

УДК: 519.85

Субградиентные методы для слабо выпуклых и относительно слабо выпуклых задач с острым минимумом

Ф. С. Стонякин^{1,2,a}, С. С. Аблаев^{1,2,b}, И. В. Баран^{2,c}, М. С. Алкуса^{1,3,d}

¹Московский физико-технический институт (национальный исследовательский университет),
Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., 9

²Крымский федеральный университет им. В. И. Вернадского,
Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, 4

³Национальный исследовательский университет «Высшая школа экономики»,
Россия, 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: ^a fedyor@mail.ru, ^b seydamet.ablaev@yandex.ru, ^c matemain@mail.ru,
^d mohammad.alkousa@phystech.edu

Получено 19.02.2023.

Принято к публикации 23.02.2023.

Работа посвящена исследованию субградиентных методов с различными вариациями шага Б. Т. Поляка на классах задач минимизации слабо выпуклых и относительно слабо выпуклых функций, обладающих соответствующим аналогом острого минимума. Оказывается, что при некоторых предположениях о начальной точке такой подход может давать возможность обосновать сходимость субградиентного метода со скоростью геометрической прогрессии. Для субградиентного метода с шагом Б. Т. Поляка доказана уточненная оценка скорости сходимости для задач минимизации слабо выпуклых функций с острым минимумом. Особенность этой оценки — дополнительный учет сокращения расстояния от текущей точки метода до множества решений по мере роста количества итераций. Представлены результаты численных экспериментов для задачи восстановления фазы (которая слабо выпукла и имеет острый минимум), демонстрирующие эффективность предложенного подхода к оценке скорости сходимости по сравнению с известным ранее результатом. Далее, предложена вариация субградиентного метода с переключениями по продуктивным и непродуктивным шагам для слабо выпуклых задач с ограничениями-неравенствами и получен некоторый аналог результата о сходимости со скоростью геометрической прогрессии. Для субградиентного метода с соответствующей вариацией шага Б. Т. Поляка на классе относительно липшицевых и относительно слабо выпуклых функций с относительным аналогом острого минимума получены условия, которые гарантируют сходимость такого субградиентного метода со скоростью геометрической прогрессии. Наконец, получен теоретический результат, описывающий влияние погрешности доступной субградиентному методу информации о (суб)градиенте и целевой функции на оценку качества выдаваемого приближенного решения. Доказано, что при достаточно малой погрешности $\delta > 0$ можно гарантировать достижение точности решения, сопоставимой с δ .

Ключевые слова: субградиентный метод, острый минимум, липшицева функция, относительная липшицевость, относительный острый минимум, задача восстановления фазы

Работа выполнена при поддержке гранта Российского научного фонда и города Москвы № 22-21-20065 (<https://rscf.ru/project/22-21-20065/>).

UDC: 519.85

Subgradient methods for weakly convex and relatively weakly convex problems with a sharp minimum

F. S. Stonyakin^{1,2,a}, S. S. Ablav^{1,2,b}, I. V. Baran^{2,c}, M. S. Alkousa^{1,3,d}

¹Moscow Institute of Physics and Technology,
9 Institutskiy per., Dolgoprudny, Moscow region, 141701, Russia

²V. I. Vernadsky Crimean Federal University,
4 Academician Vernadsky Avenue, Simferopol, Republic of Crimea, 295007, Russia

³HSE University,
20 Myasnitskaya st., Moscow, 101000, Russia

E-mail: ^a fedyor@mail.ru, ^b seydamet.ablav@yandex.ru, ^c matemain@mail.ru,
^d mohammad.alkousa@phystech.edu

Received 19.02.2023.

Accepted for publication 23.02.2023.

The work is devoted to the study of subgradient methods with different variations of the Polyak stepsize for minimization functions from the class of weakly convex and relatively weakly convex functions that have the corresponding analogue of a sharp minimum. It turns out that, under certain assumptions about the starting point, such an approach can make it possible to justify the convergence of the subgradient method with the speed of a geometric progression. For the subgradient method with the Polyak stepsize, a refined estimate for the rate of convergence is proved for minimization problems for weakly convex functions with a sharp minimum. The feature of this estimate is an additional consideration of the decrease of the distance from the current point of the method to the set of solutions with the increase in the number of iterations. The results of numerical experiments for the phase reconstruction problem (which is weakly convex and has a sharp minimum) are presented, demonstrating the effectiveness of the proposed approach to estimating the rate of convergence compared to the known one. Next, we propose a variation of the subgradient method with switching over productive and non-productive steps for weakly convex problems with inequality constraints and obtain the corresponding analog of the result on convergence with the rate of geometric progression. For the subgradient method with the corresponding variation of the Polyak stepsize on the class of relatively Lipschitz and relatively weakly convex functions with a relative analogue of a sharp minimum, it was obtained conditions that guarantee the convergence of such a subgradient method at the rate of a geometric progression. Finally, a theoretical result is obtained that describes the influence of the error of the information about the (sub)gradient available by the subgradient method and the objective function on the estimation of the quality of the obtained approximate solution. It is proved that for a sufficiently small error $\delta > 0$, one can guarantee that the accuracy of the solution is comparable to δ .

Keywords: subgradient method, sharp minimum, Lipschitz-continuous function, relatively Lipschitz-continuous function, relatively sharp minimum, phase retrieval problem

Citation: *Computer Research and Modeling*, 2023, vol. 15, no. 2, pp. 393–412 (Russian).

This work was partially supported by Russian Science Foundation and Moscow city, project 22-21-20065 (<https://rscf.ru/project/22-21-20065/>).

1. Введение

Известно, что многие возникающие в приложениях невыпуклые функции слабо выпуклы. Это касается разных типов задач нелинейной регрессии, восстановления фазы, восстановления матрицы (matrix recovery) и др. Класс слабо выпуклых оптимизационных задач интересен, ему посвящены всё новые работы [Huang, Lin, 2023; Дудов, Осипцев, 2021; Li et al., 2020; Davis, Drusvyatskiy, Paquette, 2020; Davis et al., 2018; Davis, Drusvyatskiy, Kellie, 2018], но без дополнительных предположений скоростные гарантии достаточно пессимистичны. Однако в случае острого минимума недавно получен результат о сходимости субградиентного метода со скоростью геометрической прогрессии при условии достаточной близости начальной точки к точному решению [Davis et al., 2018].

Настоящая работа посвящена развитию результата [Davis et al., 2018] в плане исследования субградиентных методов с различными вариациями шага Б. Т. Поляка на классе слабо выпуклых и относительно слабо выпуклых функций [Davis, Drusvyatskiy, Kellie, 2018], обладающих острым минимумом или его аналогом — относительным острым минимумом. Будем рассматривать задачи минимизации вида

$$\min_{x \in Q} f(x), \quad (1)$$

где Q — выпуклое замкнутое подмножество \mathbb{R}^n . Напомним, что функция $f(x)$ имеет острый минимум, если для всех $x \in Q$ верно неравенство

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

где $\alpha > 0$, $f^* = f(x_*)$ — минимальное значение функции для задачи (1), норма $\|\cdot\|_2$ — евклидова. Сделаем предположение о компактности множества искомых точных решений X_* задачи (1). Напомним, что функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (аналогично и для функций $f: Q \rightarrow \mathbb{R}$) называется μ -слабо выпуклой ($\mu \geq 0$), если функция $x \rightarrow f(x) + \frac{\mu}{2}\|x\|^2$ выпукла. Для недифференцируемых μ -слабо выпуклых функций f под субдифференциалом $\partial f(x)$ в точке x можно понимать (см. [Davis et al., 2018] и цитируемую там литературу) множество всех векторов $v \in \mathbb{R}^n$, удовлетворяющих неравенству

$$f(y) \geq f(x) + \langle v, y - x \rangle + o(\|y - x\|_2) \quad \text{при } y \rightarrow x. \quad (2)$$

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

$$f(y) \geq f(x) + \langle v, y - x \rangle - \frac{\mu}{2}\|y - x\|_2^2 \quad \forall x, y \in \mathbb{R}^n, v \in \partial f(x). \quad (3)$$

Можно проверить, что слабо выпуклые функции локально липшицевы, и поэтому в качестве субградиентов можно использовать произвольный вектор из субдифференциала Кларка (см., например, [Дудов, Осипцев, 2021]). Мы будем далее следовать этому подходу. Известно немало примеров прикладных невыпуклых, но слабо выпуклых задач. Так, слабо выпуклыми будут задачи вида (см., например, [Davis et al., 2018])

$$\min_x f(x) := h(c(x)),$$

где $h: \mathbb{R}^m \rightarrow \mathbb{R}$ выпукло и M -липшицево, а $c: \mathbb{R}^n \rightarrow \mathbb{R}^m$ является C^1 -гладким отображением с β -липшицевой матрицей Якоби. Нетрудно проверить, что при указанных допущениях f $M\beta$ -слабо выпукла. В частности, в указанный класс задач входят задачи нелинейной регрессии с $h(x) = \|x\|_1$, где $\|\cdot\|_1$ — 1-норма.

