received May 14, 2023

Revised July 4, 2023

Accepted July 10, 2023

Funding Agency: The research of F.S.Stonyakin in Sections 2 and 5 was supported by the Strategic Academic Leadership Program “Priority 2030” (agreement no. 075-02-2021-1316 of September 30, 2021). The research of A. V. Gasnikov in Section 4 was carried out under a state task of the Ministry of Science and Higher Education of the Russian Federation (project no. 0714-2020-0005).

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

Fedor Sergeevich Stonyakin, Dr. Phys.-Math. Sci., Prof., Moscow Institute of Physics and Technologies, Dolgoprudny, Moscow Region, 141701 Russia; V.I.Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: fedyor@mail.ru .

Mohammad S. Alkousa, Cand. Phys.-Math. Sci., Moscow Institute of Physics and Technology (National Research University); Dolgoprudny, Moscow Region, 141701 Russia, e-mail: mohammad.alkousa@phystech.edu .

Aleksandr Vladimirovich Gasnikov, Dr. Phys.-Math. Sci., Prof., Moscow Institute of Physics and Technology (National Research University); Dolgoprudny, Moscow Region, 141701 Russia; Caucasus Mathematical Center, Adyghe State University, Maikop, 385000 Russia; Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, 127051 Russia, e-mail: gasnikov@yandex.ru .

Cite this article as: S. S. Ablaeu, F. S. Stonyakin, M. S. Alkousa, A. V. Gasnikov. Adaptive subgradient methods for mathematical programming problems with quasi-convex functions. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2023, vol. 29, no. 3, pp. 7–25.