Работа состоит из введения, заключения и четырех основных параграфов.

Во втором параграфе получено уточнение оценки скорости сходимости субградиентного метода с шагом Б. Т. Поляка на классе μ -слабо выпуклых функций с острым минимумом [Davis et al., 2018] (теорема 1). Такое уточнение позволяет учесть приближение траектории метода ко множеству точных решений задачи по мере роста количества итераций. Более того, получено обобщение данного результата для субградиентного метода с переключениями по продуктивным и непродуктивным шагам (теорема 2) на задачи с ограничениями-неравенствами для соответствующей ε -вариации острого минимума (определение 1).

В третьем параграфе исследован субградиентный метод для относительно липшицевых и относительно слабо выпуклых функций, обладающих аналогом острого минимума — относительным острым минимумом [Макаренко, 2022]:

$$f(x) - f^* \geq \alpha \sqrt{\min_{x_* \in X^*} V(x_*, x)}.$$

Здесь $V(y, x)$ — дивергенция (или расхождение) Брэгмана, которая определяется следующим образом:

$$V(y, x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle,$$

где $d(x)$ — некоторая выпуклая дифференцируемая прокс-функция на множестве Q . Доказано, что при некоторых ограничениях на начальную точку скорость сходимости метода будет линейной. Отметим, что класс относительно слабо выпуклых задач введен в [Davis, Drusvyatskiy, Kellie, 2018]. Но в работе [Davis, Drusvyatskiy, Kellie, 2018] получены только сублинейные оценки скорости сходимости, никаких предположений типа острого минимума и результатов о линейной скорости сходимости не рассматривалось. В нашей работе доказан результат о сходимости со скоростью геометрической прогрессии (теорема 3).

В четвертом параграфе изучен вопрос о влиянии неточной информации о целевой функции и ее (суб)градиенте на вычислительные гарантии метода, рассмотренного в третьем параграфе настоящей работы для относительно слабо выпуклых задач. Доказано, что при достаточно малой погрешности $\delta > 0$ можно гарантировать достижение точности решения, сопоставимой с δ (теорема 4).

Наконец, в пятом параграфе приведены результаты проведенных численных экспериментов для иллюстрации полученных результатов. Первый пример выпуклой задачи с острым минимумом показывает возможность применения разработанной методики для задач с неточной информацией о целевой функции и ее субградиенте в текущей точке. Второй пример связан с известной задачей восстановления фазы, которая слабо выпукла и имеет острый минимум. Для этого примера продемонстрирован смысл предложенного в § 2 подхода к уточнению оценки скорости сходимости субградиентного метода. Полученные результаты расчетов показали неплохую эффективность разработанной методики для рассматриваемого класса задач.

В заключении даны некоторые комментарии по поводу полученных результатов, в том числе и о возможности использования разных типов субдифференциалов для слабо выпуклых функций.

Здесь и всюду далее под обозначением $\text{Proj}_Q(x)$ будем понимать евклидову проекцию точки x на множество Q .

2. Уточненная оценка скорости сходимости субградиентного метода для слабо выпуклых задач с острым минимумом и обобщение на задачи с ограничениями-неравенствами

В данном пункте мы считаем, что если функция $f(x)$ дифференцируема в точке x , то $\nabla f(x)$ — это ее градиент. Если функция $f(x)$ слабо выпукла и недифференцируема в точке x , то под субградиентом $\nabla f(x)$ будем понимать произвольный элемент субдифференциала

Кларка функции f в точке x . Начнем со следующего уточнения результата [Davis et al., 2018, Theorem 3.1] о линейной скорости сходимости субградиентного метода для слабо выпуклых задач с острым минимумом, которое предполагает дополнительный учет сокращения расстояния от текущей точки метода до множества решений с ростом количества итераций.

Теорема 1. Пусть f μ -слабо выпукла и имеет α -острый минимум ($\alpha, \mu > 0$), а точка $x_0: \min_{x_* \in X_*} \|x_0 - x_*\|_2 \leq \frac{\alpha\gamma}{\mu}$ для некоторого фиксированного $\gamma \in (0; 1)$. Тогда для метода

$$x_{k+1} := Proj_Q(x_k - h_k \nabla f(x_k)), \quad k = 0, 1, 2, \dots, \tag{4}$$

где

$$h_k := \frac{f(x_k) - f^*}{\|\nabla f(x_k)\|_2^2} \quad (\text{при } \nabla f(x_k) \neq 0), \tag{5}$$

верно неравенство

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq \prod_{i=0}^k \left(1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla f(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2, \tag{6}$$

где последовательность $\{\gamma_i\}_{i=0}^k$ такова, что $\gamma_0 := \gamma \in (0; 1)$, а

$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla f(x_i)\|_2^2}} \tag{7}$$

при всяком $i = 0, 1, 2, \dots, k$. Ясно, что $\gamma_0 > \gamma_1 > \gamma_2 > \dots$

Доказательство проведем методом математической индукции по i . Базис индукции ($i = 0$) очевиден ввиду $\gamma_0 = \gamma$ и $\min_{x_* \in X_*} \|x_0 - x_*\|_2 \leq \frac{\alpha\gamma}{\mu}$. Если допустить, что

$$\min_{x_* \in X_*} \|x_{i+1} - x_*\|_2^2 \leq \left(1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla f(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_i - x_*\|_2^2 \tag{8}$$

при $i = k - 1$, то применение оценки (6) приведет к (8) при $i = k$. Покажем это.

Для всякого $k \geq 0$ и ближайшего к x_k точного решения $x_* \in X_*$ с учетом μ -слабой выпуклости f имеем следующие неравенства (см., например, [Поляк, 1969], соотношения (3) из доказательства теоремы 1):

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

Используем то, что $\|x_k - x_*\|_2 \leq \frac{\alpha\gamma_k}{\mu}$ ввиду (8) при некотором $\gamma_k \in (0; 1)$. Следовательно,

$$\|x_{k+1} - x_*\|_2^2 \leq \|x_k - x_*\|_2^2 + \frac{f(x_k) - f^*}{\|\nabla f(x_k)\|_2^2} (\alpha\gamma_k - \alpha) \|x_k - x_*\|_2 \leq \left(1 - \frac{(1 - \gamma_k)\alpha^2}{\|\nabla f(x_k)\|_2^2}\right) \|x_k - x_*\|_2^2.$$

Поэтому неравенство (6) вытекает из того, что $\forall i \geq 0$ верно (8). □

ЗАМЕЧАНИЕ 1. Если f удовлетворяет условию Липшица ($\|\nabla f(x)\|_2 \leq M$ при некотором $M > 0$) на множестве $B_2(X_*, \frac{\gamma\alpha}{\mu})$, где $B_2(X_*, \frac{\gamma\alpha}{\mu})$ есть $\frac{\gamma\alpha}{\mu}$ -окрестность X_* , то $\forall k \geq 0$

$$\gamma_{k+1}^2 \leq \left(1 - \frac{(1-\gamma_k)\alpha^2}{M^2}\right) \gamma_k^2 \leq \left(1 - \frac{(1-\gamma)\alpha^2}{M^2}\right) \gamma_k^2 \leq \left(1 - \frac{(1-\gamma)\alpha^2}{M^2}\right) \gamma_{k-1}^2 \leq \dots \leq \left(1 - \frac{(1-\gamma)\alpha^2}{M^2}\right)^{k+1} \gamma^2.$$

Поэтому γ_k может оказаться существенно меньшим по сравнению с γ , используемым в оценке [Davis et al., 2018, Theorem 3.1]. Далее это наглядно показано для одного из примеров в пятом параграфе настоящей работы.

ЗАМЕЧАНИЕ 2. По аналогии с рассуждениями [Davis et al., 2018, Lemma 2.1] можно проверить, что траектория субградиентного метода (4) из теоремы 1 при условии отмеченного выбора начальной точки лежит в $\frac{\gamma\alpha}{\mu}$ -окрестности множества точных решений X_* и не содержит стационарных точек, отличных от элементов X_* .

Сделаем комментарий по поводу применимости на практике доказанной оценки (6). Главной проблемой в этом плане видится необходимость знать параметр острого минимума α , который сложно оценить во многих прикладных задачах (для параметра слабой выпуклости функции ситуация существенно лучше). Контроль качества решения исключительно по значению функции тоже может не вполне удовлетворить, поскольку f^* и так часто известно и ставится цель найти хорошее приближение итеративной последовательности $\{x_k\}_{k=1}^\infty$ к X_* по аргументу. Помочь на практике может следующее наблюдение.

ЗАМЕЧАНИЕ 3. Если для всякого x_k ($k \geq 0$) удастся подобрать такие $\alpha_k > 0$, что

$$f(x_k) - f^* \geq \alpha_k \min_{x_* \in X_*} \|x_k - x_*\|_2,$$

то неравенство (6) можно уточнить следующим образом:

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq \prod_{i=0}^k \left(1 - \frac{\alpha_i^2 (1-\gamma_i)}{\|\nabla f(x_i)\|_2^2}\right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2. \quad (9)$$

Оценка (9) потенциально может даже оказаться лучшей по сравнению с (6), если для некоторых $i \in \{0, 1, \dots, k\}$ $\alpha_i > \alpha$ (α – глобальный параметр острого минимума). Например, при $f(x) = \|x\|_2^2 + \|x\|_2$ ясно, что при $\|x\|_2 \gg 1$ $\frac{f(x)}{\|x\|_2} = \|x\|_2 + 1 \gg 1$ (здесь 1 – параметр острого минимума). Отметим, что в выпуклом случае оценки (6) и (9) верны при $\gamma = 0$ без ограничений на выбор начальной точки x_0 .

Но помощь замена α на α_i ($i \leq k$) может и в ситуации, когда известно, что множество X_* точных решений содержит 0 (в случае другой точки вместо 0 можно применить сдвиг). Это может быть верно для отдельных геометрических задач (проектирование на множество, нахождение общей точки множеств). Но принадлежность 0 множеству X_* позволяет заметить, что

$$\min_{x_* \in X_*} \|x - x_*\|_2 \leq \|x\|_2 \quad \forall x \in Q.$$

Тогда

$$\frac{f(x) - f^*}{\|x\|_2} \leq \frac{f(x) - f^*}{\min_{x_* \in X_*} \|x - x_*\|_2},$$

и поэтому $\forall x_i$ ($i \geq 0$) можно выбрать

$$\alpha_i := \frac{f(x_i) - f^*}{\|x_i\|_2}.$$

Ясно, что $f(x_i) - f^* = \alpha_i \|x_i\|_2 \geq \alpha_i \min_{x_* \in X_*} \|x_i - x_*\|_2$, и к таким параметрам $\{\alpha_i\}_{i=0}^k$ можно применить оценку (9). Если X_* — многогранник с малыми расстояниями между вершинами, то вместо нулевой точки можно выбрать какую-то вершину (хотя совсем не очевидно, ближайшая ли это точка).

Полученный выше результат можно обобщить и на задачи с ограничениями-неравенствами. Похожая постановка рассматривалась недавно в [Huang, Lin, 2023], однако без предположений типа острого минимума и результатов о сходимости со скоростью геометрической прогрессии. Введем следующий аналог острого минимума для задач вида

$$\min_{g(x) \leq 0, x \in Q} f(x). \tag{10}$$

Определение 1. Будем говорить, что задача (10) обладает ε -острым минимумом ($\varepsilon > 0$), если для фиксированного $\alpha > 0$ верна альтернатива

- при $g(x) > \varepsilon$

$$g(x) \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \forall x \in Q;$$

- при $g(x) \leq \varepsilon$

$$f(x) - f^* \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \text{или} \quad f(x) - f^* \leq \varepsilon \quad \forall x \in Q.$$

Для задач вида (10) предложим следующий алгоритм 1, предполагая (суб)градиенты f и g отличными от 0 в соответствующих точках.

Algorithm 1. Схема с переключениями по продуктивным и непродуктивным шагам для задач с ограничениями-неравенствами

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

1: $I := \emptyset, J := \emptyset.$

2: **if** $g(x_k) \leq \varepsilon$ **then**

3:

$$x_{k+1} := \text{Proj}_Q \left(x_k - \frac{f(x_k) - f^*}{\|\nabla f(x_k)\|_2^2} \nabla f(x_k) \right), \quad \text{«продуктивные шаги»}.$$

4: $k \rightarrow I$

5: **else**

6: **if** $g(x_k) > \varepsilon, k \in J$ **then**

7:

$$x_{k+1} := \text{Proj}_Q \left(x_k - \frac{g(x_k)}{\|\nabla g(x_k)\|_2^2} \nabla g(x_k) \right), \quad \text{«непродуктивные шаги»}.$$

8: $k \rightarrow J$

9: **end if**

10: **end if**

Если f и g μ -слабо выпуклы, то

a) при $g(x_k) \leq \varepsilon$ либо $f(x_k) - f^* \leq \varepsilon$, либо $f(x_k) - f^* \geq \alpha \min_{x_* \in X_*} \|x_k - x_*\|_2$, и тогда

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

б) при $g(x_k) > \varepsilon$ верно $g(x_k) \geq \alpha \min_{x_* \in X_*} \|x_k - x_*\|_2$, и тогда, ввиду $g(x_*) \leq 0$,

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

Откуда

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

Таким образом, справедлива следующая

Теорема 2. Пусть f и g μ -слабо выпуклы и имеют ε -острый минимум ($\alpha > 0$, $\mu > 0$), а начальная точка x_0 такова, что $\min_{x_* \in X_*} \|x_0 - x_*\|_2 \leq \frac{\alpha \gamma}{\mu}$ для некоторого фиксированного $\gamma \in (0; 1)$. Тогда после k итераций алгоритма 1 справедлива следующая альтернатива: или достигнуто ε -точное решение задачи (10), т. е.

$$f(x_{k+1}) - f^* \leq \varepsilon, \quad g(x_{k+1}) \leq \varepsilon, \quad (11)$$

или верно следующее неравенство:

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leq \prod_{i \in I} \left(1 - \frac{(1 - \gamma_i)\alpha^2}{\|\nabla f(x_i)\|_2^2} \right) \times \prod_{i \in J} \left(1 - \frac{(1 - \gamma_i)\alpha^2}{\|\nabla g(x_i)\|_2^2} \right) \min_{x_* \in X_*} \|x_0 - x_*\|_2^2, \quad (12)$$

где последовательность $\{\gamma_i\}_{i=0}^k$ такова, что $\gamma_0 := \gamma \in (0; 1)$, а при $i = 0, 1, 2, \dots, k$

$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla f(x_i)\|_2^2}}$$

для всякого продуктивного шага ($i \in I$) и

$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{\alpha^2(1 - \gamma_i)}{\|\nabla g(x_i)\|_2^2}}$$

для всякого непродуктивного шага ($i \in J$).

ЗАМЕЧАНИЕ 4. По аналогии с рассуждениями [Davis et al., 2018, Lemma 2.1] можно проверить, что траектория субградиентного метода (4) из теоремы 1 при условии указанного выбора начальной

точки лежит в $\frac{\gamma\alpha}{\mu}$ -окрестности множества точных решений X_* и не содержит стационарных точек f и g , отличных от элементов X_* . Тем самым сделанное выше предположение о том, что субградиенты f и g ненулевые, вполне оправдано.

ЗАМЕЧАНИЕ 5. Можно заметить, что если функции f и g M -липшицевы ($M > 0$), то можно гарантировать достижение ε -точного решения задачи (10) не более чем за $O(\log \varepsilon^{-1})$ итераций алгоритма 1. Действительно, если евклидовы нормы (суб)градиентов f и g равномерно ограничены константой $M > 0$ в $\frac{\alpha\gamma}{\mu}$ -окрестности множества X_* , то неравенство (12) означает, что

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

Поэтому либо после итераций $O\left(\frac{M^2}{\alpha^2} \log \varepsilon^{-1}\right)$ получится ε -точное решение (11) задачи (10), либо можно будет гарантировать $\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2 \leq \frac{\varepsilon}{M}$. Последнее неравенство влечет, что x_{k+1} есть ε -точное решение задачи (10) в силу M -липшицевости f и g .

ЗАМЕЧАНИЕ 6. В заключение данного параграфа отметим, что все полученные результаты верны и на классе выпуклых задач. В такой ситуации можно полагать все множители γ_i равными 0 и отказаться от ограничений на выбор начальной точки $x_0 \in Q$.

3. Метод градиентного типа для минимизации относительно слабо выпуклого функционала с относительным острым минимумом

Теперь рассмотрим класс относительно μ -слабо выпуклых задач, который был введен в работе [Davis, Drusvyatskiy, Kellie, 2018]. В данном параграфе будет доказан результат о линейной скорости сходимости субградиентного метода с вариацией шага Б. Т. Поляка на этом классе задач при условии острого минимума целевой функции. Итак, рассмотрим задачу

$$\min_{x \in Q} f(x), \tag{13}$$

где f — дифференцируемая функция и

- относительно слабо выпуклая, т. е.

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

где $V(y, x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle$ — дивергенция Брэгмана, порожденная некоторой выпуклой дифференцируемой на множестве Q прокс-функцией d ;

- относительно липшицева [Nesterov, 2019; Lu, 2019], т. е.

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

В качестве аналога острого минимума используем следующее понятие, введенное в [Макаренко, 2022]. Отметим, что поскольку X_* компактно, то для всякого $x \in Q$ существует $\min_{x_* \in X_*} V(x_*, x)$.

Определение 2. Функция $f(x)$ удовлетворяет условию относительного острого минимума, если верно неравенство

$$f(x) - f^* \geq \alpha \sqrt{\min_{x_* \in X_*} V(x_*, x)} \quad \forall x \in Q, \tag{14}$$

где $f^* = f(x_*)$, $\alpha > 0$.

Для решения задачи (13) используем метод

$$x_{k+1} := \operatorname{argmin}_{x \in Q} \{h_k \langle \nabla f(x_k), x \rangle + V(x, x_k)\} \quad (15)$$

с шагом (ранее он уже приводился в [Аблаев и др., 2022])

$$h_k = \frac{f(x_k) - f^*}{M^2}. \quad (16)$$

Теорема 3. Пусть f — относительно слабо выпуклая и относительно липшицева функция, удовлетворяющая условию относительного острого минимума, причем для начальной точки x_0 справедливо неравенство

$$\sqrt{\min_{x_* \in X^*} V(x_*, x_0)} \leq \frac{\gamma \alpha}{\mu}, \quad \gamma \in \left(0, \frac{1}{2}\right). \quad (17)$$

Тогда для метода (15) с шагом (16) справедлива оценка

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2}\right) \min_{x_* \in X^*} V(x_*, x_k). \quad (18)$$

Доказательство. Согласно лемме 1 из [Стонякин и др., 2022] для всякого $x \in Q$ имеем

$$\begin{aligned} h_k \langle \nabla f(x_k), x_k - x \rangle &\leq \frac{h_k^2 M^2}{2} + V(x, x_k) - V(x, x_{k+1}), \\ V(x, x_{k+1}) &\leq \frac{h_k^2 M^2}{2} + V(x, x_k) - h_k \langle \nabla f(x_k), x_k - x \rangle \leq \end{aligned}$$

ввиду относительной слабой выпуклости

$$\leq \frac{h_k^2 M^2}{2} + V(x, x_k) + h_k (f(x) - f(x_k) + \mu V(x, x_k))$$

при $x = x_*$ получим следующие равенства:

$$\begin{aligned} \min_{x_* \in X^*} V(x_*, x_{k+1}) &\leq \frac{h_k^2 M^2}{2} + \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(f(x_*) - f(x_k) + \mu \min_{x_* \in X^*} V(x_*, x_k) \right) = \\ &= \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\frac{f(x_k) - f(x_*)}{2} + f(x_*) - f(x_k) + \mu \min_{x_* \in X^*} V(x_*, x_k) \right) = \\ &= \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\mu \min_{x_* \in X^*} V(x_*, x_k) - \frac{1}{2}(f(x_k) - f(x_*)) \right) \leq \end{aligned}$$

с учетом острого минимума получим

$$\leq \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\mu \min_{x_* \in X^*} V(x_*, x_k) - \frac{\alpha}{2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} \right) \leq$$

согласно условию (17)

$$\begin{aligned} &\leq \min_{x_* \in X^*} V(x_*, x_k) + h_k \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} \left(\gamma \alpha - \frac{\alpha}{2} \right) = \\ &= \min_{x_* \in X^*} V(x_*, x_k) + \alpha \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} h_k \frac{2\gamma - 1}{2} \leq \\ &\leq \min_{x_* \in X^*} V(x_*, x_k) - \frac{1-2\gamma}{2} \frac{\alpha^2}{M^2} \min_{x_* \in X^*} V(x_*, x_k) = \\ &= \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2}\right) \min_{x_* \in X^*} V(x_*, x_k). \end{aligned}$$

□

ЗАМЕЧАНИЕ 7. Очевидно, что доказанный результат означает сходимость метода со скоростью геометрической прогрессии. Более того, можно по аналогии с теоремой 1 получить оценку скорости сходимости, учитывающую приближение траектории метода ко множеству X_* . Если в теореме 3 дополнительно известно, что последовательность $\{\gamma_i\}_{i=0}^k$ такова, что $\gamma_0 := \gamma \in (0, \frac{1}{2})$,

$$\gamma_{i+1} := \gamma_i \sqrt{1 - \frac{(1 - 2\gamma_i)\alpha^2}{2M^2}},$$

то

$$\min_{x_* \in X_*} V(x_*, x_{k+1}) \leq \prod_{i=0}^k \left(1 - \frac{(1 - 2\gamma_i)\alpha^2}{2M^2}\right) \min_{x_* \in X_*} V(x_*, x_0), \tag{19}$$

причем

$$\gamma_{i+1}^2 \leq \left(1 - \frac{(1 - 2\gamma)\alpha^2}{2M^2}\right)^{i+1} \gamma^2.$$

По-видимому, посредством замены квадратов норм градиентов f и g на константы $M_f > 0$ и $M_g > 0$ можно получить и аналог теоремы 2 для задач минимизации относительно слабо выпуклого и M_f -относительно липшицева функционала f с относительно слабо выпуклым и относительно M_g -липшицевым функционалом ограничения g .

ЗАМЕЧАНИЕ 8. Отметим, что в условиях теоремы 3 нулевой градиент возможен только в точке $x_k \in X_*$. Дело в том, что задача минимизации относительно слабо выпуклой функции с относительным острым минимумом не имеет стационарных точек, удовлетворяющих условию

$$0 < \sqrt{\min_{x_* \in X_*} V(x_*, x)} < \frac{\alpha}{\mu}. \tag{20}$$

Действительно, зафиксируем стационарную точку $x \in Q \setminus X^*$ функции $f(x)$. Выберем $x_* \in X^*$ такое, что $V(x_*, x) = \mu \min_{x_* \in X_*} V(x_*, x)$. Тогда

$$\alpha \sqrt{\min_{x_* \in X_*} V(x_*, x)} \leq f(x) - f^* \leq \mu V(x_*, x) = \mu \min_{x_* \in X_*} V(x_*, x),$$

разделив на $\sqrt{\min_{x_* \in X_*} V(x_*, x)}$, получим

$$\sqrt{\min_{x_* \in X_*} V(x_*, x)} \geq \frac{\alpha}{\mu}.$$

Следовательно, в окрестности (20) нет стационарных точек, не принадлежащих множеству X^* .

Отметим, что сделанное выше предположение о дифференцируемости прокс-функции позволяет использовать вместо градиента произвольный элемент субдифференциала Фреше [Davis, Drusvyatskiy, Kellie, 2018, лемма 2.1].

4. Вариация субградиентного метода в случае доступности неточной информации о целевой функции или субградиенте

В предыдущих параграфах были использованы методы с шагом Б. Т. Поляка, работа которых предполагала знания точного значения минимума f^* , а также доступность значений субградиента и целевой функции во всякой текущей точке метода. Однако возможна ситуация, когда эти значения известны лишь с некоторой погрешностью. В данном параграфе будет исследована ситуация, в которой методу потенциально доступно лишь приближенное значение функции и/или субградиента в текущих точках. При этом предполагается доступным лишь некоторое его приближение $f_\delta(x)$, такое, что

$$f(x) - \delta \leq f_\delta(x) \leq f(x) \tag{21}$$

при некотором $\delta > 0$.

Рассмотрим метод (15) с шагом вида (см. [Аблаев и др., 2022])

$$h_k = \frac{f_\delta(x_k) - f^* - \delta}{M^2}. \quad (22)$$

Отметим, что если $h_k < 0$, то мы можем прекратить работу метода и полагать нужное качество решения уже достигнутым ($f_\delta(x_k) < f^* + \delta$). Соответственно, можно предложить следующий вариант понятия относительного острого минимума: функция $f(x)$ удовлетворяет условию относительного δ -острого минимума (в частном случае евклидовой прокс-структуры похожее понятие рассматривалось в [Аблаев и др., 2022]), если

$$f_\delta(x) - f^* \geq \alpha \sqrt{\min_{x_* \in X^*} V(x_*, x)} - \delta, \quad (23)$$

где $f^* = f_\delta(x_*)$, $\alpha > 0$, $\delta > 0$.

Далее, для относительно слабо выпуклых задач имеем при всяких x, x_k

$$\langle \nabla f(x_k), x_k - x \rangle \geq f(x_k) - f(x) - \mu V(x, x_k)$$

и предположим, что нам доступен неточный субградиент $\tilde{\nabla} f(x)$ и $f_\delta(x)$ такое, что $f(x) \geq f_\delta(x) - \delta$. Тогда верно неравенство

$$\langle \tilde{\nabla} f(x_k), x_k - x \rangle \geq f_\delta(x_k) - f(x) - \mu V(x, x_k) - \delta, \quad (24)$$

которое мы условимся называть δ -аналогом относительной μ -слабой выпуклости для функции $f(x)$.

Справедлив следующий аналог теоремы 3.

Теорема 4. Пусть f — относительно липшицева функция, удовлетворяющая условию относительного δ -острого минимума (23) и условию (24), и для начальной точки x_0 справедливо неравенство

$$\sqrt{\min_{x_* \in X^*} V(x_*, x_0)} \leq \frac{\gamma\alpha}{\mu}, \quad \gamma \in \left(0, \frac{1}{2}\right),$$

а также $\delta \leq \frac{(1-2\gamma)\alpha^2}{2\mu}$. Тогда для метода (15) с шагом (22) верна альтернатива:

1) либо

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2}\right)^{k+1} \min_{x_* \in X^*} V(x_*, x_0) + C(\delta) \frac{2M^2}{(1-2\gamma)\alpha^2},$$

где $C(\delta) = \frac{\alpha^2\delta\gamma(1-\gamma)}{\mu M^2} + \frac{\sqrt{2}\alpha\delta\gamma}{\mu M} - \frac{\delta^2}{M^2}$, при этом $C(\delta) = O(\delta)$;

2) либо

$$\min_{x_* \in X^*} V(x_*, x_k) < \left(\frac{-2\alpha\delta(1-2\gamma) - 2\sqrt{2}\delta M}{2\delta\mu - (1-2\gamma)\alpha^2}\right)^2. \quad (25)$$

Доказательство. Согласно лемме 1 из [Стонякин и др., 2022] имеем

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \frac{h_k^2 M^2}{2} + \min_{x_* \in X^*} V(x_*, x_k) - h_k \langle \tilde{\nabla} f(x_k), x_k - x_* \rangle \leq$$

воспользуемся условием (24)

$$\begin{aligned} &\leq \frac{h_k^2 M^2}{2} + \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(f(x_*) - f_\delta(x_k) + \mu \min_{x_* \in X^*} V(x_*, x_k) + \delta \right) = \\ &= \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\frac{f_\delta(x_k) - f(x_*) - \delta}{2} + f(x_*) - f_\delta(x_k) + \mu \min_{x_* \in X^*} V(x_*, x_k) + \delta \right) = \\ &= \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\mu \min_{x_* \in X^*} V(x_*, x_k) - \frac{1}{2} (f_\delta(x_k) - f(x_*) - \delta) \right) \leq \end{aligned}$$

согласно условию (23)

$$\begin{aligned} &\leq \min_{x_* \in X^*} V(x_*, x_k) + h_k \left(\mu \min_{x_* \in X^*} V(x_*, x_k) - \frac{\alpha}{2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + \delta \right) \leq \\ &\leq \min_{x_* \in X^*} V(x_*, x_k) - h_k \frac{\alpha(1-2\gamma)}{2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + h_k \delta \leq \end{aligned}$$

с учетом того, что $h_k \geq \frac{\alpha \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} - 2\delta}{M^2}$

$$\begin{aligned} &\leq \min_{x_* \in X^*} V(x_*, x_k) - \frac{\alpha(1-2\gamma)}{2} \frac{\alpha \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} - 2\delta}{M^2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + h_k \delta = \\ &= \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + \frac{\alpha\delta(1-2\gamma)}{M^2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + h_k \delta. \end{aligned}$$

Учтем, что

$$\begin{aligned} h_k = \frac{f_\delta(x_k) - f^* - \delta}{M^2} &\leq \frac{f(x_k) - f^* - \delta}{M^2} \leq \frac{\langle \nabla f(x_k), x_k - x_* \rangle + \mu \min_{x_* \in X^*} V(x_*, x_k) - \delta}{M^2} \leq \\ &\leq \frac{M \sqrt{2 \min_{x_* \in X^*} V(x_*, x_k)} + \mu \min_{x_* \in X^*} V(x_*, x_k) - \delta}{M^2}. \end{aligned}$$

Итак, имеем

$$\begin{aligned} \min_{x_* \in X^*} V(x_*, x_{k+1}) &\leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + \frac{\alpha\delta(1-2\gamma)}{M^2} \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + \\ &+ \frac{M\delta \sqrt{2 \min_{x_* \in X^*} V(x_*, x_k)} + \mu\delta \min_{x_* \in X^*} V(x_*, x_k) - \delta^2}{M^2}. \end{aligned}$$

Выясним, при каких условиях можно гарантировать, что

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \min_{x_* \in X^*} V(x_*, x_k), \tag{26}$$

$$\begin{aligned} \min_{x_* \in X^*} V(x_*, x_{k+1}) &\leq \min_{x_* \in X^*} V(x_*, x_k) + \left(\frac{\delta\mu}{M^2} - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + \\ &+ \left(\frac{\alpha\delta(1-2\gamma)}{M^2} + \frac{\sqrt{2}\delta}{M} \right) \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} - \frac{\delta^2}{M^2}. \end{aligned}$$

Условие (26) выполняется, если

$$\left(\frac{\delta\mu}{M^2} - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + \left(\frac{\alpha\delta(1-2\gamma)}{M^2} + \frac{\sqrt{2}\delta}{M} \right) \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} \leq 0,$$

то есть

$$\sqrt{\min_{x_* \in X^*} V(x_*, x_k)} \left(\frac{\delta\mu}{M^2} - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \sqrt{\min_{x_* \in X^*} V(x_*, x_k)} + \frac{\alpha\delta(1-2\gamma)}{M^2} + \frac{\sqrt{2}\delta}{M} \leq 0,$$

причем равенство возможно только в случае $V(x_*, x_k) = 0$.

По условию,

$$\delta \leq \frac{(1-2\gamma)\alpha^2}{2\mu}.$$

Тогда

$$\sqrt{\min_{x_* \in X^*} V(x_*, x_k)} \geq \frac{-2\alpha\delta(1-2\gamma) - 2\sqrt{2}\delta M}{2\delta\mu - (1-2\gamma)\alpha^2},$$

причем обе части этого неравенства — неотрицательные величины, и поэтому можно возвести его в квадрат, т. е.

$$\min_{x_* \in X^*} V(x_*, x_k) \geq \left(\frac{-2\alpha\delta(1-2\gamma) - 2\sqrt{2}\delta M}{2\delta\mu - (1-2\gamma)\alpha^2} \right)^2.$$

Следовательно, либо

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \min_{x_* \in X^*} V(x_*, x_k),$$

либо

$$\min_{x_* \in X^*} V(x_*, x_k) < \left(\frac{-2\alpha\delta(1-2\gamma) - 2\sqrt{2}\delta M}{2\delta\mu - (1-2\gamma)\alpha^2} \right)^2.$$

В случае когда неравенство (26) верно, справедливо соотношение

$$\min_{x_* \in X^*} V(x_*, x_{k+1}) \leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + C(\delta), \quad (27)$$

где $C(\delta) = \frac{\alpha^2\delta\gamma(1-\gamma)}{\mu M^2} + \frac{\sqrt{2}\alpha\delta\gamma}{\mu M} - \frac{\delta^2}{M^2}$. Далее, по рекурсии получим

$$\begin{aligned} \min_{x_* \in X^*} V(x_*, x_{k+1}) &\leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right) \min_{x_* \in X^*} V(x_*, x_k) + C(\delta) \leq \dots \leq \\ &\leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right)^{k+1} \min_{x_* \in X^*} V(x_*, x_0) + C(\delta) \sum_{i=0}^k \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right)^i \leq \\ &\leq \left(1 - \frac{(1-2\gamma)\alpha^2}{2M^2} \right)^{k+1} \min_{x_* \in X^*} V(x_*, x_0) + C(\delta) \frac{2M^2}{(1-2\gamma)\alpha^2}. \end{aligned}$$

□

5. Численные эксперименты

В настоящем параграфе для иллюстрации работоспособности предложенного выше алгоритма (4)–(6) и полученных для него результатов (в частности, уточненной оценки скорости сходимости из теоремы 1) приведем некоторые результаты вычислительных экспериментов для двух примеров. Все эксперименты проводились на Python 3.4 на компьютере с Intel(R) Core(TM) i7-8550U CPU (1,80 GHz, 4 ядра, 8 потоков). Оперативная память компьютера составляла 8 ГБ.

ПРИМЕР 1. Пусть P — гиперкуб в \mathbb{R}^n с m вершинами $(\pm 1, \pm 1, \dots, \pm 1)$, с разным распределением знаков $+$ и $-$ (оно подобрано случайно для m точек). Целевая функция имеет следующий вид:

$$f(x) = \text{dist}(x, P) + \text{dist}^2(x, P), \tag{28}$$

где $\text{dist}(x, P) = \min_{y \in P} \|x - y\|_2$. Ясно, что целевая функция (28) имеет α -острый минимум с $\alpha = 1$, а множество решений задачи (13) для этой целевой функции есть $X_* = P$, причем $f^* = 0$. Отметим, что шаг метода при реализации находился приближенно, что соответствует идеологии теоретических результатов, полученных в предыдущем пункте. Точнее говоря, чтобы найти $\text{dist}(x, P) = \min_{y \in P} \|x - y\|_2 = \|x - \widehat{x}\|_2$, где $\widehat{x} = \text{Proj}_P(x)$, находим проекцию приближенно следующим образом:

- 1) находим v_1, v_2 — две ближайшие вершины гиперкуба P к точке x ;
- 2) строим отрезок $v_1 v_2 = tv_1 + (1 - t)v_2$ для $t \in [0, 1]$;
- 3) берем N точек на отрезке $v_1 v_2$, а затем берем (из этих N точек) точку, дающую минимальное расстояние до точки x .

ПРИМЕР 2 (ЗАДАЧА ВОССТАНОВЛЕНИЯ ФАЗЫ). Восстановление фазы — распространенная вычислительная задача, имеющая приложения в различных областях, таких как визуализация, рентгеновская кристаллография и обработка речи [Davis, Drusvyatskiy, Paquette, 2020; Duchi, Ruan, 2019; Eldar, Mendelson, 2014]. Для простоты остановимся на версии задачи с вещественными числами. Задача (вещественная) восстановления фазы предполагает нахождение точки $x \in \mathbb{R}^n$, удовлетворяющей условиям

$$|\langle a_i, x \rangle|^2 \approx b_i \quad \text{для } i = 1, \dots, m,$$

где $a_i \in \mathbb{R}^n$ и $b_i \in \mathbb{R}$ заданы для каждого $i = 1, \dots, m$.

Таким образом, рассмотрим задачу (13) с целевой функцией вида

$$f(x) = \frac{1}{m} \sum_{i=1}^m |\langle a_i, x \rangle^2 - b_i|. \tag{29}$$

Как известно, эта функция μ -слабо выпукла при $\mu = 2 \max_{1 \leq i \leq m} \|a_i\|_2^2$ [Li et al., 2020].

Для примера 1 алгоритм (4)–(6) запускался с начальной точкой $x_0 = (10, \dots, 10) \in \mathbb{R}^n$. Результаты проведенных экспериментов для этого примера представлены на рис. 1 ниже. Эти результаты отражают динамику изменения приближенного значения целевой функции (28) на каждой итерации, где $\widehat{f}(x_k) = f(x_k) + \delta$ для небольшого значения шума δ , полученного в результате приближенного вычисления проекции $\text{Proj}_P(x_k)$ (см. рис. 1 слева) и оценки качества решения задачи (см. рис. 1 справа), которая является правой частью неравенства (9) с изменяющимися значениями $\alpha_i = \frac{f(x_i)}{\min_{v \in P} \text{dist}(x_i, v)}$ и фиксированным $\alpha_i = \alpha = 1 \ \forall 0 \leq i \leq k$.

Также на рис. 2 отражено время работы алгоритма (4)–(6) в секундах как функции размерности задачи n и m (количество вершин P). Заметим, что $\gamma_i = 0 \ \forall 0 \leq i \leq k$, так как целевая функция (28) выпукла.

Для примера 2 алгоритм (4)–(6) запускался для случайно сгенерированной начальной точки в \mathbb{R}^n с равномерным распределением на $[0, 1)$.

Векторы a_i и константы b_i для каждого $i = 1, \dots, m$ генерируются случайным образом по следующей процедуре: мы генерируем a_i и целевой сигнал x_* с нормальным (гауссовским)

распределением со средним значением (центром) в 0 и стандартным отклонением (шириной), равным 1. Затем полагаем $b_i = \langle a_i, x_* \rangle^2$ для каждого $i = 1, \dots, m$.

Результаты проведенных экспериментов для этого примера представлены на рис. 3. Эти результаты отражают сравнение оценки (9) с оценкой в [Davis et al., 2018], которая аналогична (9), но с фиксированным значением $\gamma_i = \gamma > 0 \forall i = 1, 2, \dots, k$. Отметим, что выбираются значения параметров $\gamma = \gamma_0 = 0,9$ и $\alpha_i = \min_{i \leq N} \frac{f(x_i)}{\min_{i \leq N} \|x_i - x_*\|_2}$ при $N = 100$.

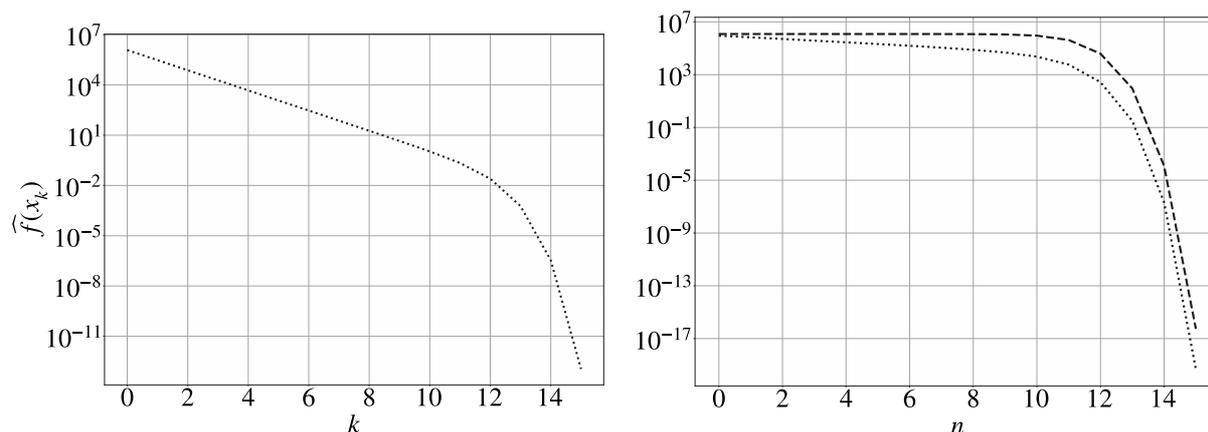


Рис. 1. Результаты алгоритма (4)–(6) для примера 1 с $n = 10\,000$ и $m = 16\,384$. На рисунке слева пунктирная линия показывает динамику приближенных значений $\widehat{f}(x_k)$ целевой функции (28) с ростом количества итераций. На правом рисунке пунктирная линия отражает теоретическую оценку с изменяющимися значениями α_i , а штриховая линия отражает теоретическую оценку с фиксированными значениями $\alpha_i = \alpha = 1$

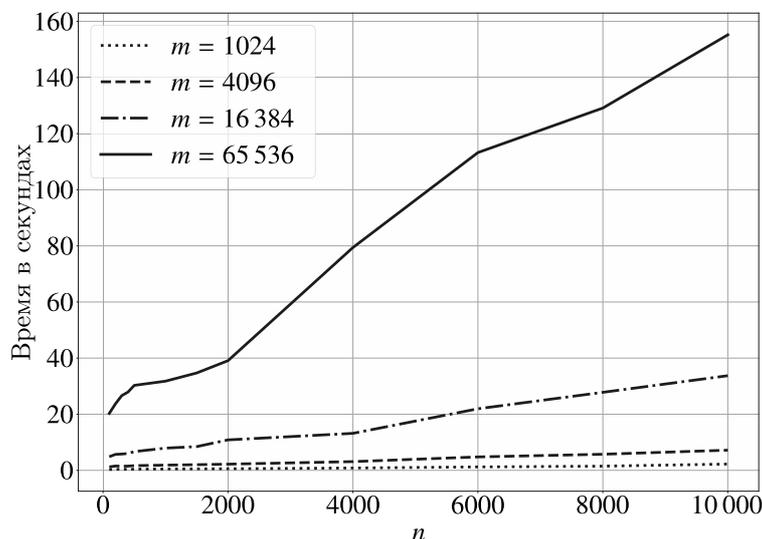


Рис. 2. Результаты алгоритма (4)–(6) для примера 1 для различных значений размерности пространства n и количества вершин гиперкуба m

На рис. 1 видно, что алгоритм (4)–(6) для примера 1 достиг решения рассматриваемой задачи с высоким качеством при использовании адаптивной процедуры вычисления α_i , т. е. без какой-либо заранее известной информации о параметре α . Также из рис. 2 видно, что время работы алгоритма (4)–(6) не сильно возрастает при увеличении размерности рассматриваемой задачи минимизации.

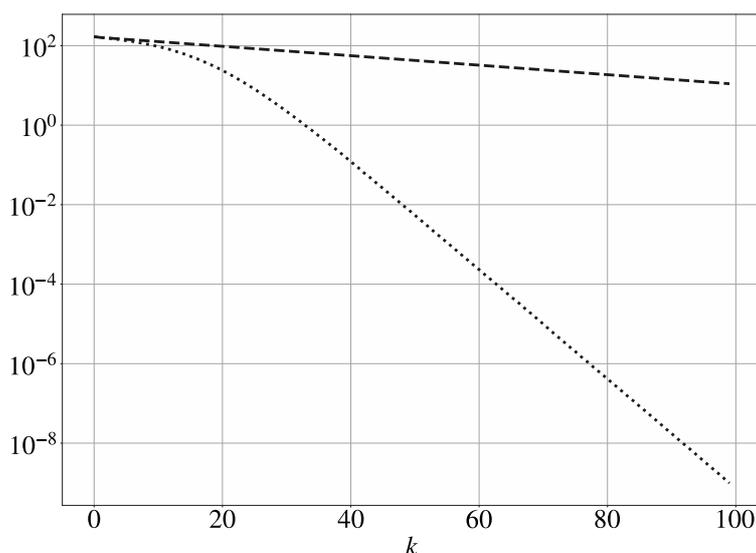


Рис. 3. Результаты алгоритма (4)–(6) для примера 2 с $n = 1000$ и $m = 5000$. Пунктирная линия отражает динамику теоретической оценки (9) с изменяющимися значениями параметров γ_i , а штриховая линия отражает динамику изменения теоретической оценки с фиксированными γ_i

На рис. 3 видно, что предложенная процедура подбора (7) может оказаться существенно лучше по сравнению с ранее известным аналогом [Davis et al., 2018, Theorem 3.1].

Теперь посмотрим на эффективность предложенной схемы (алгоритм 1) для задач с функциональным ограничением (10), где целевые функции f – (28) и (29) (которые приведены в примерах 1 и 2 соответственно). Функциональное ограничение g задается следующим образом:

$$g(x) = \max_{1 \leq i \leq K} \{g_i(x) = \langle \beta_i, x \rangle + c_i\}. \tag{30}$$

Здесь векторы $\beta_i \in \mathbb{R}^n$ и константы $c_i \in \mathbb{R} \forall 1 \leq i \leq K$ выбираются случайным образом из равномерного распределения по $[0, 1)$ таким образом, что для примера 2 имеем $g(x_*) < 0$.

Предложенный алгоритм 1 запускался для целевой функции (28) с $K = 1000, n = 5000, m = 8192$ вершин с начальной точкой $x_0 = (2, \dots, 2) \in \mathbb{R}^n$. Для целевой функции (29) алгоритм 1 запускался со случайно сгенерированной начальной точкой из равномерного распределения по $[0, 1), K = 5000, n = 1000, m = 5000$. Для обоих примеров мы берем $\varepsilon = 0,001$. Результаты проведенных экспериментов представлены на рис. 4. Эти результаты демонстрируют значение целевой функции (28) (слева) и (29) (справа) на продуктивных шагах $\{x_k; k \in I\}$. Для обоих примеров множество непродуктивных шагов непусто. У нас есть 47 непродуктивных шагов для примера 1 и 4 непродуктивных шага для примера 2.

6. Заключение

Итак, в статье получен ряд новых результатов о сходимости субградиентных методов с шагом Б. Т. Поляка и его аналогами на классе слабо выпуклых липшицевых задач с острым минимумом.

Во-первых, получено уточнение оценки теоремы 3.1 из [Davis et al., 2018], которое учитывает приближение траектории субградиентного метода по мере роста количества итераций (теорема 1).

Во-вторых, предложено обобщение острого минимума на класс задач со слабо выпуклыми функционалами-ограничениями и доказана возможность обобщения на соответствующий класс

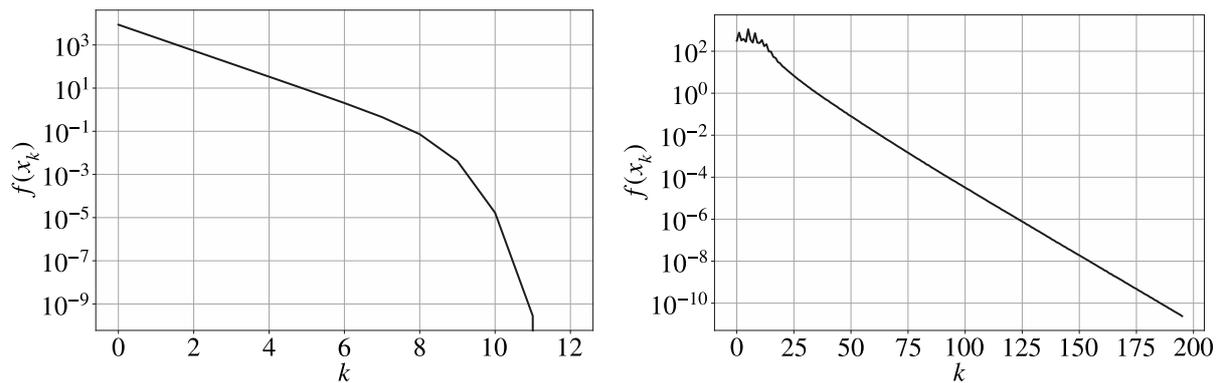


Рис. 4. Результаты алгоритма 1 для примера 1 (слева) и примера 2 (справа) с функциональным ограничением (30). Показана динамика значений $f(x_k)$

задач результата о сходимости субградиентного метода со скоростью геометрической прогрессии (теорема 2). Правда, при этом такой метод связан с заданием фиксированной точности ε , для повышения точности может потребоваться его перезапуск. Отметим, что, по-видимому, возможно рассматривать и вариации этого результата для других аналогов введенного в настоящей работе ε -острого минимума для задач с ограничениями-неравенствами (определение 1). Например, если (содержащее X_*) множество $Q_g = \{x \in Q \mid g(x) \leq 0\}$ есть выпуклый компакт в \mathbb{R}^n , то при $g(x) > \varepsilon$ можно рассматривать условие $g(x) \geq \alpha_g \min_{x' \in Q_g} \|x - x'\|_2$.

В-третьих, рассмотрен класс слабо выпуклых задач относительно дивергенции/расхождения Брэгмана и при соответствующих предположениях о начальной точке, относительно аналоге острого минимума, получен результат о линейной скорости для субградиентного метода с подходящим аналогом шага Б. Т. Поляка (теорема 3). Наконец, в теореме 4 описано влияние погрешности, доступной методу градиентного типа информации о (суб)градиенте и целевой функции, на гарантии качества выдаваемого методом решения. Доказано, что при достаточно малой погрешности $\delta > 0$ можно гарантировать достижение точности решения, сопоставимой с δ . Представляется на будущее интересной задачей исследование влияния погрешностей задания целевой функции и/или (суб)градиента как для метода с обычным шагом Б. Т. Поляка (5), так и для других возможных субградиентных методов на классе слабо выпуклых задач [Davis et al., 2018] с острым минимумом, с прицелом на доказательство оценок качества выдаваемого решения и для стохастических методов.

Отметим в заключение такой существенный момент. В статье мы опустили вопрос выбора субградиента для недифференцируемых слабо выпуклых задач, ограничившись лишь рассмотрением дифференцируемого случая. Однако все результаты можно обобщить на класс субдифференцируемых (например, по Кларку) функций для всех рассмотренных типов задач при условии произвольного выбора субградиента в текущей точке метода. Действительно, хорошо известно (см. [Дудов, Осипцев, 2021; Vial, 1983] и цитируемую там литературу), что слабо выпуклые функции являются локально липшицевыми. Поэтому они (как и выпуклые функции) дифференцируемы по всем направлениям в произвольной точке $x \in Q$, причем производная по направлению

$$f'(x, g) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha g) - f(x)}{\alpha},$$

для любых $g \in \mathbb{R}^n$ совпадает с обобщенной производной по направлению в смысле Ф. Кларка [Кларк, 1983]:

$$f'(x, g) = f^0(x, g) = \lim_{y \rightarrow x, \alpha \downarrow 0} \sup \frac{f(y + \alpha g) - f(y)}{\alpha}.$$

Следовательно,

$$f'(x, g) = \max_{v \in \partial f(x)} \langle v, g \rangle \quad \forall g \in \mathbb{R}^n,$$

где $\partial f(x)$ — субдифференциал Кларка функции $f(\cdot)$ в точке x :

$$\partial f(x) \equiv \{v \in \mathbb{R}^n : f^0(x, g) \geq \langle v, g \rangle \quad \forall g \in \mathbb{R}^n\}.$$

Таким образом, слабо выпуклые функции субдифференцируемы, причем субдифференциал Кларка совпадает с субдифференциалом Демьянова–Рубинова [Демьянов, Васильев, 1981; Демьянов, Рубинов, 1990].

Список литературы (References)

- Аблаев С. С., Макаренко Д. В., Стонякин Ф. С., Алкуса М. С., Баран И. В. Субградиентные методы для задач негладкой оптимизации с некоторой релаксацией условия острого минимума // Компьютерные исследования и моделирование. — 2022. — Т. 14, № 2. — С. 473–495.
- Ablaev S. S., Makarenko D. V., Stonyakin F. S., Alkousa M. S., Baran I. V. Subgradiyentnyye metody dlya zadach negladkoj optimizatsii s nekotoryy relaksatsiyey usloviya ostrogo minimuma [Subgradient methods for non-smooth optimization problems with some relaxation of sharp minimum] // Computer Research and Modeling. — 2022. — Vol. 14, No. 2. — P. 473–495 (in Russian).
- Демьянов В. Ф., Васильев Л. В. Недифференцируемая оптимизация. — М.: Наука, 1981.
- Dem'yanov V. F., Vasil'ev L. V. Nedifferentsiruemyaya optimizatsiya [Nondifferentiable optimization]. — Moscow: Nauka, 1981 (in Russian).
- Демьянов В. Ф., Рубинов А. М. Основы негладкого анализа и квазидифференциальное исчисление. — М.: Наука, 1990.
- Dem'yanov V. F., Rubinov A. M. Osnovy negladkogo analiza i kvazidifferencial'noe ischislenie [Fundamentals of nonsmooth analysis and quasi-differential calculus]. — Moscow: Nauka, 1990 (in Russian).
- Дудов С. И., Осипцев М. А. Характеризация решения задач сильно-слабо выпуклого программирования // Матем. сб. — 2021. — Т. 112, № 6. — С. 43–72.
- Dudov S. I., Osipcev M. A. Kharakterizatsiya resheniya zadach sil'no-slabo vypuklogo programmirovaniya [Characterization of solutions of strong-weak convex programming problems] // Sbornik: Mathematics. — 2021. — Vol. 112, No. 6. — P. 43–72 (in Russian).
- Кларк Ф. Оптимизация и негладкий анализ. — М.: Наука, 1988.
- Clarke F. Optimizatsiya i negladkii analiz [Optimization and nonsmooth analysis]. — Moscow: Nauka, 1988 (in Russian).
- Макаренко Д. В. Разработка численных методов решения задач оптимизации при ослабленных условиях гладкости: дисс. ... канд. физ.-мат. наук. — М.: Моск. физ.-техн. ин-т, 2022. — 90 с.
- Makarenko D. V. Razrabotka chislennykh metodov resheniya zadach optimizatsii pri oslablennykh usloviyakh gladkosti [Development of numerical methods for solving optimization problems under weakened smoothness conditions]. — Diss. ... kand. fiz.-mat. nauk. — Moscow: Mosk. fiz.-tekhn. in-t, 2022. — 90 p. (in Russian).
- Поляк Б. Т. Минимизация негладких функционалов // Ж. вычисл. матем. и матем. физ. — 1969. — Т. 9, № 3. — С. 509–521.
- Polyak B. T. Minimizatsiya negladkikh funktsionalov [Minimization of nonsmooth functionals] // Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki. — 1969. — Vol. 9, No. 3. — P. 509–521 (in Russian).
- Стонякин Ф. С., Титов А. А., Макаренко Д. В., Алкуса М. С. Численные методы для некоторых классов вариационных неравенств с относительно сильно монотонными операторами // Матем. заметки. — 2022. — Т. 112, № 6. — С. 879–894.
- Stonyakin F. S., Titov A. A., Makarenko D. V., Alkousa M. S. Chislennyye metody dlya nekotorykh klassov variatsionnykh neravenstv s otноситel'no sil'no monotonnymi operatorami [Numerical Methods for Some Classes of Variational Inequalities with Relatively Strongly Monotone Operators] // Math Notes. — 2022. — Vol. 112, No. 6. — P. 965–977 (in Russian).
- Davis D., Drusvyatskiy D., Kellie M. Stochastic model-based minimization under high-order growth // arXiv preprint. — 2018. — <https://arxiv.org/pdf/1807.00255.pdf>
- 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.

- Davis D., Drusvyatskiy D., Paquette C.* The nonsmooth landscape of phase retrieval // IMA Journal of Numerical Analysis. — 2020.— Vol. 40, No. 4. — P. 2652–2695.
- Duchi J., Ruan F.* Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval Information and Inference // A Journal of the IMA. — 2019.— Vol. 8, No. 3. — P. 471–529.
- Eldar Y., Mendelson S.* Phase retrieval: stability and recovery guarantees // Appl. Comput. Harmon. Anal. — 2014.— Vol. 36, No. 3. — P. 473–494.
- Huang Y., Lin Q.* Single-loop switching subgradient methods for non-smooth weakly convex optimization with non-smooth convex constraints // arXiv preprint. — 2023. — <https://arxiv.org/pdf/2301.13314.pdf>
- Li X., Zhu Z., Man-Cho So A., Lee J.* Incremental methods for weakly convex optimization // OPT2020: 12th Annual Workshop on Optimization for Machine Learning. — 2020. — <https://arxiv.org/pdf/1907.11687v1.pdf>
- Lu H.* Relative continuity for non-Lipschitz nonsmooth convex optimization using stochastic (or deterministic) mirror descent // Informs Journal on Optimization. — 2019.— Vol. 1, No. 4. — P. 288–303.
- Mai V., Johansson M.* Stability and convergence of stochastic gradient clipping: Beyond Lipschitz continuity and smoothness // arXiv preprint. — 2021. — <https://arxiv.org/pdf/2102.06489.pdf>
- Nesterov Yu.* Relative smoothness: New paradigm in convex optimization // EUSIPCO-2019, A Coruna, Spain, September 4. — Conference report. — 2019.
- Vial J.P.* Strong and weak convexity of sets and functions // Math. Oper. Res. — 1983. — Vol. 8, No. 2. — P. 231–259.